

**UNIVERSIDADE DO VALE DO RIO DOS SINOS
UNIDADE ACADÊMICA DE GRADUAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

LUCAS MAYER CESCHINI

**AUXILIANDO O DIAGNÓSTICO CLÍNICO DE GLAUCOMA A PARTIR DA
SEGMENTAÇÃO E CLASSIFICAÇÃO AUTOMATIZADA DE IMAGENS DE FUNDO
DE OLHO**

São Leopoldo
2021

LUCAS MAYER CESCHINI

**AUXILIANDO O DIAGNÓSTICO CLÍNICO DE GLAUCOMA A PARTIR DA
SEGMENTAÇÃO E CLASSIFICAÇÃO AUTOMATIZADA DE IMAGENS DE FUNDO
DE OLHO**

Artigo apresentado como requisito parcial para
obtenção do título de Bacharel em Ciência da
Computação, pelo Curso de Ciência da Compu-
tação da Universidade do Vale do Rio dos Sinos
(UNISINOS)

Orientador(a): Prof. Dr. Gabriel de Oliveira Ra-
mos

São Leopoldo
2021

AUXILIANDO O DIAGNÓSTICO CLÍNICO DE GLAUCOMA A PARTIR DA SEGMENTAÇÃO E CLASSIFICAÇÃO AUTOMATIZADA DE IMAGENS DE FUNDO DE OLHO

Lucas Mayer Ceschini¹ Gabriel de Oliveira Ramos²

Resumo: Glaucoma é o maior responsável pela perda irreversível de visão, sendo a segunda maior causa de cegueira no mundo. Sua natureza discreta faz com que os primeiros sinais surjam apenas em um estado avançado da doença, quando não há mais volta. O diagnóstico prévio é de suma importância, e atualmente é realizado, principalmente, via exames de fundo de olho, de pressão intraocular, ou de tomografia da retina, sendo que o primeiro é o mais popular e acessível. O exame de fundo de olho é um procedimento manual e tedioso, propenso a erros que podem resultar em falsos negativos, por sua vez implicando em perda de visão. Para combater este problema, soluções tecnológicas estão sendo estudadas. Abordagens clássicas de aprendizado de máquina, como Support Vector Machine, são utilizadas para classificar entre glaucoma ou normal com auxílio de técnicas de visão computacional. Abordagens de aprendizado profundo são utilizadas para classificar puramente a partir da imagem de fundo de olho ou de tomografia computadorizada. Aparelhos capazes de registrar digitalmente a imagem de fundo de olho já são largamente utilizados e, em conjunto com sistemas inteligentes capazes de aprender a classificar sozinhos essas imagens, torna-se possível projetar uma ferramenta de auxílio ao diagnóstico de doenças oculares, como o glaucoma por exemplo. Este estudo tem como base uma arquitetura composta por duas redes de segmentação e uma rede de classificação, que, além de darem o seu diagnóstico, apresentam ao especialista uma representação visual das estruturas oculares detectadas, o cálculo da relação escavação/disco, métrica de suma importância para o diagnóstico de glaucoma, e também a confiança do sistema naquela previsão. As contribuições do presente estudo são duplas: A adição de um segundo subsistema classificador, e a união dos dois subsistemas de segmentação. Os resultados são promissores e abrem espaço para discussões futuras. A adição de um novo sistema de classificação, e o posterior agrupamento deste com o classificador atual, aumentou a sensibilidade em 3%, significando uma expressiva redução na taxa de falsos negativos. A união das redes de segmentação, mesmo que sem apresentar melhorias nas métricas avaliadas, foi capaz de simplificar a arquitetura do modelo, reduzindo a quantidade total de parâmetros do modelo em 50%, de 4.899.206 para 2.449.603. Reduziu também o seu custo de armazenamento e processamento, alcançando os mesmos níveis de eficácia que as redes individuais, tendo como consequência apenas um aumento no número de épocas necessárias de treinamento, indo de 100 para 300 épocas.

Palavras-chave: Glaucoma. Aprendizado de Máquina. Visão Computacional. Fundoscopia.

¹Graduando em de Ciência da Computação pela Unisinos. Email: lceschini@edu.unisinos.br

²<http://lattes.cnpq.br/9281736089055094>. Email: gdoramos@unisinos.br

1 INTRODUÇÃO

Glaucoma é uma doença ocular responsável pelo maior número de perda irreversível de visão do mundo, sendo a segunda maior causa de cegueira, ficando atrás somente da catarata (MANTRAVADI; VADHAR, 2015). Os tipos mais comuns de glaucoma são o glaucoma de ângulo aberto e o glaucoma de ângulo fechado. O glaucoma de ângulo fechado é precedido de um súbito aumento na pressão intraocular, causando uma dor aguda. Já o glaucoma de ângulo aberto se desenvolve de forma silenciosa e indolor, apresentando como primeiros sintomas a perda da visão periférica, em um estágio da doença onde já não há mais volta (MANTRAVADI; VADHAR, 2015). Por conta disso, este tipo de glaucoma é o mais comum e o mais perigoso, sendo responsável por até 80% dos casos (MANTRAVADI; VADHAR, 2015).

O diagnóstico prévio da doença é de suma importância, para que assim se possa evitar a perda irreversível da visão. Este grupo de doenças possui a degeneração do nervo óptico como sua principal patologia, e é causado principalmente devido ao entupimento ou bloqueio parcial da malha trabecular, que resulta em um aumento na pressão intraocular (NICE, 2017). Este aumento de pressão causa uma característica ampliação no ângulo do copo do disco óptico, erosão no tecido de borda e, conseqüentemente, defeitos no campo de visão. Ocorre normalmente em adultos a partir de 40 anos, sendo intensificada a sua presença em idosos com mais de 75 anos (NICE, 2017). Em 2013, o número mundial de pessoas entre 40 à 80 anos com glaucoma era estimado em 64.3 milhões, aumentando para 76.0 milhões em 2020 e 111.8 milhões em 2040, consequência direta do comportamento insidioso da doença e das condições socioeconômicas da população, que não conseguem bancar idas rotineiras ao médico (THAM et al., 2014).

A identificação do glaucoma pode ser feita a partir de diversos exames diferentes, tal como a tonometria, fundoscopia (oftalmoscopia), perimetria e paquimetria (CIVIT-MASOT et al., 2020b). A avaliação da cabeça do nervo óptico a partir da fundoscopia é um dos exames mais familiares e acessíveis de serem executados pelos peritos, devido à sua abordagem pouco invasiva e sua capacidade de gerar imagens de alta resolução das estruturas internas do olho (CIVIT-MASOT et al., 2020b). Isso não quer dizer que os outros exames não estejam sendo utilizados pela comunidade científica para avaliar glaucoma automaticamente a partir de aprendizado de máquina, porém a fundoscopia é a abordagem mais comum e mais explorada.

A avaliação de imagens de fundo de olho obtidas a partir de ferramentas de fundoscopia possibilita a identificação da escavação responsável por ampliar o ângulo do copo do disco óptico, característica mais comum do glaucoma de ângulo aberto (AHN et al., 2018). O exame de fundoscopia é um processo manual, tedioso e complexo, propenso a erros capazes de culminar na perda irreversível de visão de pacientes que receberam um prognóstico falso negativo de médicos cansados ou inexperientes (AHN et al., 2018).

A automação deste procedimento é discutida na comunidade científica, e abordagens clássicas de aprendizado de máquina, como Support Vector Machine, estão sendo aplicadas com auxílio de técnicas de visão computacional (DEY; BANDYOPADHYAY, 2016) para realizar

este diagnóstico. Mais recentemente, modelos de aprendizado profundo estão sendo explorados para classificar glaucoma a partir das imagens de fundo de olho, obtendo resultados extremamente promissores.

Este estudo tem como objetivo aperfeiçoar um modelo de auxílio no diagnóstico médico de glaucoma elaborado por Civit-Masot et al. (2020b), indo além de um obscuro sistema de diagnóstico, este projeto tem o intuito de apoiar o especialista em sua decisão, apresentando representações visuais das estruturas do olho, além de uma classificação direta, o grau de confiança nela e um cálculo da relação escavação/disco, fundamental para o prognóstico. A arquitetura do modelo base consiste em uma rede de classificação direta baseada em MobileNetV2 (SANDLER et al., 2018), e duas redes de segmentação, uma para o disco e outra para o copo do olho, ambas baseadas em um modelo U-net (RONNEBERGER; FISCHER; BROX, 2015) modificado. A arquitetura deste estudo consiste em uma rede de classificação baseada em *Ensemble Learning* das arquiteturas MobileNetV2 e EfficientNet, e uma rede de segmentação unificada para o disco e copo.

As contribuições deste estudo são: A) a adição de uma segunda rede de classificação leve, EfficientNetB1 (TAN; LE, 2019), que é agrupada com a rede já existente e que resulta em um *aumento de 3% na sensibilidade*, reduzindo assim expressivamente a quantidade de falsos negativos, tão perigosos para o ambiente clínico. E B) a união de duas redes de segmentação, uma para o disco e outra para o copo, resultando na mesma eficácia porém de uma forma mais simples e leve, indo de quatro milhões de parâmetros para dois milhões, de suma importância para a eventual implantação deste modelo na clínica médica.

As seções a seguir estão organizadas da seguinte forma: 1) Fundamentação teórica, com a apresentação dos conceitos chave; 2) Trabalhos relacionados levantados a partir de uma revisão da literatura; 3) Proposta de melhoria do modelo base; 4) Estudos preliminares com as metodologias adotadas e os resultados da implementação e 5) Conclusão do projeto e discussão de trabalhos futuros.

2 FUNDAMENTAÇÃO

Nesta seção, será explicado os conceitos necessários para que seja elaborado um classificador automático de glaucoma. Inicialmente, o contexto será explanado, discorrendo sobre a doença, a patologia e como é feito o seu diagnóstico. Em seguida, será abordados os conceitos de aprendizado de máquina, redes neurais artificiais, aprendizado profundo e redes neurais convolucionais.

2.1 Glaucoma

Glaucoma é um grupo de doenças que possuem a degeneração do nervo óptico como sua principal patologia, podendo resultar em perda de visão e cegueira (MANTRAVADI; VADHAR, 2015). O seu tipo mais comum é o glaucoma primário de ângulo aberto, causado devido ao entupimento ou bloqueio parcial da malha trabecular, um canal que fica no fundo do olho e é responsável por drenar o humor aquoso que passa por ali (MANTRAVADI; VADHAR, 2015).

Esse bloqueio acaba aumentando a pressão intraocular, conseqüentemente degenerando os tecidos do olho, causando uma característica ampliação no ângulo do copo do disco óptico, erosão no tecido de borda e, conseqüentemente, defeitos no campo de visão (MANTRAVADI; VADHAR, 2015). Na maioria dos casos, esse processo é lento e silencioso, só demonstrando sinais, como a perda progressiva no campo de visão, em um estágio avançado da doença, quando não há mais a possibilidade de tratamento (MANTRAVADI; VADHAR, 2015).

Glaucoma é a segunda maior causa de cegueira no mundo, sendo a maior causa de perda irreversível da visão (MANTRAVADI; VADHAR, 2015). Na grande maioria dos casos, mas nem sempre, é precedida de hipertensão ocular. Ocorre normalmente em adultos a partir de 40 anos, sendo intensificada a sua presença em idosos com mais de 75 anos (NICE, 2017).

Em 2013, o número mundial de pessoas entre 40 à 80 anos com glaucoma era estimado em 64.3 milhões, aumentando para 76.0 milhões em 2020 e 111.8 milhões em 2040, consequência direta do comportamento insidioso da doença e das condições socioeconômicas da população, que não conseguem bancar idas rotineiras ao médico (THAM et al., 2014).

2.2 Diagnóstico

De acordo com Gupta et al. (2015), o nervo óptico deve ser avaliado por sinais característicos de glaucoma. Dependendo do grau de dano ao nervo, diferentes tratamentos deverão ser executados.

Danos iniciais ao nervo óptico podem incluir o copo do nervo óptico maior ou igual a 0.5, defeitos na camada de fibra nervosa da retina, afinamento de borda focal, escavação vertical, assimetria de disco, escavação focal e hemorragia de disco (BARROS et al., 2020).

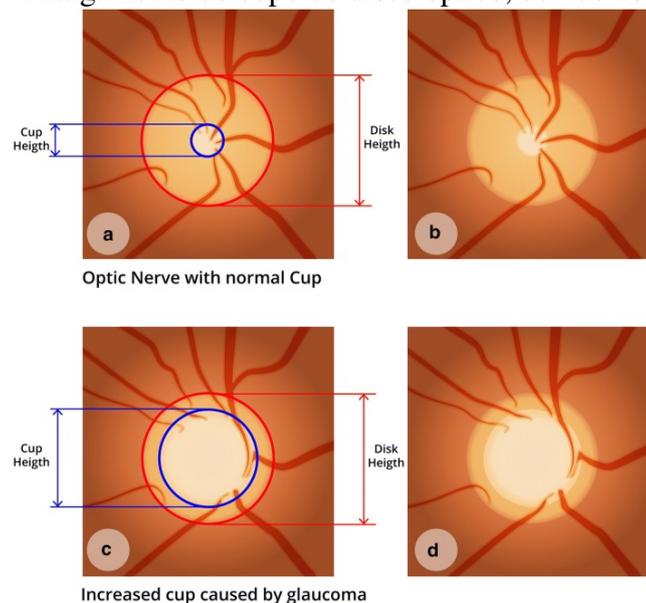
Danos moderados à avançados podem incluir um grande aumento no ângulo do copo do

nervo óptico, maior ou igual a 0.7, defeitos difusos na camada de fibra nervosa da retina, afinamento difuso de borda focal, escavação do nervo óptico, surgimento de um fosso no nervo óptico e hemorragia de disco (BARROS et al., 2020). A figura 1 mostra o alargamento do copo do nervo óptico, comum nos casos de glaucoma de ângulo aberto.

Essas características podem ser medidas de forma qualitativa e quantitativa por diversos exames diferentes, tal como tonometria, fundoscopia, perimetria, gonioscopia e paquimetria (CIVIT-MASOT et al., 2020b).

A avaliação da cabeça do nervo óptico a partir da fundoscopia é um dos exames mais familiares e acessíveis de serem executados pelos peritos, sendo especialmente eficaz devido à sua abordagem menos invasiva e suas imagens de alta resolução da estrutura interna do olho (AHN et al., 2018). Com isso é possível de se detectar a doença nos seus estágios iniciais, consequentemente aumentando as oportunidades de tratamentos (AHN et al., 2018).

Figura 1 – Alargamento do copo do disco óptico, devido ao glaucoma.



Fonte: (BARROS et al., 2020)

2.3 Aprendizado de Máquina Supervisionado

Aprendizado de máquina (AM) é uma técnica muito comum em sistemas de inteligência artificial onde a solução de um problema é desconhecida em um primeiro momento (RUSSELL; NORVIG, 2010). Nesses casos, agentes inteligentes com aprendizado formulam sua solução a partir de características e exemplos sobre o problema que estão enfrentando, continuamente melhorando seu desempenho a partir de novas experiências (RUSSELL; NORVIG, 2010).

O aprendizado de máquina supervisionado é uma das três grandes categorias desta técnica, tais modelos recebem um conjunto de exemplos e características rotulados, aprendem a partir deles e assim conseguem classificar ou prever dados não rotulados (RUSSELL; NORVIG,

2010).

Tais técnicas estão sendo altamente investigadas pelo setor da saúde (CABITZA; RASOINI; GENSINI, 2017). A acurácia de modelos de AM está sendo maior que a de alguns especialistas, que estão tirando proveito de tais métodos para auxiliá-los na detecção de doenças e de padrões patológicos (CABITZA; RASOINI; GENSINI, 2017). É previsto mudanças substanciais no setor médico nos próximos anos, devido ao crescimento de sistemas automatizados que utilizam técnicas de aprendizado de máquina (CABITZA; RASOINI; GENSINI, 2017).

Para a detecção automatizada de glaucoma, medidas oculares computadas por um aparelho de tomografia óptica, como a grossura da camada de fibra nervosa da retina e a cabeça do nervo óptico, foram passadas para modelos de AM, que aprenderam a classificar os olhos como glaucomatosos ou saudáveis (HUANG; CHEN, 2005).

Uma segunda abordagem utilizou imagens de fundo de olho e técnicas de segmentação para identificar o disco e o copo óptico, assim calculando a relação escavação/disco (cup-to-disc ratio), uma das métricas mais relevantes no diagnóstico de glaucoma (YU et al., 2019).

Uma imagem gerada por aparelhos de telemedicina pode conter milhares de pixels, o que a torna inviável para alguns modelos de aprendizado de máquina. Para isso, técnicas de segmentação de imagem em conjunto com visão computacional foram aplicadas em diversos estudos para extrair características relevantes, como cor, textura ou formato, e assim otimizar e simplificar os modelos (BARROS et al., 2020). O anexo A exemplifica uma arquitetura genérica comumente utilizada para extração de características de uma imagem.

2.4 Redes Neurais Artificiais

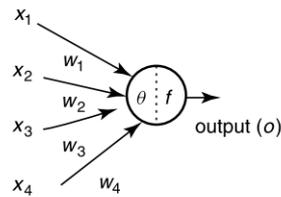
Uma rede neural artificial é a união de componentes matemáticos, chamados neurônios artificiais, distribuídos e conectados em camadas, chamadas perceptrons, a fim de formar uma rede interconectada (ABRAHAM, 2005). Esses neurônios artificiais foram inspirados na biologia do cérebro, e seu funcionamento se dá de forma parecida, mesmo que extremamente simplificada.

Em um neurônio artificial, sinais de entrada são recebidos, seja do exterior, como dados de entrada, ou de outros neurônios, para que assim possam ser computados por uma função matemática em conjunto com um vetor de pesos únicos, que são previamente definidos (ABRAHAM, 2005).

Assim que um determinado sinal de entrada é recebido, é computado a somatória do mesmo em conjunto com os pesos do neurônio, passando essa somatória para uma função não linear, chamada de função de ativação. Caso o resultado ultrapasse um determinado valor de corte, esse neurônio será ativado, por consequência passando para frente os seus valores computados (ABRAHAM, 2005). Por outro lado, caso esse limiar não tenha sido atingido, o neurônio não será ativado e seu retorno será nulo (ABRAHAM, 2005). A figura 2 exemplifica sua estrutura.

Cada um desses neurônios são interconectados para formar uma rede neural artificial. Ela é composta por camadas de entrada, camadas escondidas e camadas de saída. A camada de

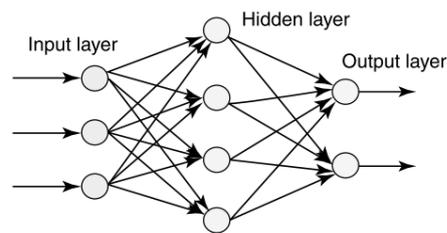
Figura 2 – Exemplo de um neurônio artificial.



Fonte: (ABRAHAM, 2005)

entrada é alimentada por dados pertinentes ao problema a ser resolvido, as camadas escondidas processarão esses dados, ativando seus neurônios ou não, e passarão seus resultados para a camada final, que retornará o resultado final (ABRAHAM, 2005). Esse tipo de rede é unilateral, sendo exemplificada na figura 3. Nesses casos o fluxo segue da direita para a esquerda, terminando seu processo com a saída final.

Figura 3 – Exemplo de uma rede neural artificial.



Fonte: (ABRAHAM, 2005)

A partir desses conceitos iniciais, é possível inserir técnicas de aprendizado de máquina, para assim usufruir das capacidades computacionais da arquitetura de redes neurais artificiais, com a possibilidade de otimização a partir de aprendizado. Seguindo o aprendizado supervisionado apresentado acima, uma rede neural pode receber como entrada os dados acompanhados de seus rótulos ou classificações (ABRAHAM, 2005).

Depois que os dados passam pelas camadas escondidas e a sua saída é retornada, uma função derivativa pode ser empregada, para que assim os pesos de cada um dos neurônios possam ser ajustados, modificando assim os neurônios que serão ativados. Diversas rodadas desse procedimento serão empregadas para que assim o erro da classificação ou previsão da rede neural seja minimizado, ajustando os pesos no seu valor otimizado para resolver uma determinada tarefa (ABRAHAM, 2005).

2.5 Aprendizado Profundo

Modelos clássicos ou tradicionais de aprendizado de máquina supervisionado podem alcançar uma eficácia muito boa, porém dependem de diversos fatores. O pré-processamento e a

segmentação de imagens são alguns deles, porém muitas vezes são processos dispendiosos que, com adição da variância na qualidade da imagem, estão sujeitos à erros e dependem de uma engenharia manual e especializada (MAETSCHKE et al., 2019).

Para resolver as dificuldades encontradas na extração de características úteis em grande volumes de dados não classificados, como por exemplo uma imagem contendo milhares de pixels, foi desenvolvido uma abordagem chamada aprendizado profundo. Estes modelos nada mais são do que uma rede neural artificial com múltiplas camadas, isso é, um conjunto de camadas de processamento, cada uma realizando uma determinada transformação nos dados de entrada, a fim de extrair e elaborar características ou classificações na sua saída (LECUN; BENGIO; HINTON, 2015).

Tais métodos revolucionaram áreas como visão computacional e processamento de linguagem natural, pois tiram vantagem da quantidade cada vez maior de dados e processamento disponíveis, e são capazes de aprender com a mínima influência ou expertise humana (LECUN; BENGIO; HINTON, 2015).

Diferente dos modelos de aprendizado de máquina clássicos, que precisam de uma engenharia manual, as arquiteturas de aprendizado profundo, comumente chamadas de redes neurais profundas, recebem como entrada um dado "crú", sem nenhuma marcação prévia, e em cada camada vão aplicando pequenas modificações e transformações nos dados, que vão sendo passados para as camadas seguintes até que por fim uma classificação ou marcação é computada pela camada final (LECUN; BENGIO; HINTON, 2015).

Por ser uma abordagem supervisionada, esse valor de saída computado é então comparado com o valor esperado, e de acordo com a diferença entre eles, é realizado operações matemáticas derivativas, desde as últimas camadas até as primeiras, para assim ajustar as variáveis, ou pesos, das funções. Repetindo esse processo inúmeras vezes com o intuito de aprender os valores ideais para reduzir ao máximo o erro (LECUN; BENGIO; HINTON, 2015).

2.6 Redes Neurais Convolucionais

Quando o assunto é o reconhecimento de imagens, existe um tipo de redes neurais profunda que se destaca. É o caso das Redes Neurais Convolucionais (RNC), um modelo com os mesmos fundamentos das redes neurais, que normalmente é utilizado em conjunto com elas, porém modificado de tal forma à simplificar e otimizar arquiteturas de reconhecimento e classificação de imagens. Pois, quando se trata de imagens em alta qualidade e com três canais de cores, por exemplo RGB, uma rede neural profunda acaba tendo que lidar com milhões de variáveis, tornando-a inviável, tanto computacionalmente como temporalmente (O'SHEA; NASH, 2015).

Em uma RNC, cada camada da rede consiste em operações matemáticas de convolução, feitas em cima da entrada e de um filtro, normalmente uma matriz de tamanho variado, que irá processar e devolver uma imagem levemente transformada, realçando contornos ou exibindo partições relevantes da imagem original, enquanto deixa de lado todo o resto (O'SHEA; NASH,

2015).

Depois de algumas operações como essa, junto com outras operações responsáveis por alterar os pedaços da imagem que serão computados ou até alterar o tamanho da imagem de saída de cada camada, as camadas mais avançadas da rede neural irão receber como entrada somente os traços relevantes da imagem, ou trechos específicos da imagem que se destacaram, englobando assim um objeto, por exemplo, e assim, por fim, serão capazes de classificar e reconhecer uma imagem com uma acurácia e performance impressionante (ALBAWI; MOHAMMED; AL-ZAWI, 2018).

A grande vantagem de uma arquitetura como essa é que, ao invés de olhar para a imagem inteira com seus milhares de pixels e três canais de cores, cada camada da rede pode focar em uma região da imagem, reconhecendo padrões e realçando certas características independente de onde elas estão posicionadas na imagem (ALBAWI; MOHAMMED; AL-ZAWI, 2018). O anexo C demonstra a arquitetura de um modelo classificador de glaucoma que utiliza redes neurais convolucionais.

2.7 Discussão

Glaucoma é a principal causa de perda de visão no mundo, sendo a segunda maior causa de cegueira no mundo. Sua natureza silenciosa, somada com fatores sociais e culturais que reduzem a frequência de revisões oculares, fazem com que a mesma só seja detectada, na maioria dos casos, nos seus estágios mais avançados, onde ocorre perdas no campo de visão e não há mais oportunidades de reversão do quadro clínico. Devido a esses fatores, a estimativa de crescimento no número de casos dessa doença é cada vez mais alarmante.

Para combater isso, é de suma importância a elaboração de estudos e modelos computacionais capazes de auxiliar médicos oftalmos no diagnóstico prévio da doença. Fundoscopia com imagens de fundo de olho se popularizou no diagnóstico de doenças oculares devido à suas imagens de alta resolução, baixo custo e capacidade de análise das estruturas do olho.

Modelos computacionais de aprendizado de máquina foram treinados tanto com as métricas (modelos clássicos), como com as imagens de fundo de olho (modelos de aprendizado profundo). Sua acurácia está nos mesmos patamares daquela dos especialistas, comprovando assim sua eficácia e capacidade de auxílio aos médicos que possuem a difícil tarefa de diagnosticar doenças como glaucoma em seus estágios iniciais.

Conclui-se, então, que para reduzir as incidências de casos de perda de visão e até cegueira, é de vital importância estudos que elaborem modelos e técnicas computacionais capazes de auxiliar os especialistas no diagnóstico prévio de glaucoma, sendo os mais promissores e eficazes aqueles que utilizam abordagens de aprendizado de máquina em conjunto com imagens de fundo de olho.

3 TRABALHOS RELACIONADOS

Afim de nortear e auxiliar na condução desta pesquisa, uma breve revisão sistemática da literatura foi realizada, seguindo os protocolos elaborados por Kitchenham e Charters (2007). Tal protocolo pode ser dividido em quatro etapas principais: **1)** Definir as questões de pesquisa; **2)** Definir o processo de pesquisa; **3)** Definir os critérios para seleção dos artigos e **4)** Analisar os artigos selecionados para assim responder as questões levantadas.

3.1 Questões de Pesquisa

Foram definidas seis questões de pesquisa, são elas: **1)** Quais são as técnicas ou tecnologias comumente utilizadas para classificação automatizada de glaucoma? **2)** Que tipo de métricas são utilizadas para medir a eficácia e acurácia dos modelos classificadores de glaucoma? **3)** Quais são os conjuntos de dados (datasets) utilizados para treinar e testar os modelos classificadores de glaucoma? **4)** Quais são as questões ainda em aberto quando se trata de modelos classificadores de glaucoma? **5)** Quais são os desafios encontrados na modelagem de algoritmos classificadores de glaucoma? **6)** Quais são os modelos, utilizados atualmente, que apresentam a maior acurácia e eficácia?

3.2 Processo de Pesquisa

O processo de pesquisa elaborado foi o seguinte: **1)** Especificar a *string* de busca; **2)** Definir as bases de dados e **3)** Aplicar a *string* nessas bases para colher os artigos retornados. A *string* de busca foi estruturada como: *((Machine OR Deep) AND Learning AND Glaucoma AND (Detection OR Classification))*. Em seguida foi definido *Google Scholar* como a base digital que será utilizada, por ser abrangente e reunir artigos dos principais periódicos.

3.3 Critérios de Inclusão e Exclusão

Foi definido critérios de inclusão (CI) e de exclusão (CE) para filtrar os estudos coletados das bases de pesquisa e assim melhorar a qualidade da pesquisa.

Os critérios de inclusão são: **1)** O estudo deve ser publicado em uma conferência, revista ou periódico; **2)** O estudo deve ser um artigo completo e **3)** o estudo deve ser relacionado a classificadores de glaucoma automatizados.

Já os critérios de exclusão definidos foram: **1)** Estudo publicado antes de 2011; **2)** Estudo não está escrito em inglês; **3)** Estudo tese ou publicado em *workshops*; **4)** Estudo não relacionado às questões de pesquisa; **5)** Estudo relacionado ao glaucoma, porém não ao seu diagnóstico automatizado e **6)** Estudos duplicados.

3.4 Resultados

A primeira busca nas bases, a partir da *string* de busca e aplicando filtros nos resultados para que assim fosse apresentado apenas aqueles publicados após 2011 e que são artigos completos, resultou em 252 artigos. Em cima desses 252 artigos foi feito um filtro por título, onde foram coletados 98 artigos. Em seguida foi feita a revisão dos selecionados a partir do seu resumo, reduzindo assim os artigos de interesse para 25. Esses últimos 25 artigos foram lidos por completo, resultando por fim em 13 artigos selecionados.

A partir desses artigos será respondido as questões de pesquisa previamente levantadas. O apêndice A lista os artigos selecionados.

3.4.1 Quais são as técnicas ou tecnologias comumente utilizadas para classificação automatizada de glaucoma?

A classificação normalmente é efetuada a partir de duas etapas: A segmentação do disco óptico e a classificação a partir da imagem segmentada. Para a segmentação, Yu et al. (2019) utilizou uma U-net modificada com a ResNet-34 como encoder, conseguindo a performance de ponta em termos de valores dice, para o copo e o disco.

Além desse autor, Civit-Masot et al. (2020b) e Xu et al. (2021) também utilizaram uma rede U-shaped para a segmentação. Sendo que Civit-Masot et al. (2020b) utilizou também RANSAC para calcular e melhorar a forma elíptica esperada e Xu et al. (2021) segmentou além de disco e copo óptico, também os defeitos da camada de fibra nervosa da retina.

Em Sreng et al. (2020), utilizou-se uma rede que ainda não tinha sido explorada pelos pesquisadores para segmentação, a DeepLabV3+, testando a performance da mesma com 5 CNNs diferentes, sendo a melhor combinação DeepLabV3+ com MobileNet.

Fora U-net, Fully Convolutional Network (FCN) (CHAI; LIU; XU, 2018) e ResNet-50 (PHASUK et al., 2019) também foram exploradas para segmentação. Para a classificação, Sallam et al. (2021) pegou várias arquiteturas de RNCs pré-treinadas e treinou todas usando *transfer learning* a partir do mesmo conjunto de dados, comparando assim as suas performances. Nesse estudo foram avaliadas as redes AlexNet, VGG11-19, InceptionV1, ResNet-18,50,101 e 152.

Já em Civit-Masot et al. (2020b), focou-se na aplicabilidade clínica ao fazer um modelo embarcado que explica sua previsão, utilizando para isso MobileNetV2, uma rede leve que conseguiu performance comparável com o estado da arte porém com um custo computacional até 20 vezes menor.

Os autores de Chai et al. (2021) usufruíram de dados heterogêneos, como imagens de retina, indicadores médicos e textos de reclamações dos pacientes para a classificação. Para isso, integraram todos esses dados a partir de um método de *representation integration* e realizaram a previsão com um modelo Bayesiano de aprendizado profundo.

Em Ahn et al. (2018) foi elaborado três modelos para classificar glaucoma, o primeiro usava regressão logística, o segundo InceptionV3 com *transfer learning* e o último foi uma RNC arquitetada manualmente pelos autores. O modelo clássico de regressão logística obteve uma boa performance na classificação de glaucoma avançado, porém não performou bem nos casos de glaucoma leve ou inicial. O modelo InceptionV3 teve uma boa performance em ambos os casos de glaucoma, devido à importância que tem a imagem de fundo de olho para a classificação de glaucoma inicial, mas o melhor dentre os três foi o modelo elaborado pelos autores.

Isso se dá devido ao fato de que o modelo foi refinado manualmente na base da tentativa e erro, afim de descobrirem os melhores parâmetros. O estudo de Phasuk et al. (2019) combinou os resultados de diversas redes integradas com a DenseNet-121 para a extração de características e uma rede neural para a classificação final. Uma dessas redes usava a imagem inteira de fundo de olho, outra usava a imagem focada no disco óptico e a última usava a imagem polarizada.

De um modo geral, os modelos foram implementados em Python a partir do framework Keras com Tensorflow como seu Backend. Abordagens de *transfer learning* foram comumente utilizadas para que se pudesse utilizar modelos robustos e conhecidos do meio de classificação de imagens, sendo que os pesos iniciais normalmente eram os pré-treinados do desafio ImageNet (DENG et al., 2009). Em diversos casos também, esses modelos tinham seus resultados fundidos em abordagens de *ensemble learning*.

3.4.2 Que tipo de métricas são utilizadas para medir a eficácia e acurácia dos modelos classificadores de glaucoma?

Para a segmentação, normalmente é utilizado o coeficiente DICE. Porém, acurácia e o índice Jaccard, também conhecido como Intersect over Union (IoU), também estão presentes. Para a classificação, as medidas mais comuns são especificidade, sensibilidade e Area Under the ROC Curve (AUC). Porém outras medidas também são utilizadas, como F1-score, acurácia e precisão. Abaixo, será brevemente explanado cada uma dessas métricas.

Acurácia mede a razão das predições corretas sobre o número total de instâncias avaliadas. É uma métrica simples e de fácil computação, porém possui limitações nos casos em que o conjunto de dados não é balanceado e por conta disso não é a melhor métrica a ser utilizada quando se espera selecionar ou determinar o classificador ótimo (HOSSIN; SULAIMAN, 2015).

Dice, também chamado de índice de sobreposição, é a métrica mais usada para validar a segmentação de volumes médicos (TAHA; HANBURY, 2015), sendo comumente utilizado como forma de validação estatística das anotações.

Equivalente ao coeficiente Dice definido acima, Jaccard, também conhecido como Intersect over Union (IoU) é definido como a intersecção entre dois conjuntos, dividido pela sua união.

A medida de precisão é usada para medir os padrões positivos que são corretamente previstos a partir do número total de previsões de uma classe positiva (HOSSIN; SULAIMAN, 2015).

A medida de sensibilidade mede a fração dos padrões positivos corretamente classificados (HOSSIN; SULAIMAN, 2015).

Ao contrário da sensibilidade, a medida de especificidade mede a fração dos padrões negativos corretamente classificados (HOSSIN; SULAIMAN, 2015).

Por fim, Area under the ROC Curve (AUC) reflete a performance geral de um classificador. Foi provada tanto empírica como teoricamente melhor do que a métrica de acurácia para avaliar a performance de classificadores (HOSSIN; SULAIMAN, 2015).

3.4.3 Quais são os conjuntos de dados utilizados para treinar e testar os modelos classificadores de glaucoma?

Diversos estudos utilizaram conjuntos de dados privados, elaborados pelos autores com auxílio de hospitais e clínicas de oftalmologia. Por outro lado, alguns autores optaram pelos conjuntos públicos disponíveis na rede. São eles HRF (BUDAI et al., 2013), RIM-ONE (FUMERO et al., 2011), ORIGA (ZHANG et al., 2010), ACRIMA (DIAZ-PINTO et al., 2019), DRISHTI (SIVASWAMY et al., 2014), REFUGE (ORLANDO et al., 2020) e LAG (LI et al., 2020).

Em todos os conjuntos, os dados são imagens de fundo de olho, elaborados especificamente para pesquisas relacionadas à patologias oculares, contendo as estruturas principais para o diagnóstico das mesmas. Esses conjuntos variam principalmente na resolução e na etnia das amostras. A maioria dos conjuntos possui um número reduzido de imagens, sendo alguns deles em proporções desbalanceadas de classes. Por exemplo, o conjunto HRF possui somente 15 amostras de olhos saudáveis, 15 amostras de olhos com retinopatia diabética e 15 amostras com glaucoma. ORIGA possui somente 750 amostras, RIM-ONE 485 e DRISHTI 101.

Afim de resolver tais problemas, os estudos empregaram técnicas de pré-processamento de dados para harmonizar e homogeneizar esses conjuntos, além de técnicas de aumento de dados e *transfer learning* para que assim seja possível utilizar modelos profundos e robustos que performam melhor com volumes grandes de dados.

3.4.4 Quais são as questões ainda em aberto quando se trata de modelos classificadores de glaucoma?

Em Xu et al. (2021) e Civit-Masot et al. (2020b), foi levantado a discussão de que os modelos atuais são "oráculos caixas pretas", que não explicam como chegaram nos seus resultados e nem com quanta confiança ou certeza eles possuem na sua predição.

Para isso, Xu et al. (2021) criou um modelo a partir do raciocínio clínico para realizar a classificação, com o foco em auxiliar os médicos no diagnóstico, fazendo assim com que cada decisão do modelo tenha uma explicação.

Já Civit-Masot et al. (2020b) focou em fazer um modelo embarcado que possa ser implan-

tado com facilidade nas clínicas e que possa auxiliar os médicos explicando a sua previsão e quantificando a sua confiança naquele resultado.

Os autores de Sreng et al. (2020) e Chai, Liu e Xu (2018) ressaltaram as características dos conjuntos de dados públicos, sendo todos eles imagens de alta qualidade, sem ruído nem outras patologias, focadas no disco óptico e na maioria das vezes a partir do mesmo ângulo.

Tirando uma ou outra característica específica de cada conjunto, essa homogeneidade está longe da realidade encontrada hoje nas clínicas, onde os pacientes possuem outras patologias além de glaucoma e os aparelhos utilizados para capturar a foto do fundo do olho variam.

Praticamente todos estudos comentam na dificuldade de se produzir um conjunto de dados robusto com anotações. Pois esse é um processo manual, tedioso e cansativo, que requer uma quantia considerável de tempo.

Para combater essa questão, Hemelings et al. (2020) utilizou uma abordagem de *Active Learning* para predizer quais imagens sem anotação o classificador se beneficiaria mais caso as mesmas fossem anotadas, reduzindo assim o custo de anotação em até 60 por cento.

O estudo de Chai et al. (2021) comenta que a maioria dos pesquisadores focam somente em um tipo de dados, as imagens de fundo de olho, porém na clínica o médico possui à sua disposição diferentes tipos de informações qualitativas e quantitativas, como por exemplo a ficha médica, a descrição do paciente e da sua condição. Além disso, também focaram em explicitar a confiança do seu modelo na predição, e comentaram a possibilidade de explorarem, no futuro, *blockchain* para implementar um modelo capaz de diagnosticar outras doenças oculares além de glaucoma.

No trabalho de Phasuk et al. (2019), ressaltou-se a importância da região do disco óptico, demonstrando como a relação escavação/disco calcula uma das mais importantes características de glaucoma, o aumento do copo, ou escavação. Isso foi possível a partir da comparação de performance do modelo que usou a imagem cortada no disco óptico e o modelo que utilizou a imagem inteira.

Por fim, Kim et al. (2019) levantou a questão de como incorporar diversos tipos de imagens de fundo de olho, por exemplo aquelas obtidas por equipamentos diferentes e com ângulos diferentes, para assim aumentar a diversidade e reduzir o viés e o sobre-ajuste, já que as características do olho variam de acordo com a etnia dos pacientes.

3.4.5 Quais são os desafios encontrados na modelagem de algoritmos classificadores de glaucoma?

Os autores de Hemelings et al. (2020) apontam que a distribuição das classes dos conjuntos de dados privados e coletados em hospitais está fora da realidade (no casos deles, 70 por cento das imagens eram glaucoma). Isso se dá devido à falta de disponibilidade de imagens de olhos normais, sem glaucoma, pois os hospitais não as guardam.

Além disso, devido à natureza insidiosa da doença, a maioria das imagens eram de glaucoma

moderado ou avançado, sendo de crucial importância a capacidade de treinar e avaliar o modelo computacional em cima de imagens de glaucoma leve ou inicial. Por fim, os modelos treinados e validados nesse estudo usaram majoritariamente imagens de pessoas caucasianas, capturadas a partir de um único modelo de câmera. Para resolver essa questão eles estão estendendo o seu estudo para validar e refinar o modelo atual usando dados mais heterogêneos, obtidos a partir de uma colaboração internacional, com o objetivo de desenvolver um modelo capaz de realizar exames em escala global.

Ainda nessa linha dos conjuntos de dados, Sreng et al. (2020) comenta sobre a dificuldade de generalização devido aos conjuntos de dados públicos utilizarem câmeras diferentes para capturar amostras de pacientes de etnias diferentes. Explica que cada etnia tem a sua pigmentação específica e que isso varia a imagem de fundo de olho. Além disso, todas as imagens são de alta resolução, que não é a realidade das clínicas.

Os autores de Civit-Masot et al. (2020b) comentam a expectativa dos médicos na segmentação do disco e do copo óptico, onde esperam ver uma forma elíptica que nem sempre é aquela segmentada pelos algoritmos. Para tentar resolver isso, utilizaram RANSAC. Ainda nessa linha de transmitir confiança, descrevem o seu foco em explicabilidade ao implementarem um relatório final para complementar a classificação e assim ajudar o médico a tomar a decisão de diagnóstico final.

Já em Yu et al. (2019), comenta-se como a qualidade da imagem e a presença de outras patologias impactam no desempenho do sistema. Para melhorar a acurácia da segmentação é necessário treiná-lo e aperfeiçoá-lo em diferentes conjuntos de dados com qualidades diferentes.

3.4.6 Quais são os modelos, utilizados atualmente, que apresentam a maior acurácia e eficácia?

O trabalho de Sallam et al. (2021) reuniu 9 dos modelos mais bem sucedidos em desafios de classificação de imagens, com seus pesos pré-treinados, adaptou os mesmos para a classificação de glaucoma e comparou seus resultados em um único conjunto de dados, o Large-scale Attention based Glaucoma (LAG) dataset. Os modelos escolhidos foram AlexNet, VGG11, VGG16, VGG19, GoogleNet (Inception V1), ResNET-18, ResNET- 50, ResNET-101 e ResNet-152. Dentre eles, ResNet-152 se destacou, com acurácia, precisão e recall de 86.9.

Em Serte e Serener (2019), comparou-se o desempenho de quatro redes conhecidas da área de classificação de imagem em cada um dos cinco conjuntos de dados publicamente disponíveis para a tarefa de classificação de glaucoma. Enquanto quem um dos conjuntos de dados era escolhido para teste, os outros quatro eram utilizados para treinar os modelos. Assim, compararam seus desempenhos para cada um dos conjuntos. As redes foram Xception, ResNet-50, GoogLeNet e ResNet-152. Os conjuntos de dados foram HRF, Drishti-GS1, RIM-ONE, sjchoi86-HRF e ACRIMA. De acordo com o estudo, cada uma das redes se destacou em um conjunto específico.

Em Civit-Masot et al. (2020b), os autores combinaram os resultados de dois sub-sistemas,

um de segmentação composto por duas redes U-net, uma para o disco e outra para o copo, e outro de classificação baseado em MobileNetV2. Comparando com alguns modelos comumente utilizados como VGG16, ResNet50 e Xception, eles obtiveram performance e acurácia comparável, AUC 0.96, acurácia 0.88, porém com custo computacional até vinte vezes menor. O conjunto de dados nesse caso foi a união de RIM-ONE e Drishti com técnicas de aumento de dados, que resultaram em 9360 imagens.

A partir da integração de RNCs e Multi-layer Perceptrons (MLPs) Bayesianas, em um modelo chamado de Bayesian Deep Multisource Learning model (BDMSL), Chai et al. (2021) classificou glaucoma a partir de múltiplas fontes de informação, como imagens da retina, indicadores médicos e indicadores de reclamações dos pacientes. Para avaliar o desempenho do seu modelo, compararam o mesmo com diversas abordagens tradicionais de visão computacional e modelos de aprendizado profundo.

O apêndice B relaciona as técnicas utilizadas e as métricas obtidas.

3.5 Discussão

É possível classificar imagens de fundo de olho como normais ou glaucomatosas de forma automatizada a partir de modelos de aprendizado profundo.

Além de avaliar a viabilidade e obter resultados cada vez melhores, estudos estão sendo feitos com foco na clínica, onde questões além de acurácia são de suma importância, como a explicabilidade e transparência dos resultados obtidos pelos modelos.

Os modelos mais performáticos são também aqueles modelos que necessitam de um volume grande de dados para que possam extrair características e padrões relevantes. Infelizmente no meio médico tais informações são de difícil obtenção, havendo um esforço considerável da comunidade em modelar sistemas capazes de trabalhar com quantidades pequenas de dados sem perder em performance.

Os conjuntos de dados disponíveis publicamente são escassos e homogêneos, sendo necessário técnicas de aumento de dados para que se possa obter uma quantidade expressiva de informações. Tal homogeneidade está longe da realidade das clínicas, que possuem imagens de qualidades variadas com a possibilidade de haver mais de uma patologia envolvida.

Além disso, nas clínicas os médicos normalmente possuem à sua disposição informações das mais variadas fontes e aparelhos, como por exemplo a pressão intraocular, a ficha médica do paciente com suas informações básicas e seus comentários sobre o seu estado de saúde atual.

4 METODOLOGIA

A partir da análise feita em cima dos trabalhos relacionados, foi selecionado um dos estudos como o modelo de base, sendo a partir dele que este artigo fará as suas contribuições. O trabalho selecionado foi o de (CIVIT-MASOT et al., 2020b), e o seu modelo consiste em uma ferramenta de auxílio diagnóstico de glaucoma, e é composto por dois sub-sistemas classificadores. Este trabalho se destacou por ser um sistema moderno, que utiliza técnicas de aprendizado profundo, sendo que o seu objetivo final é elaborar uma ferramenta de auxílio ao diagnóstico médico de glaucoma, criticando assim os trabalhos que desenvolvem um sistema "oráculo caixa-preta", que apenas retornam uma classificação sem justificativa, e consequentemente sem a confiança dos especialistas.

O primeiro sub-sistema é composto por duas redes de segmentação baseados em uma arquitetura U-net (RONNEBERGER; FISCHER; BROX, 2015) modificada, responsáveis por localizar o disco e o copo do olho.

Com a segmentação das estruturas do copo e do disco, o cálculo dos seus diâmetros é realizado para assim computar a relação escavação/disco (do inglês Cup-to-disc Ratio, ou CDR), classificando como glaucoma resultados acima de um determinado limite. Além deste cálculo, o sistema também verifica se a elipse anotada como o contorno do disco e do copo está de acordo com aquela esperada pelos usuários, aplicando Random Sample Consensus (RANSAC) (CANTZLER, 1981) para melhorar os resultados. Este algoritmo é capaz de interpretar e suavizar dados que possuem uma porcentagem significativa de erros.

O segundo sub-sistema é composto por uma rede de classificação direta baseada na arquitetura de MobileNetV2 (SANDLER et al., 2018), que recebe a imagem como entrada e tem como saída a classificação final de glaucoma ou normal.

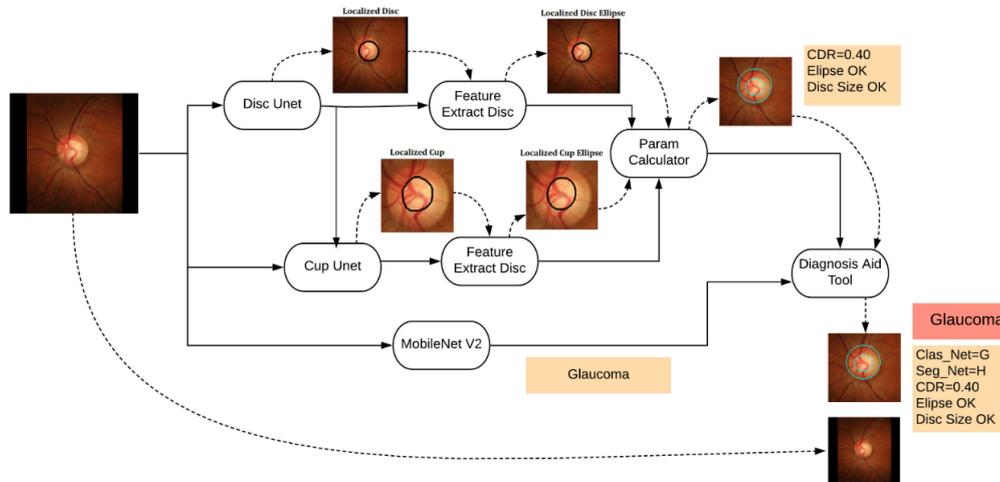
Por fim, a ferramenta faz um *ensemble*, ou agrupamento dos resultados de ambos sub-sistemas, classificando como glaucoma se pelo menos um deles assim o classificou. Além disso, apresenta a anotação de ambas as segmentações, o valor do CDR e se a elipse encontrada estava ok. A figura 4 apresenta esta arquitetura do modelo base.

O modelo proposto no presente trabalho modifica o modelo de base nos seus dois sub-sistemas de classificação.

No sistema de segmentação, foi feita a união das duas redes, fazendo com que a mesma seja capaz de segmentar tanto o disco como o copo presente na imagem. Para que isso fosse possível, foram realizadas modificações na preparação dos dados e nos hiper-parâmetros de treinamento, detalhados nas seções 4.2 e 4.4.

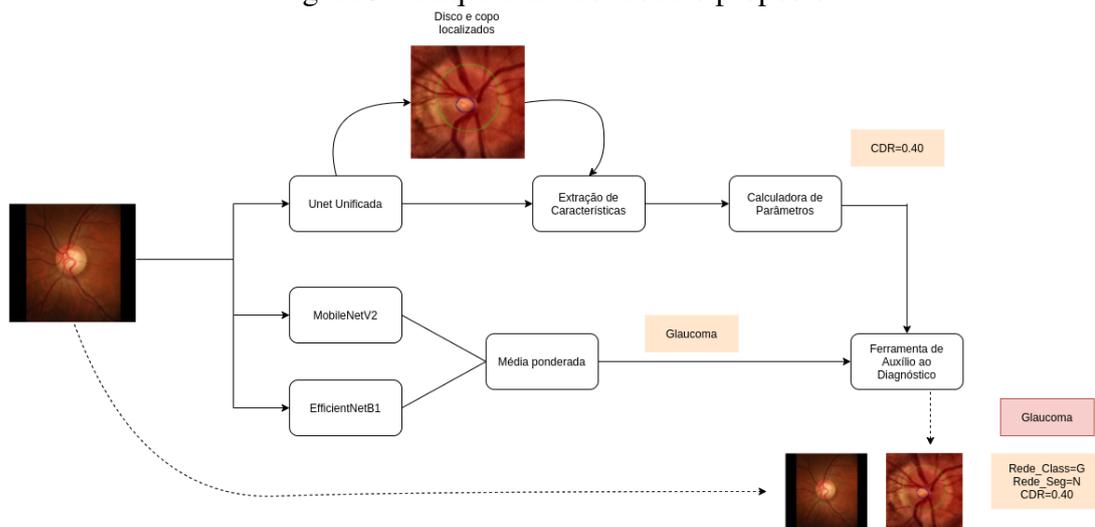
No sistema de classificação direta, foi aplicada uma técnica de agrupamento de algoritmos, adicionando uma nova rede de classificação direta, a EfficientNetB1 (TAN; LE, 2019). Os dados de entrada são repassados tanto para a rede original MobileNetV2 quanto para a nova rede EfficientNetB1, sendo que a última camada retorna uma classificação binária baseada na média da saída de ambas as redes. A Figura 5 detalha a arquitetura do modelo proposto.

Figura 4 – Arquitetura do modelo base.



Fonte: (CIVIT-MASOT et al., 2020b)

Figura 5 – Arquitetura do modelo proposto.



Fonte: Autor

A realização do projeto consistiu em três atividades macro: A replicação do modelo base; a implementação da contribuição 1 (agrupamento dos classificadores); e a implementação da contribuição 2 (união das redes de segmentação).

Para replicar o modelo base, foi necessário as seguintes etapas: 1) Coletar e analisar o conjunto de dados, 2) Preparar os dados, 3) Elaborar a arquitetura das redes neurais, 4) Treinar as redes e, por fim, 5) Avaliar seus resultados. A seguir será detalhado cada uma dessas etapas.

Para a implementação foi utilizado o ambiente Google Colab Pro, com as bibliotecas TensorFlow e Keras, o treinamento foi feito com GPUs habilitadas.

A figura 6 apresenta o resultado final da ferramenta de diagnóstico do modelo base, replicada pelo autor deste artigo.

Figura 6 – Saída final do modelo base.



Fonte: Autor

4.1 Conjunto de Dados

Enquanto que no trabalho base se fez necessário manipular dois conjuntos de dados, devido à escassez de amostras, no presente estudo não se fez necessário tal prática. Os autores utilizaram 149 amostras do conjunto RIM-ONE e 101 amostras do conjunto Drishti, disponíveis na época, sendo que ao replicar o modelo base, estavam disponíveis 485 amostras somente no conjunto RIM-ONE. Afim de homogeneizar as amostras e economizar com pré-processamento, foi adotado somente este conjunto de dados.

O conjunto de dados RIM-ONE, elaborado por Fumero et al. (2011), consistiu originalmente em 169 imagens de fundo de olho, recortadas na área de interesse da cabeça do nervo óptico, e faz parte de um projeto de pesquisa desenvolvido em colaboração com três hospitais espanhóis: Hospital Universitario de Canarias [12], Hospital Clínico San Carlos [13] e Hospital Universitario Miguel Servet [14].

O propósito principal do trabalho é oferecer um conjunto de dados de imagens de referência oftalmológica, especificamente desenvolvidas para o diagnóstico de glaucoma. As anotações foram feitas por 5 especialistas na área, com uma segmentação de referência elaborada a partir delas.

O apêndice E apresenta exemplos de imagens do conjunto de dados, sendo que a imagem da esquerda é de um olho normal, e a da direita um olho com glaucoma severo.

Com o passar do tempo, versões novas foram sendo liberadas, com mais imagens e algumas revisões. Em uma recente versão, os autores elaboraram um conjunto de dados modificado e específico para projetos de aprendizado profundo, o RIM-ONE DL (BATISTA et al., 2020), onde além das imagens de fundo de olho, também estão presentes máscaras de segmentação binárias do disco e do copo do olho. O apêndice E apresenta as máscaras binárias do copo e do disco correspondentes à uma imagem deste dataset.

Atualmente, o conjunto de dados RIM-ONE DL consiste em 485 imagens de fundo de olho, com suas correspondentes segmentações binárias do disco e do copo. Dentre estas 485 amostras, 313 (65%) são anotadas como pacientes normais e 172 (35%) são anotadas como pacientes com

glaucoma. Este desbalanceamento entre as classes influencia nas escolhas de hiper-parâmetros, como a função de perda, o otimizador e as métricas avaliadas.

O conjunto de dados foi dividido em conjuntos de treinamento e de teste, com duas variações, uma particionada aleatoriamente, com estes dois sub conjuntos compostos de imagens de todo o dataset, e uma outra variação particionada por hospital, onde as imagens do Hospital Universitario de Canarias foram usadas para treinamento, enquanto que as imagens restantes dos dois hospitais foram usadas para teste.

4.2 Preparação dos Dados

Seguindo as descrições de pré-processamento e aumento de dados presente no artigo base, os dados foram preparados da seguinte forma.

Primeiro, as imagens e suas respectivas máscaras foram concatenadas em uma lista, em seguida foi transformado esta lista em uma estrutura de dados *tf.dataset*. Com isso, funções de pré-processamento da biblioteca TensorFlow puderam ser aplicadas no conjunto de dados.

O conjunto de dados foi dividido em um sub-conjunto de treino e outro de validação, para este último reservado 25 por cento das imagens, separados de forma aleatória.

As imagens possuíam dimensões variadas, e por isso foram redimensionadas para 224x224 no sub-sistema de classificação e 128x128 no sub-sistema de segmentação. Além disso, as imagens foram normalizadas para o intervalo $[0, 1]$, facilitando assim o processamento da rede neural.

Além da dimensão, a qualidade das imagens também variam, e, para resolver aqueles casos em que as imagens possuíam uma qualidade muito baixa, foi aplicado equalização adaptativa do histograma de contraste limitado (do inglês, contrast limited adaptive histogram equalization, ou CLAHE), assim realçando as estruturas morfológicas presentes na imagem.

O número de amostras do RIM-ONE DL não é suficiente para treinar um robusto modelo de aprendizado de máquina. Para resolver isso, técnicas de aumento de dados foram aplicadas. O processo consistiu em aplicação de um filtro de brilho randômico com delta máximo de até 0.3, ou contraste randomizado com fator no intervalo $[0.2, 0.4]$, seguido de uma rotação leve de até 15 graus. Com isso, o número de amostras foi de 485 para 9215.

Para a implementação da rede de segmentação unificada, foi necessário unir também as máscaras do disco e do copo de cada imagem. Nesta nova máscara, cada uma das máscaras corresponde a um canal de cor do sistema RGB, sendo a máscara do disco correspondendo ao canal vermelho e a máscara do copo ao canal verde. O apêndice C demonstra visualmente este processo.

4.3 Arquiteturas das redes neurais

A ferramenta de diagnóstico do estudo base consiste em três redes neurais, duas redes U-net generalizadas para segmentação, uma do disco e outra do copo, e uma rede de classificação direta baseada em MobileNetV2. O modelo proposto neste estudo consiste em duas redes neurais, uma rede U-net generalizada para a segmentação tanto do disco como do copo, e uma rede de classificação direta composta pelo agrupamento da rede baseada em MobileNetV2, com uma nova rede baseada em EfficientNetB1. Todas essas redes serão descritas a seguir.

4.3.1 Redes de Segmentação: U-net

Para segmentar o disco e o copo, o modelo base utilizou uma arquitetura de U-net generalizada (CIVIT-MASOT et al., 2020a). Esta rede possui 6 níveis de codificação e decodificação, com 64 canais no primeiro estágio de codificação e uma taxa de incremento de canais por camada de 1.1. Mesmo com um nível a mais que a U-net original, a redução da taxa de incremento de 2 para 1.1 diminuiu o número de parâmetros de 138M para 2.5M (CIVIT-MASOT et al., 2020b). O anexo A demonstra o diagrama da arquitetura proposta no artigo base e replicada neste estudo.

A rede de segmentação unificada proposta neste estudo seguirá a mesma arquitetura de U-net descrita acima, substituindo somente o otimizador de Adam para RMSprop. Devido ao aumento de complexidade, Será incrementado o número de épocas, de 100 para 300.

4.3.2 Redes de Classificação: MobileNetV2 e EfficientNetB1

Para o sub-sistema de classificação direta, uma rede baseada em MobileNetV2 foi elaborada pelos autores do modelo base. As camadas superiores desta arquitetura foram removidas, sendo adicionado um classificador final baseado em uma camada de *average polling* cuja saída é repassada para uma camada densa de 64 nós, um estágio de *dropout* e uma camada final densa de 2 nós para distinguir entre as duas classes (CIVIT-MASOT et al., 2020b). Este sistema pode ser visto no anexo B.

O sub-sistema de classificação proposto neste estudo consiste em um agrupamento (ensemble learning) do modelo base MobileNetV2 com um novo modelo baseado em EfficientNetB1, seguindo a mesma estrutura de camadas finais. Cada um dos modelos fará a sua predição e em seguida uma camada de média ponderada será responsável por agrupar os resultados em uma classificação final. O apêndice D exemplifica o sistema proposto.

4.4 Treinamento das redes

Para treinar as redes neurais, diversos hiper-parâmetros devem ser definidos. Aqueles que foram descrito no trabalho base foram utilizados, nos casos em que não foi explicitado, a escolha dos hiper-parâmetros foi feita pelo autor deste artigo. A seguir será descrito como as redes foram e serão treinadas.

4.4.1 Redes de Segmentação: U-net

No treinamento das redes de segmentação foi empregado o otimizador Adam com taxa de aprendizado adaptativo entre $1e-3$ (0.001) e $2e-4$ (0.0002), função de custo DICE e treinado por 100 épocas com o tamanho do batch de 120 amostras. A última camada consistiu em uma densa com ativação sigmoide, e as métricas avaliadas foram IoU e DICE.

A rede de segmentação unificada proposta utilizará os mesmo hiper-parâmetros do modelo base, porém substituindo o otimizador Adam pelo RMSprop, e treinando por 300 épocas ao invés de 100.

4.4.2 Redes de Classificação: MobileNetV2 e Rede Agrupada

Para o treinamento da rede de classificação MobileNetV2 foi utilizado os pesos pré-treinados do desafio ImageNet, com todas as camadas iniciais congeladas. O otimizador empregado foi o RMSprop com taxa de aprendizado de $1e-03$, ou 0.001. A função de custo adotada foi a Entropia Cruzada (Binary Cross Entropy em inglês).

A rede foi treinada primeiramente por 50 épocas, como descrito em Civit-Masot et al. (2020b), porém não atingiu as métricas esperadas. Por conta disso, foi elevado o número de épocas para 100. O tamanho do batch foi de 64 amostras.

As métricas avaliadas foram AUC, acurácia, especificidade e sensibilidade.

O classificador com o agrupamento das redes, proposto por este estudo, foi treinado com os mesmos hiper-parâmetros descritos acima.

4.5 Discussão

A rede elaborada pelo artigo base alcança resultados similares em comparação com os trabalhos relacionados, porém com uma carga de processamento reduzida, sendo o principal responsável por isso a rede MobileNetV2.

A segmentação do copo e do disco feita por redes distintas facilitou na adoção dos conjuntos de dados, pois possuem máscaras distintas para cada estrutura, além da modelagem da arquitetura das redes, sendo mais simples elaborar uma rede de segmentação binária do que uma multi-classe. O problema desta abordagem é o aumento da carga de armazenamento e proces-

samento, sendo necessário carregar duas redes com a mesma arquitetura porém com os pesos diferentes.

O foco dos autores é garantir a confiança dos especialistas na ferramenta, além de mantê-la leve o suficiente para ser implementada em um ambiente embarcado. Ambas as propostas apresentadas neste estudo levam em conta esses objetivos, tendo como impacto esperado o aumento das métricas avaliadas e a otimização do sistema no que tange requerimentos de armazenamento e processamento.

5 EXPERIMENTOS E RESULTADOS

Aqui serão descritos os experimentos realizados e os resultados obtidos em cada etapa do processo de implementação, começando pelo desenvolvimento do modelo base, seguindo para a adição da segunda rede de classificação e seu posterior agrupamento com a rede original, e por fim a união das redes de segmentação.

5.1 Replicando o Modelo Base

Com o intuito de aperfeiçoar a ferramenta de diagnóstico apresentada por Civit-Masot et al. (2020b), foram propostas duas contribuições principais, a adição de uma segunda rede de classificação, a EfficientNetB1 (TAN; LE, 2019), e a união das duas redes de segmentação, do copo e do disco. O objetivo dessas modificações é aumentar a acurácia e a sensibilidade das redes, sendo essa última um dos focos dos autores do artigo base, pois, de acordo com eles, falsos negativos são os resultados mais críticos e portanto devem ser minimizados.

Em um primeiro momento, foi replicado o modelo base, de Civit-Masot et al. (2020b). A tabela 1 compara as métricas obtidas pelos autores do artigo base e desta replicação, no que tange ao coeficiente DICE das redes de segmentação. A tabela 2 compara as métricas obtidas pelos autores do artigo base e desta replicação, no que tange à AUC, acurácia, sensibilidade e especificidade da rede de classificação.

Tabela 1: Comparação da segmentação do trabalho base e de sua réplica

Métrica	Trabalho Base	Trabalho Replicado
DICE do Disco	0.92	0.99
DICE do Copo	0.84	0.94

Tabela 2: Comparação da classificação do trabalho base e de sua réplica

Métrica	Trabalho Base	Trabalho Replicado
AUC	0.93	0.95
Acurácia	0.88	0.88
Especificidade	0.86	0.84
Sensibilidade	0.91	0.91

5.2 Adicionando uma Segunda Rede de Classificação

Afim de melhorar ainda mais as predições da rede de classificação, foi proposto adicionar uma segunda rede, igualmente leve e com o mesmo propósito de ser aplicável em um sistema embarcado. Com isso, foi escolhido a EfficientNetB1 (TAN; LE, 2019), composta de pouco menos que 8 milhões de parâmetros, é considerado um modelo leve e capaz de prover resultados tão eficazes quanto aqueles das arquiteturas mais robustas, como VGG ou Inception, por exemplo.

Foi treinado com os mesmos hiper-parâmetros que o classificador do artigo base, e seus

resultados se mostraram promissores, pois mesmo não aumentando a AUC, aumentou a sensibilidade de 0.9140 para 0.9462. A tabela 3 compara os resultados.

Tabela 3: Métricas MobileNetV2, EfficientNetB1 e do seu agrupamento

Métrica	MobileNetV2	EfficientNetB1	Modelo Agrupado
AUC	0.9567	0.9394	0.9512
Acurácia	0.8897	0.8345	0.8897
Especificidade	0.8461	0.6538	0.7884
Sensitividade	0.9140	0.9355	0.9462

5.3 Unificando as Redes de Segmentação

Com o propósito de simplificar a arquitetura e torná-la mais leve para uma possível implantação embarcada, foi desenvolvido uma única rede de segmentação, capaz de segmentar tanto o disco como o copo. Devido ao aumento de complexidade, foi necessário treinar por mais épocas para chegar no mesmo patamar das redes individualizadas, indo de 100 épocas para 300. Mesmo assim, ficou evidente que uma rede unificada é capaz de alcançar os mesmos resultados que as redes de segmentação binárias, porém de uma forma mais simples e leve, de suma importância para a implantação do modelo em um sistema embarcada. A tabela 4 compara os resultados.

Tabela 4: Métricas da segmentação do disco, copo e disco+copo

Métrica	Rede Disco	Rede Copo	Rede Unificada
IoU	0.9872	0.8892	0.9455
DICE	0.9936	0.9406	0.9502

5.4 Discussão

O foco dos autores do artigo base são expressados como auxiliar o médico no diagnóstico de glaucoma, para tanto, empenham-se em desenvolver um modelo leve e eficiente, com o menor número de falsos negativos possíveis e que possa ser executado em sistemas embarcados, capazes de serem implantados na clínica médica. Levando isso em consideração, é possível concluir que ambas as modificações desenvolvidas neste presente estudo contribuem para o objetivo final do artigo e do modelo apresentado.

A adição de uma segunda rede de classificação obteve êxito em *melhorar a sensibilidade do modelo final* sem acarretar em nenhum custo adicional de complexidade ou computação, assim reduzindo o número de possíveis falsos negativos, isso é, pacientes classificados como saudáveis porém com glaucoma. A união das redes de segmentação não conseguiu melhorar os coeficientes das redes individuais, porém simplificou a arquitetura e a quantidade de parâmetros a serem treinados e carregados, facilitando assim na implantação em dispositivos de baixa capacidade de armazenamento e processamento.

Comparando com o trabalho base e os trabalhos relacionados, é possível concluir que o modelo proposto alcançou desempenho equiparável, porém de modo mais simples e leve. Para

isso, porém, algumas ressalvas são necessárias. Primeiro, os trabalhos relacionados utilizaram conjuntos de dados distintos, alguns inclusive utilizaram dados além de imagens de fundo de olho. Segundo, a comparação só seria fiel se os trabalhos relacionados fossem replicados, treinados no mesmo conjunto de dados, neste caso RIM-ONE DL, e avaliados posteriormente.

6 CONCLUSÃO

Glaucoma é uma doença silenciosa que só demonstra sinais em estados avançados e irreversíveis da doença. Seu diagnóstico consiste em medições da pressão intraocular e principalmente exames de fundo de olho, onde o especialista analisa as estruturas oculares como o disco óptico e sua escavação, conhecida como o copo do olho.

Diversos exames e testes são utilizados para avaliar glaucoma, inclusive com abordagens automatizadas de aprendizado de máquina. O principal destes exames é a fundoscopia, sendo este um processo manual, tedioso e sujeito à falhas, sendo necessário alguns anos de expertise e até especialização por parte dos médicos, sem contar com possíveis falhas ou a precariedade dos equipamentos utilizados para análise.

Nos casos em que não há médicos especializados ou equipamentos adequados, soluções automatizadas são muito bem vindas e podem auxiliar e até diagnosticar sem a interferência humana. Técnicas de visão computacional e aprendizado de máquina já estão sendo empregadas em larga escala na medicina, e para o diagnóstico de glaucoma não é diferente.

Abordagens clássicas como SVM são utilizadas em conjunto com técnicas de visão computacional para dar o seu diagnóstico. Aprendizado profundo também é utilizado para analisar imagens de fundo de olho e de tomografia óptica, classificando glaucoma a partir das características morfológicas visualmente presentes.

Esses sistemas muitas vezes pecam em transparência e justificativa, sendo considerados "oráculos caixa preta", sendo assim um grande foco da comunidade científica os sistemas explicativos, que apresentam além de uma classificação, ou diagnóstico, o por que daquele resultado e o quão confiante o algoritmo está dele.

Todas estas considerações foram levadas em conta ao escolher o estudo base, sendo ele uma ferramenta de auxílio ao oftalmo que utiliza aprendizado profundo e imagens de fundo de olho para classificar diretamente e segmentar a imagem, providenciando informações robustas para embasar o seu diagnóstico final, além de utilizar módulos leves o suficiente para serem instalados em sistemas embarcados de baixo poder de armazenamento e processamento.

As contribuições deste presente estudo foram a adição de uma segunda rede de classificação e a união das redes de segmentação. A primeira contribuição foi capaz de *aumentar a sensibilidade*, reduzindo ainda mais o número de falsos negativos e assim garantindo ainda mais valor para seu uso na clínica. A segunda contribuição fez com que a arquitetura fosse simplificada, ocupando menos espaço de armazenamento e utilizando menos recursos de processamento, sem perder acurácia, tendo como única consequência a necessidade de mais tempo de treinamento inicial. Neste sentido, a principal novidade em relação à literatura se encontra na validação do modelo apresentado no trabalho base, e sua posterior otimização a partir das duas contribuições.

Mesmo fornecendo informações valiosas para o usuário, o sistema ainda não pode ser considerado um sistema de inteligência artificial explicável, pois falta dar as porcentagens de confiança na classificação, além de ressaltar quais áreas da imagem influenciaram mais na classi-

ficação, em especial daquela realizada pelo sub-sistema de classificação direta. Mesmo assim, há um esforço considerável em aumentar a confiança do especialista no modelo, apresentando para o mesmo a imagem original, as estruturas do disco e do copo localizadas, a relação escavação/disco calculada e o resultado de cada uma das redes.

Este esforço se traduz na dificuldade de aceitação dos especialistas em sistemas de aprendizado de máquina. De acordo com Civit-Masot et al. (2020b), sistemas de aprendizado profundo podem ser classificados como "oráculos caixas pretas", devido à sua capacidade de classificar e diagnosticar patologias sem se preocupar em explicar a lógica da solução encontrada. Com isso, o objetivo dos sistemas são substituir o especialista pelo sistema automatizado, dificultando ainda mais a adoção do sistema na clínica.

Neste sentido a explicabilidade é crucial para uma futura implantação do sistema na clínica, para isso, será necessário evoluir o sistema em uma aplicação independente em ambiente de produção, com ciclos de aprendizado e de aplicação direta, sendo inserida no fluxo do processo a partir de um sistema embarcado que recebe a imagem diretamente do sistema de captura, ou que recebe a imagem de fundo de olho posteriormente a partir de uma API presente no computador da clínica.

Com isso, o especialista pode ter uma sugestão de diagnóstico de forma clara e explicativa, sem o intuito de substituí-lo e, conseqüentemente, sem gerar desconfiança.

Isso requer um avanço e uma futura colaboração entre os pesquisadores e as instituições médicas, sendo um dos principais trabalhos futuros.

Sendo o objetivo principal do estudo auxiliar o diagnóstico clínico, é de suma importância levar em conta a realidade dos pacientes. Isso inclui outras patologias além de glaucoma, etnias variadas com características oculares específicas, além de imagens com resoluções e especificações diferentes.

Trabalhos futuros são diversos e incluem: a adição de outros conjuntos de dados, públicos e principalmente privados, para assim garantir uma maior heterogeneidade e um contato mais próximo das clínicas reais, onde não é utilizado sempre o mesmo tipo de equipamento responsável pela aquisição das imagens digitais de fundo de olho; a otimização e preparação do sistema para que o mesmo possa ser implantado em um sistema embarcado e assim testar sua atuação prática na clínica; a expansão das capacidades do sistema de categorizar outras doenças oculares, inclusive mais de uma patologia em uma mesma amostra, sendo este um dos casos da maioria dos pacientes com doenças oculares e, por fim, adotar aquelas técnicas de sistemas de aprendizado profundo que o tornam totalmente explicativos e assim aumentar tanto a confiança do especialista na tecnologia, como também o seu apoio na adoção e desenvolvimento da ferramenta para a clínica.

Referências

- ABRAHAM, A. Artificial neural networks. **Handbook of measuring system design**, John Wiley and Sons, Ltd Chichester, UK, 2005.
- AHN, J. M. et al. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. **PLoS ONE**, v. 13, 2018.
- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. **Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017**, v. 2018-January, 2018.
- BARROS, D. M. et al. Machine learning applied to retinal image processing for glaucoma detection: Review and perspective. **BioMedical Engineering Online**, v. 19, n. 1, 2020.
- BATISTA, F. J. F. et al. Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. **Image Analysis & Stereology**, v. 39, n. 3, p. 161–167, 2020.
- BUDAI, A. et al. Robust vessel segmentation in fundus images. **International Journal of Biomedical Imaging**, v. 2013, 2013.
- CABITZA, F.; RASOINI, R.; GENSINI, G. F. Unintended consequences of machine learning in medicine. **JAMA - Journal of the American Medical Association**, v. 318, n. 6, 2017.
- CANTZLER, H. Random sample consensus (ransac). **Institute for Perception, Action and Behaviour, Division of Informatics, University of Edinburgh**, Citeseer, 1981.
- CHAI, Y. et al. Glaucoma diagnosis in the chinese context: An uncertainty information-centric bayesian deep learning model. **Information Processing and Management**, v. 58, 2021.
- CHAI, Y.; LIU, H.; XU, J. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. **Knowledge-Based Systems**, v. 161, 2018.
- CIVIT-MASOT, J. et al. Multidataset incremental training for optic disc segmentation. In: SPRINGER. **International Conference on Engineering Applications of Neural Networks**. [S.l.], 2020. p. 365–376.
- CIVIT-MASOT, J. et al. Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction. **IEEE Access**, v. 8, 2020.
- DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. **2009 IEEE conference on computer vision and pattern recognition**. [S.l.], 2009. p. 248–255.
- DEY, A.; BANDYOPADHYAY, S. K. Automated glaucoma detection using support vector machine classification method. **Journal of Advances in Medicine and Medical Research**, p. 1–12, 2016.
- DIAZ-PINTO, A. et al. Cnns for automatic glaucoma assessment using fundus images: An extensive validation. **BioMedical Engineering Online**, v. 18, 2019.
- FUMERO, F. et al. Rim-one: An open retinal image database for optic nerve evaluation. **Proceedings - IEEE Symposium on Computer-Based Medical Systems**, 2011.

GÓMEZ-VALVERDE, J. J. et al. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. **Biomedical Optics Express**, v. 10, 2019.

GUPTA, N. et al. Ico guidelines for glaucoma eye care. **International Council of Ophthalmology**, v. 32, 2015.

HEMELINGS, R. et al. Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. **Acta Ophthalmologica**, v. 98, 2020.

HOSSIN, M.; SULAIMAN, M. N. A review on evaluation metrics for data classification evaluations. **International Journal of Data Mining and Knowledge Management Process**, v. 5, 2015.

HUANG, M.-L.; CHEN, H.-Y. Development and comparison of automated classifiers for glaucoma diagnosis using stratus optical coherence tomography. **Investigative Ophthalmology and Visual Science**, v. 46, n. 11, 2005.

KIM, M. et al. Medinoid: Computer-aided diagnosis and localization of glaucoma using deep learning. **Applied Sciences (Switzerland)**, v. 9, 2019.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. 2007.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, 2015.

LI, L. et al. A large-scale database and a cnn model for attention-based glaucoma detection. **IEEE Transactions on Medical Imaging**, v. 39, 2020.

MAETSCHKE, S. et al. A feature agnostic approach for glaucoma detection in oct volumes. **PLoS ONE**, v. 14, n. 7, 2019.

MANTRAVADI, A.; VADHAR, N. Glaucoma. **Primary Care - Clinics in Office Practice**, v. 42, n. 3, 2015.

NICE. Glaucoma: diagnosis and management. **NICE guideline**, v. 81, n. 3, 2017.

ORLANDO, J. I. et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. **Medical Image Analysis**, v. 59, 2020.

O'SHEA, K.; NASH, R. **An Introduction to Convolutional Neural Networks**. 2015.

PHASUK, S. et al. Automated glaucoma screening from retinal fundus image using deep learning. **Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS**, 2019.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. [S.l.], 2015. p. 234–241.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence — A Modern Approach**. Nova Jersey, EUA: Pearson Education, 2010. v. 3 ed.

- SALLAM, A. et al. Early detection of glaucoma using transfer learning from pre-trained cnn models. **2021 International Conference of Technology, Science and Administration (ICTSA)**, 2021.
- SANDLER, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2018. p. 4510–4520.
- SERTE, S.; SERENER, A. A generalized deep learning model for glaucoma detection. **3rd International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2019 - Proceedings**, 2019.
- SIVASWAMY, J. et al. Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation. **2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014**, 2014.
- SRENG, S. et al. Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. **Applied Sciences (Switzerland)**, v. 10, 2020.
- TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: Analysis, selection, and tool. **BMC Medical Imaging**, v. 15, 2015.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. **36th International Conference on Machine Learning, ICML 2019**, v. 2019-June, 2019.
- THAM, Y.-C. et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. **Ophthalmology**, v. 121, 2014.
- XU, Y. et al. A hierarchical deep learning approach with transparency and interpretability based on small samples for glaucoma diagnosis. **npj Digital Medicine**, v. 4, 2021.
- YU, S. et al. Robust optic disc and cup segmentation with deep learning for glaucoma detection. **Computerized Medical Imaging and Graphics**, v. 74, 2019.
- ZHANG, Z. et al. Origa-light : An online retinal fundus image database for glaucoma analysis and research. **2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10**, 2010.

APÊNDICE A – LISTA DE ARTIGOS SELECIONADOS

Neste apêndice encontram-se os trabalhos relacionados selecionados na revisão da literatura.

Tabela 5: Lista de artigos selecionados

Citação	Título	Periódico	Ano
(YU et al., 2019)	Robust optic disc and cup segmentation with deep learning for glaucoma detection	Computerized Medical Imaging and Graphics	2019
(CIVIT-MASOT et al., 2020b)	Dual Machine-Learning System to Aid Glaucoma Diagnosis Using Disc and Cup Feature Extraction	IEEE Access	2020
(XU et al., 2021)	A hierarchical deep learning approach with transparency and interpretability based on small samples for glaucoma diagnosis	npj Digital Medicine	2021
(SRENG et al., 2020)	Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images	Applied Sciences (Switzerland)	2020
(PHASUK et al., 2019)	Automated Glaucoma Screening from Retinal Fundus Image Using Deep Learning	IEEE Engineering in Medicine and Biology Society	2019
(SALLAM et al., 2021)	Early Detection of Glaucoma using Transfer Learning from Pre-trained CNN Models	2021 International Conference of Technology, Science and Administration	2021
(CHAI et al., 2021)	Glaucoma diagnosis in the Chinese context: An uncertainty information-centric Bayesian deep learning model	Information Processing and Management	2021
(AHN et al., 2018)	A deep learning model for the detection of both advanced and early glaucoma using fundus photography	PLoS ONE	2018
(SERTE; SERENER, 2019)	A Generalized Deep Learning Model for Glaucoma Detection	3rd International Symposium on Multidisciplinary Studies and Innovative Technologies	2019
(CHAI; LIU; XU, 2018)	Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models	Knowledge-Based Systems	2018
(HEMELINGS et al., 2020)	Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning	Acta Ophthalmologica	2020
(GÓMEZ-VALVERDE et al., 2019)	Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning	Biomedical Optics Express	2019
(KIM et al., 2019)	Medinoid: Computer-aided diagnosis and localization of glaucoma using deep learning	Applied Sciences (Switzerland)	2019

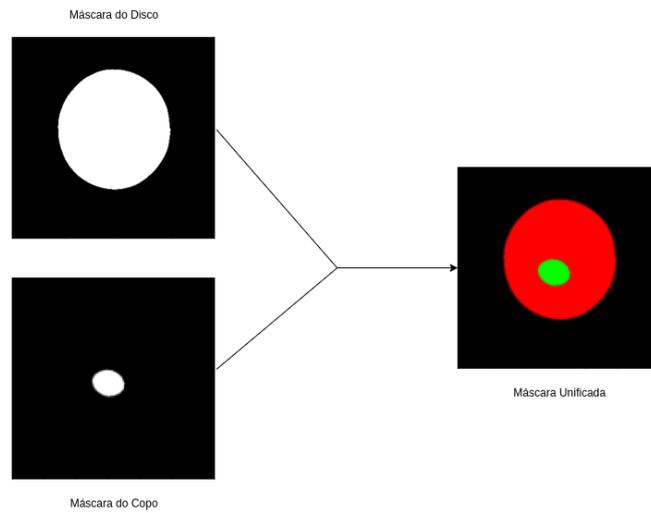
APÊNDICE B – LISTA DE TÉCNICAS E MÉTRICAS OBTIDAS

Tabela 6: Lista das técnicas e das métricas obtidas pelos modelos

Citação	Dataset	Método	Métrica
(YU et al., 2019)	RIGA, DRISHTI-GS e RIM-ONE V3.	U-net modificada	valor dice no disco 97.38% e no copo 88.77%
(CIVIT-MASOT et al., 2020b)	RIM-ONE e DRISHTI	U-net CDR + MobileNet	AUC: 0.96, Acc: 0.88, Sp: 0.86, Se: 0.91
(XU et al., 2021)	privado	U-net + InceptionV3, SVM e Decision Tree	Sp: 0.98 e Se: 0.98
(SRENG et al., 2020)	REFUGE, ACRIMA, ORIGA, RIM-ONE e DRISTI-GS1.	União de 11 redes neurais pré treinadas	AUC: 100% e Acc: 97.37% no RIM-ONE e AUC: 99.54% e Acc: 99.98 no ACRIMA
(PHASUK et al., 2019)	ORIGA-650, RIM-ONE R3 e DRISHTI-GS	ResNet-50 + DenseNet-121	Sobreposição: 0.96 e AUC: 0.94
(SALLAM et al., 2021)	LAG dataset	ResNet-152	Acc: 0.86 Sp: 0.86 e Se: 0.86
(CHAI et al., 2021)	privado	Modelo Bayesiano de Múltiplas Camadas	AUC: 0.94, Acc: 0.92, F1: 0.93
(AHN et al., 2018)	privado	CNN feita à mão	AUC: 0.94 e Acc: 87.9
(SERTE; SERENER, 2019)	HRF, Drishti-GS1, RIM-ONE, sjchoi86-HRF e ACRIMA.	ResNet-50, GoogLeNet e ResNet-152	ResNet-152 no HRF com AUC: 0.95 e Sp 0.93; ResNet-50 no Drishti com AUC: 0.75 e Sp: 0.77; GoogLeNet no RIM-ONE com AUC: 0.83 e Sp: 0.96
(CHAI; LIU; XU, 2018)	privado	Rede neural ramificada	Acc: 0.91, Sp: 0.90 e Se: 0.92
(HEMELINGS et al., 2020)	privado	ResNet-50 + active learning	AUC: 0.98, Sp: 87% e Se: 96%
(GÓMEZ-VALVERDE et al., 2019)	RIM-ONE, DRISHTI e ESPE-RANZA (privado)	VGG-19 + transfer learning & fine tuning	AUC: 0.94, Sp: 89% e Se: 87%
(KIM et al., 2019)	privado	ResNet-152 modificado	Acc: 93%, Sp: 92% e Se: 92%

APÊNDICE C – PROCESSO DE UNIFICAÇÃO DAS MÁSCARAS DE SEGMENTAÇÃO

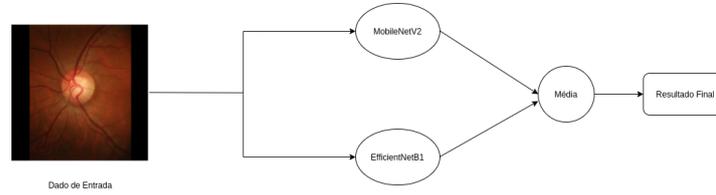
Figura 7 – Processo de Unificação das Máscaras de Segmentação.



Fonte: Autor

APÊNDICE D – SISTEMA PROPOSTO COM AGRUPAMENTO DE CLASSIFICADORES

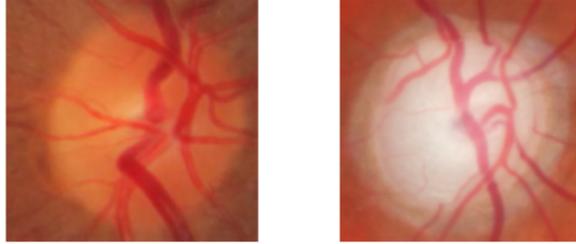
Figura 8 – Sistema de classificação com agrupamento de classificadores.



Fonte: Autor

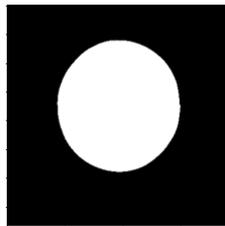
APÊNDICE E – EXEMPLO DE IMAGENS E MÁSCARAS DO CONJUNTO DE DADOS RIM-ONE

Figura 9 – Exemplo de imagens do conjunto de dados RIM-ONE DL.

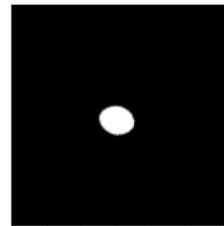


Fonte: (FUMERO et al., 2011)

Figura 10 – Exemplo das máscaras binárias do copo e do disco de uma imagem de fundo de olho.



Máscara binária do disco

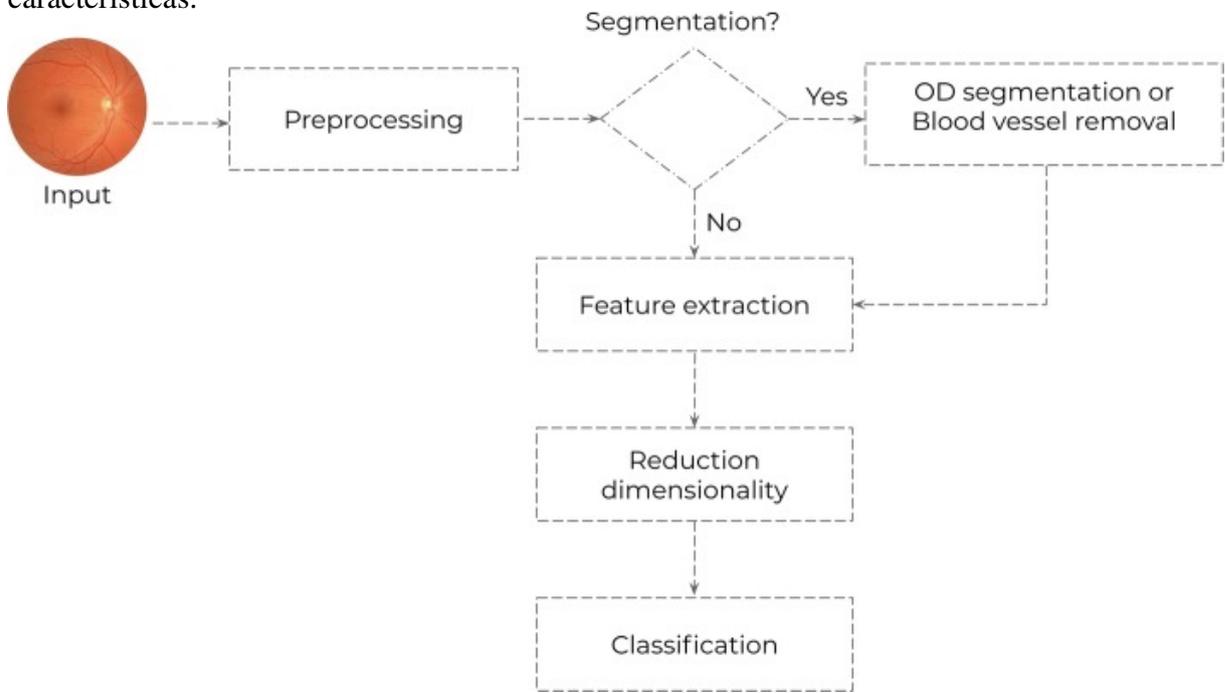


Máscara binária do copo

Fonte: Autor

ANEXO A – ARQUITETURA GENÉRICA DE APRENDIZADO DE MÁQUINA

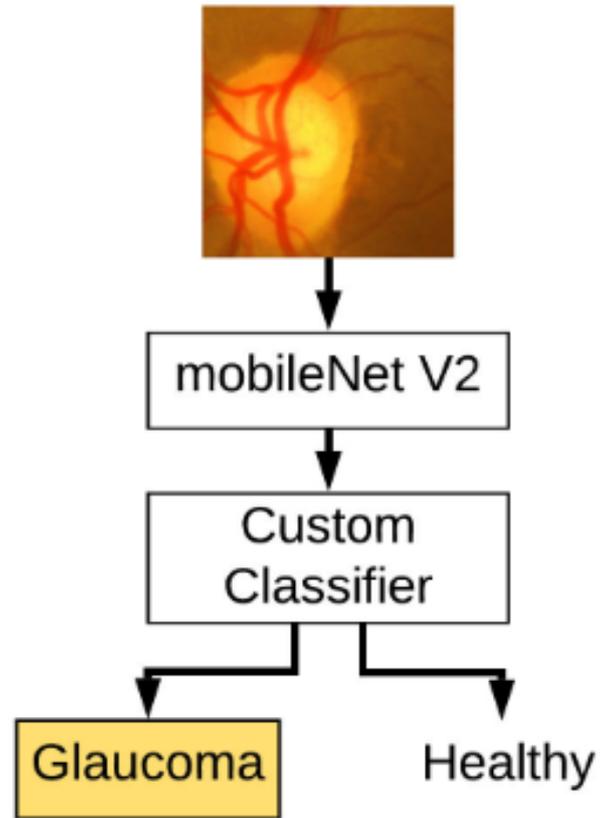
Figura 11 – Arquitetura genérica de um modelo de Aprendizado de Máquina para extração de características.



Fonte: (BARROS et al., 2020)

ANEXO B – SUB-SISTEMA DE CLASSIFICAÇÃO

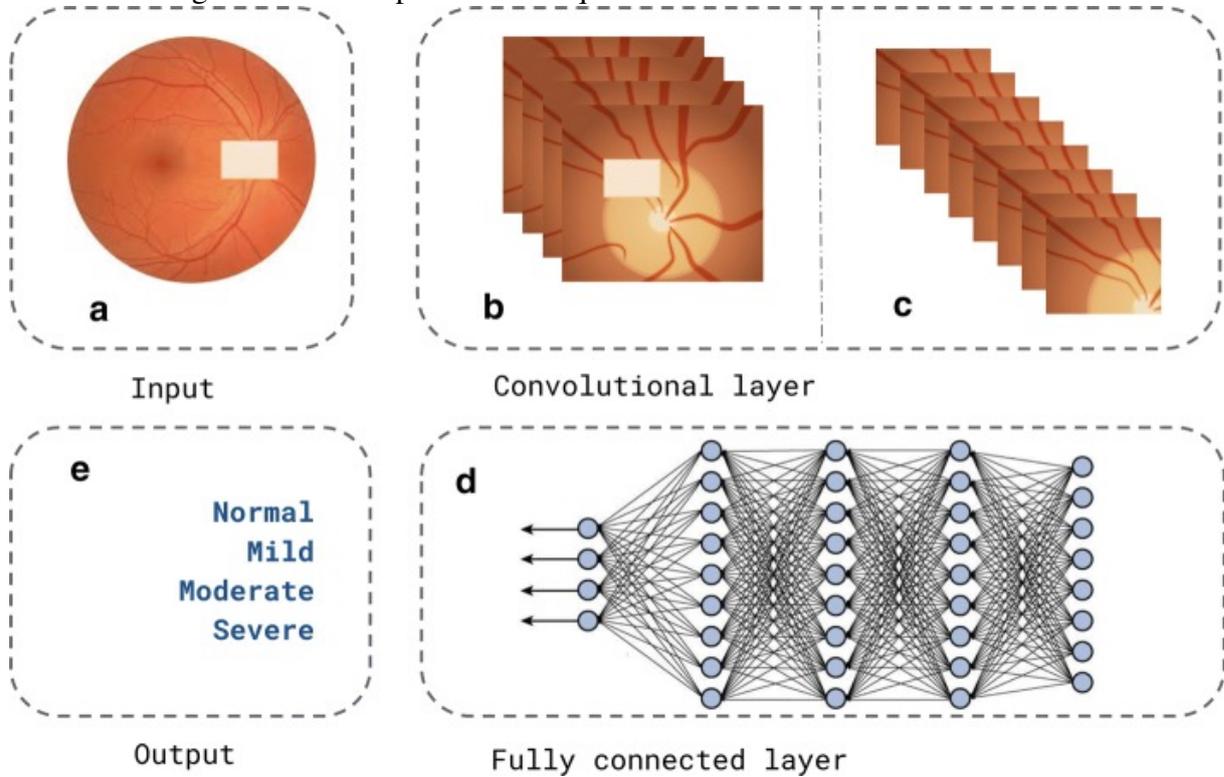
Figura 12 – Sub-sistema de classificação.



Fonte: (CIVIT-MASOT et al., 2020b)

ANEXO C – ARQUITETURA DE REDE NEURAL CONVOLUCIONAL

Figura 13 – Exemplo de uma arquitetura de rede neural convolucional.



Fonte: (BARROS et al., 2020)