



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

DÉBORA FERREIRA DE ASSIS

**AVALIAÇÃO DE MÉTODOS DE CLASSIFICAÇÃO DE GLAUCOMA EM IMAGENS
DE FUNDOSCOPIA**

FORTALEZA

2020

DÉBORA FERREIRA DE ASSIS

AVALIAÇÃO DE MÉTODOS DE CLASSIFICAÇÃO DE GLAUCOMA EM IMAGENS DE
FUNDOSCOPIA

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Processamento de Imagens.

Orientador: Prof. Dr. Paulo César Cortez

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A865a Assis, Débora Ferreira de.
Avaliação de Métodos de Classificação de Glaucoma em Imagens de Fundoscopia / Débora Ferreira de Assis. – 2020.
105 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2020.
Orientação: Prof. Dr. Paulo César Cortez.

1. Glaucoma. 2. Classificação. 3. Segmentação. 4. Rede Neural Convolucional. 5. Otimização de hiperparâmetros. I. Título.

CDD 621.38

DÉBORA FERREIRA DE ASSIS

AVALIAÇÃO DE MÉTODOS DE CLASSIFICAÇÃO DE GLAUCOMA EM IMAGENS DE
FUNDOSCOPIA

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Processamento de Imagens.

Aprovada em: 28 de janeiro de 2020

BANCA EXAMINADORA

Prof. Dr. Paulo César Cortez (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Pedro Pedrosa Rebouças Filho
Universidade Federal do Ceará (UFC)

Prof. Dr. José Marques Soares
Universidade Federal do Ceará (UFC)

Prof. Dr. Auzuir Ripardo de Alexandria
Instituto Federal de Educação, Ciência e Tecnologia
do Ceará (IFCE)

Aos meus pais Maria Lindimar e José Eleilson.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. E com apoio do Laboratório de Engenharia de Sistemas de Computação (LESC) - UFC.

Diante de minha fé, agradeço a Deus por ter me dado saúde, perseverança para superar os obstáculos e pelas pessoas que Ele colocou em minha vida. Agradeço aos meus pais, Maria Lindimar e José Eleilson, por todo amor, compreensão e paciência. Pelo apoio e incentivo para que eu sempre acreditasse que o melhor caminho a seguir é através do conhecimento.

Agradeço aos professores pela dedicação no ensino, e por todo conhecimento que compartilharam com o corpo estudantil. Em especial, ao professor Paulo César Cortez, pela excelente orientação, por todo apoio e incentivo durante a construção dessa dissertação, muito obrigada por todos os ensinamentos e conhecimento repassados. Aos professores participantes da banca examinadora pelo tempo, colaborações e sugestões.

Agradeço ao Alison por sempre acreditar no potencial que eu mesma não sabia que tinha, e apesar de todos seus compromissos, sempre teve tempo para me escutar e aconselhar. Muito obrigada pela companhia durante esses anos, comemorando cada pequena conquista e incentivando novos sonhos. Obrigada por cada palavra que me impulsionou e incentivou durante todo meu percurso até aqui.

Agradeço ao meu amigo Pedro, que sempre esteve presente durante toda a pesquisa, me auxiliou diversas vezes, e compartilhou experiências e aprendizagem. Obrigada pela dedicação e pela cooperação no projeto, sempre trabalhando em equipe. Agradeço também aos meus amigos do laboratório que de alguma maneira ajudaram e aconselharam. Obrigada aos meus amigos Renan, Lucas e Erick que estiveram disponíveis para debater e compartilhar conhecimento. Ao meu amigo John, que mesmo distante sempre apoiou em todos os momentos, acreditou e vibrou em cada conquista, e continua apoiando em todas decisões. Obrigada por tudo, vocês são incríveis!

“Você pode até ter medo, mas ande, caminhe, e só não pare, não pare nunca!”

(Rosa de Saron)

RESUMO

O glaucoma é uma doença ocular assintomática no início que, se não for tratada a tempo, pode acarretar a cegueira. A Organização Mundial de Saúde (OMS) estimou que em, 2020, o glaucoma deve afetar 80 milhões de pessoas e, em 2040, serão 111,5 milhões. Neste contexto, a presente dissertação visa comparar métodos de classificação e estudar diferentes técnicas de extração de características de imagens da base pública RIM-ONE versão 2 para auxiliar o médico especialista no diagnóstico da doença. São desenvolvidos 3 modelos com base em diferentes tipos de extração de características. O modelo 1 extrai características não geométricas previamente indicadas: Padrões Binários Locais (LBP), Histograma de Gradientes Orientados (HOG), momentos de Zernike e informações estatísticas do filtro de Gabor. No modelo 2 adicionam-se ao modelo 1 características geométricas extraídas das imagens. Já no modelo 3 são utilizadas arquiteturas pré-treinadas das redes convolucionais (MobileNet, VGG16, VGG19 e Resnet50) para extração de informações das imagens. Para cada modelo, as características obtidas são submetidas à Análise de Componentes Principais (PCA) para a redução de dimensionalidade, cujas componentes resultantes são classificadas por: Regressão Logística (RL), Árvore de decisão com aumento de gradiente (GBDT), Máquina de vetores de suporte (SVM), k vizinhos mais próximos (k-NN) e Perceptron Multicamadas (MLP). Para melhorar o desempenho de classificação, utilizam-se técnicas de otimização de hiperparâmetros através do algoritmo de pesquisa em grade. Dentre os três modelos avaliados, a extração de características do modelo 1 juntamente com a classificação utilizando o SVM, produzem as maiores taxas médias obtidas através das 100 execuções. Os resultados dos dados de teste atingem uma taxa média de acurácia igual a 89,03%, sensibilidade igual a 86,59%, especificidade igual a 91,06% e área sob a curva (AUC) igual a 88,95%.

Palavras-chave: Glaucoma. Classificação. Segmentação. Rede Neural Convolucional. Otimização de hiperparâmetros.

ABSTRACT

Glaucoma is an asymptomatic eye disease that, if not treated on time, can lead to blindness. The World Health Organization (WHO) estimates that by 2020 glaucoma should affect 80 million people and by 2040 it will be 111.5 million. In this context, the present dissertation aims to compare classification methods and study different techniques for the extraction of image characteristics, thus assisting the specialist physician in diagnosing the disease. Three models are developed based on different types of feature extraction. Model 1 extracts non-geometric characteristics: Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), Zernike moments and Gabor filter statistical information. Model 2 is similar to model 1 with the addition of geometric features. And in model 3, pre-trained convolutional network models (MobileNet, VGG16, VGG19 and Resnet50) are used to extract information from the images. For each model, the obtained characteristics are submitted to Principal Component Analysis (PCA) for dimensionality reduction, the resulting components are classified by: Logistic Regression (RL), Gradient Increasing Decision Tree (GBDT), Support Vector Machine (SVM), k-nearest neighbors (k-NN), and Multilayer Perceptron (MLP). To improve classification performance, hyperparameter optimization techniques using Grid Search are used. Of the three models evaluated, model 1 produces the best results using SVM for classification. The test results achieved an average accuracy rate of 89.03%, sensitivity of 86.59%, specificity of 91.06% and AUC (*area under a curve*) of 88.95%.

Keywords: Glaucoma. Classification. Segmentation. Convolutional Neural Network. Hyperparameter optimization.

LISTA DE FIGURAS

Figura 1 – Estrutura do olho	20
Figura 2 – Imagem típica de fundo de olho e suas principais estruturas.	23
Figura 3 – Gradientes (a) no eixo x G_x , (b) no eixo y G_y e (c) magnitude do gradiente $ G $	27
Figura 4 – (a) Imagem do fundo do olho original, (b) cálculo do HOG e (c) histograma da imagem resultante.	27
Figura 6 – Exemplos de vizinhança com p e r usados para construir LBP.	29
Figura 5 – Região em análise em (b), limiarização da imagem em (b), multiplicação em potências de 2 em (c) e resultado do pixel central em (d).	29
Figura 7 – Imagem original em (a) e representação do LBP em (b).	30
Figura 8 – Imagem do fundo do olho em (a), filtro de Gabor na frequência em (b), e imagem após a aplicação do filtro com ângulo 3 em (c).	33
Figura 10 – Operação de convolução com filtro 3×3	34
Figura 9 – Arquitetura genérica de uma rede neural convolucional.	34
Figura 11 – Operação de subamostragem com kernel 2×2	35
Figura 12 – Modelo de um bloco residual.	37
Figura 13 – Filtro convolucional padrão em (a), convolução em profundidade em (b) e convolução em ponto (c).	38
Figura 14 – Arquitetura VGG16.	40
Figura 15 – Exemplo de vizinhança para o algoritmo k-NN.	43
Figura 16 – Representação genérica de uma árvore de decisão.	44
Figura 17 – Separação linear	46
Figura 18 – Hiperplano de separação.	47
Figura 19 – Arquitetura do neurônio da rede neural Perceptron Multicamadas.	48
Figura 20 – Validação cruzada 5-fold	53
Figura 21 – Representação da curva ROC	55
Figura 22 – Imagem de fundo do olho de paciente, (a) saudável e (b) com glaucoma.	65
Figura 23 – Diagrama de Blocos do modelo proposto.	67
Figura 24 – Imagens original em (a), operação de <i>black top hat</i> na componente verde em (b), junção das componentes RGB após o processamento da componente G em (c), filtragem mediana em (d) e Método do Contorno Ativo em (e).	69

Figura 25 – Imagem pré-processada em (a), limiarização de Otsu em (b), filtragem mediana em (c) e contorno da escavação em (d).	70
Figura 26 – Contorno do disco óptico e escavação de um paciente saudável em (a) e com glaucoma em (b).	70
Figura 27 – Quadrante ISNT.	71
Figura 28 – Borda neuroretiniana e quadrantes ISNT para o cálculo NRR de um fundo de olho saudável em (a) e com glaucoma em (b).	72
Figura 29 – Segmentação dos vasos sanguíneos no disco óptico de um paciente saudável em (a) e com glaucoma em (b).	72
Figura 30 – Frequência dos hiperparâmetros do modelo de regressão logística.	76
Figura 31 – Frequência dos hiperparâmetros do modelo k-NN.	77
Figura 32 – Frequência dos hiperparâmetros do modelo GBDT.	78
Figura 33 – Frequência dos hiperparâmetros do modelo SVM.	79
Figura 34 – Frequência dos hiperparâmetros do modelo MLP.	81
Figura 35 – Boxplot da acurácia do modelo proposto para cada classificador avaliado.	84
Figura 36 – Boxplot da sensibilidade do modelo proposto para cada classificador avaliado.	84
Figura 37 – Boxplot da especificidade do modelo proposto para cada classificador avaliado.	85
Figura 38 – Classificação errônea de um falso negativo (a) e um falso positivo (b).	93
Figura 39 – Tamanhos das saídas e núcleos convolucionais da arquitetura ResNet.	107

LISTA DE TABELAS

Tabela 1 – Matriz de confusão.	54
Tabela 2 – Resumo dos trabalhos relacionados utilizando extração de características previamente indicadas.	58
Tabela 3 – Resumo dos trabalhos relacionados utilizando extração de características geométricas.	61
Tabela 4 – Resumo dos trabalhos relacionados utilizando extração de características por redes neurais convolucionais.	64
Tabela 5 – Resumo dos hiperparâmetros selecionados para o modelo 1.	82
Tabela 6 – Desempenho médio dos dados de teste referente ao modelo 1.	83
Tabela 7 – Resumo dos hiperparâmetros selecionados para o modelo 2.	87
Tabela 8 – Desempenho médio dos dados de teste do modelo 2 referente a 100 execuções.	87
Tabela 9 – Tempo para carregamentos de arquitetura e extração de características.	89
Tabela 10 – Desempenho médio dos dados de teste referente ao modelo 3.	91
Tabela 11 – Comparação do desempenho do método proposto com outros métodos.	95
Tabela 12 – Arquitetura da rede completa MobileNet.	106
Tabela 13 – Arquitetura da rede completa VGG16.	106
Tabela 14 – Arquitetura da rede completa VGG19.	107

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivos	18
1.2	Organização do trabalho	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Estrutura do Olho	19
2.1.1	<i>Glaucoma</i>	21
2.1.1.1	<i>Ângulo aberto</i>	21
2.1.1.2	<i>Ângulo Fechado</i>	22
2.1.1.3	<i>Congênito</i>	22
2.1.1.4	<i>Secundário</i>	22
2.2	Imagens do Fundo de Olho	22
2.2.1	<i>Segmentação do Nervo Óptico</i>	24
2.2.2	<i>Limiarização de Otsu</i>	25
2.2.3	<i>Método de Contornos Ativos</i>	25
2.3	Descritores de Características	25
2.3.1	<i>Histograma de Gradientes Orientado</i>	26
2.3.2	<i>Padrões binários locais</i>	28
2.3.3	<i>Momentos de Zernike</i>	30
2.3.4	<i>Filtros de Gabor</i>	32
2.4	Redes Neurais Convolucionais	33
2.4.1	<i>Aprendizado por transferência</i>	36
2.4.1.1	<i>ResNet50</i>	36
2.4.1.2	<i>MobileNet</i>	38
2.4.1.3	<i>VGG</i>	39
2.5	Métodos de Classificação	40
2.5.1	<i>Regressão Logística</i>	40
2.5.2	<i>k - Vizinhos Mais Próximos</i>	42
2.5.3	<i>Árvore de Decisão com Aumento de Gradiente</i>	44
2.5.4	<i>Máquina de Suporte Vetorial</i>	45
2.5.5	<i>Perceptron Multicamadas</i>	48

2.6	Análise de Componentes Principais	49
2.7	Otimização dos Hiperparâmetros	52
2.7.1	<i>Validação Cruzada</i>	52
2.8	Métricas de Avaliação	53
2.8.1	<i>Matriz de Confusão</i>	54
2.8.2	<i>Área sob a curva</i>	55
3	TRABALHOS RELACIONADOS	56
3.1	Características não geométricas	56
3.2	Características geométricas e não-geométricas	59
3.3	Redes Neurais Convolucionais	61
4	MATERIAIS E MÉTODOS	65
4.1	Ambiente de desenvolvimento e teste	65
4.2	Base de imagens	65
4.3	Métodos	66
4.3.1	<i>Modelo 1</i>	68
4.3.2	<i>Modelo 2</i>	69
4.3.3	<i>Modelo 3</i>	73
5	RESULTADOS E DISCUSSÕES	74
5.1	Modelo 1	74
5.1.1	<i>Regressão Logística</i>	75
5.1.2	<i>k Vizinhos Mais Próximos</i>	77
5.1.3	<i>Árvore de Decisão com Gradiente Aumentado</i>	78
5.1.4	<i>Máquinas de Vetores de Suporte</i>	79
5.1.5	<i>Perceptron Multicamadas</i>	80
5.1.6	<i>Resultados do Modelo 1</i>	81
5.2	Modelo 2	85
5.2.1	<i>Regressão logística</i>	86
5.2.2	<i>k Vizinhos mais próximos</i>	86
5.2.3	<i>Árvore de decisão com aumento de gradiente</i>	86
5.2.4	<i>Máquinas de vetores de suporte</i>	86
5.2.5	<i>Perceptron Multicamadas</i>	86
5.2.6	<i>Resultados do modelo 2</i>	87

5.3	Modelo 3	87
5.3.1	<i>Regressão logística</i>	88
5.3.2	<i>k Vizinhos mais próximos</i>	88
5.3.3	<i>Árvore de decisão com aumento de gradiente</i>	88
5.3.4	<i>Máquinas de vetores de suporte</i>	88
5.3.5	<i>Perceptron Multicamadas</i>	89
5.3.6	<i>Resultados do modelo 3</i>	89
5.4	Discussões dos resultados	91
6	CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS . . .	96
	REFERÊNCIAS	97
	APÊNDICES	106
	ANEXOS	106
	ANEXO A–ARQUITETURAS DOS MODELOS CNN PRÉ-TREINADOS	106

1 INTRODUÇÃO

O glaucoma é uma doença ocular crônica que, se não for tratada, pode acarretar a perda de visão. Dentre as doenças oculares, essa é a segunda doença que mais causa cegueira no mundo, permanecendo atrás somente da catarata que é uma doença curável. Dentre vários distúrbios oculares, o glaucoma é o que mais se associa à intensidade de pressão intraocular alterada, a qual provoca lesões no nervo óptico, comprometendo assim o campo visual (KINGMAN, 2004).

Estima-se que o glaucoma deverá afetar em 2020 aproximadamente 76 milhões de pessoas no mundo, e esse número tende a crescer em 2040, alcançando 111,8 milhões de pessoas. Por consequência, pesquisas que auxiliem no seu diagnóstico possuem grande importância para a saúde pública (THAM *et al.*, 2014).

Essa doença possui tratamento, porém não tem cura. Nos estágios iniciais, o glaucoma tende a ser assintomático, seus sintomas aparecem somente nos estágios mais avançados da doença. A descoberta precoce se torna imprescindível, pois, existem tratamentos que evitam a sua progressão, consequentemente, a perda da visão do paciente (CLARO *et al.*, 2019).

O melhor meio de detectar essa doença é através de exames periódicos realizado por um oftalmologista. Um exame comumente utilizado para obter informações acerca do fundo de olho é a Fundoscopia. A partir deste exame é possível visualizar as estruturas do fundo de olho, com especial atenção para o nervo óptico, os vasos retinianos e a retina propriamente dita, particularmente sua região central denominada mácula (PINHEIRO, 2019). Essas regiões são importantes para identificar o glaucoma, visto que possibilita a verificação de alterações no nervo óptico, avalia resposta do reflexo da pupila e acuidade visual. Além disso, conforme OLIVEIRA V. K. (2012), outros exames confirmatórios também podem ser realizados, como a tonometria que mede a pressão do olho, e a gonioscopia que determina o ângulo da câmara anterior do olho, a fim de classificar o tipo de glaucoma.

A Fundoscopia, realizada através do aparelho retinógrafo, é a técnica de triagem mais utilizada para a detecção de glaucoma, sendo preferível aos demais exames oculares, visto que é mais econômico e simples de ser realizado do que outros métodos de análise de retina disponíveis (ABRÀMOFF *et al.*, 2010). Todavia, sua análise é limitada pela visão humana que incorpora fatores subjetivos, tornando-a dependente do especialista. Nestas imagens digitais, obtidas por esse exame, ou observadas visualmente pelo especialista, a primeira avaliação do médico para detectar o glaucoma é através das estruturas do nervo óptico, principalmente as áreas do disco óptico e escavação. Entretanto, em estágios iniciais não é possível detectar claramente,

pois a avaliação visual é subjetiva, isto porque há discrepância de valores estimados, com base na observação visual de imagens de fundoscopia.

Alguns médicos qualificam a informação do Cup to Disk Ratio (CDR), que relaciona a área da escavação em relação ao disco óptico, como umas das principais características para avaliar a progressão do glaucoma. Por exemplo, um médico qualifica uma imagem da área central do nervo óptico com um valor igual a 0,3, enquanto outro médico indica para a mesma imagem um valor CDR ser igual a 0,5, ou seja, o primeiro aponta que o paciente está saudável, enquanto o segundo indica que o indivíduo tem a possibilidade de ter a doença. O valor de CDR é uma das principais características geométricas para identificar o glaucoma, porém, a análise visual sem um cálculo acurado, não a torna uma característica confiável, caso seja analisado individualmente e por inspeção visual humana.

Numa situação real, dependendo do aparelho utilizado para a aquisição das imagens, as imagens registradas não possuem a mesma qualidade e os ângulos de aquisição podem diferir. Assim, trabalhos como o de Thakur e Juneja (2017) que localizam a escavação pela região mais clara, podem não ser eficientes. Isto porque essa região é facilmente confundida com feixes de luz mais claros que aparecem na retina no momento de aquisição da imagem do disco óptico, que é normal em pacientes mais jovens. Por esse fator, é importante que haja métodos automáticos que possibilitem a classificação de imagens de fundo de olho, e que incluam uma grande variedade de imagens, auxiliando em um diagnóstico médico mais exato e eficaz.

Existem duas abordagens para classificação de imagens por método de aprendizado supervisionado: aprendizado de máquina e métodos de aprendizado profundo por redes convolucionais. O primeiro método consiste em extrair características da imagem usando descritores apropriados e que posteriormente servem como entrada para um algoritmo classificador. Já no segundo método, extraem-se características, analisam-se padrões e os classificam em um único processo. Pesquisas envolvendo este segundo método apresenta crescimento nos últimos anos, isto porque tal técnica apresenta flexibilidade, como a possibilidade de extrair apenas as características através da sua arquitetura e classificar com métodos comuns de aprendizado de máquinas (LECUN *et al.*, 2015).

Recentemente, diversos trabalhos foram desenvolvidos para detecção de glaucoma a partir de imagens de fundo do olho. Algumas abordagens utilizaram para classificação de imagens somente as características geométricas, obtidas a partir da segmentação do disco óptico, como é o caso de Zahoor e Fraz (2018), na qual segmenta o disco óptico e obtém, a partir da

regra I-Inferior S-Superior N-Nasal T-Temporal (ISNT) as informações de CDR, Neuro-Retinal Rim (NRR) e Blood Vessel Ratio (BVR). Tais informações foram usadas para classificar as imagens em glaucomatosas e saudáveis. No trabalho desenvolvido por Salam *et al.* (2016), além de usar características geométricas também fizeram uso de características não geométricas, como informações de textura e intensidades. Enquanto no artigo de Septiarini *et al.* (2018) foram utilizadas somente as informações estatísticas das imagens para a classificação.

Além disso, vários autores abordaram redes convolucionais para a análise das imagens, como no trabalho desenvolvido por Al-Bander *et al.* (2017), em que, após um pré-processamento, por meio de CNN extraem-se características do disco óptico e as classificam com o Support Vector Machine (SVM) a fim de distinguir imagens glaucomatosas e normais. Já no trabalho de Gómez-Valverde *et al.* (2019) utilizam-se técnicas de modelos pré-treinados para extrair e classificar as imagens da retina em um único processo. A vantagem de tais técnicas de aprendizado por transferência é porque evitam o treinamento da rede a partir dos pesos com inicialização aleatória, ou seja, que ainda irá aprender características importantes a serem extraídas das imagens. Com o aprendizado por transferência o treinamento se torna mais rápido e com menos épocas de treinamento (YU *et al.*, 2019).

Apesar desses trabalhos apresentarem desempenhos com altas taxas de acerto na segmentação e classificação, ainda há deficiência na avaliação dos resultados obtidos por essas técnicas. Isto porque não é possível identificar se tais metodologias englobam diferentes bases de imagens ou envolvem somente a base estudada, em sua maioria bases privadas, o que dificulta em reproduzir os resultados pela impossibilidade de acesso a essas imagens.

Neste contexto, a presente dissertação propõe, a partir de uma avaliação de vários métodos de classificação automática do glaucoma, analisar diferentes técnicas e identificar uma de fácil implementação e de menor custo computacional, capaz de distinguir pacientes com glaucoma a partir de exames do fundo do olho. Esta metodologia utiliza-se de aprendizagem de máquina supervisionada para classificação das imagens, na qual são comparados os desempenhos de cinco classificadores com base nos diferentes tipos de extração de características estudados.

São analisados três modelos: o modelo 1 que é baseado nas características não geométricas previamente selecionadas; o modelo 2 que é composto do modelo 1 com a adição das características geométricas obtidas pela segmentação de imagens; e o modelo 3 que se refere à extração de características por modelos de aprendizagem profundo pré-treinados.

1.1 Objetivos

O principal objetivo desta dissertação é avaliar métodos de classificação automática do glaucoma para auxiliar no diagnóstico por um médico especialista. Para tal, alguns objetivos específicos devem ser atingidos:

- estudar diferentes métodos de extração de características das imagens do fundo de olho;
- verificar se a inclusão de características geométricas auxilia na identificação do glaucoma;
- utilizar classificadores, de aprendizado supervisionado, com diferentes configurações de hiperparâmetros para identificar qual modelo melhor se ajusta aos dados, obtendo bom desempenho nos dados de treino e, conseqüentemente, nos dados de teste;
- avaliar os diferentes classificadores e identificar qual o modelo que distingue com maior precisão as classes de imagens, em particular o que apresenta maior sensibilidade;
- comparar os modelos de extração de características previamente indicadas com a extração de características por modelos pré-treinados; e
- analisar vantagens e desvantagens de cada um dos três modelos para resultar em uma técnica que produza os resultados mais precisos para a base de dados utilizada.

1.2 Organização do trabalho

Esta dissertação está organizada da seguinte maneira, no Capítulo 2 são apresentadas as fundamentações teóricas utilizadas, na qual contém as definições da estrutura do olho, doenças oculares (em especial o glaucoma), métodos de segmentação do disco óptico e escavação para o cálculo das características geométricas, descritores de características, métodos de classificação, modelos pré-treinados de redes convolucionais, técnicas de redução de dimensionalidade, otimização dos hiperparâmetros e por fim as métricas de avaliação dos resultados. No Capítulo 3 são apresentados os trabalhos que foram utilizados como base para o desenvolvimento desta dissertação, contendo estudos que utilizaram segmentação e classificação de imagens. Já no Capítulo 4 são descritos os materiais e métodos utilizados para o desenvolvimento da dissertação, e no Capítulo 5 são apresentados e discutidos os resultados e, por fim, no Capítulo 6 apresentam-se as conclusões, contribuições e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este Capítulo dedica-se à descrição de conceitos utilizados para o desenvolvimento desta dissertação, a divisão é feita em 8 seções. A seção 1 descreve a estrutura do olho e doenças oculares, em especial o glaucoma. A seção 2 apresenta a segmentação do disco e escavação, as quais são responsáveis pelo cálculo das características geométricas CDR, BVR e NRR. Na seção 3, são apresentados os métodos de extração de características não geométricas previamente indicadas, sendo explicada matematicamente a configuração de cada descritor: Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), momentos de Zernike e filtros de Gabor. Na seção 4, são explanadas as redes convolucionais e descritos os modelos pré-treinados: MobileNet, ResNet50, VGG16 e VGG19. A classificação é realizada com os algoritmos de aprendizagem supervisionada RL, k-NN, GBDT, SVM e MLP, descritos na seção 5. Na seção 6 discute-se o uso da Análise de Componentes Principais para redução de dimensionalidade. A seção 7 descreve o método *Grid Search* para otimização dos hiperparâmetros com validação cruzada, e na seção 8 se definem as métricas de avaliação dos resultados utilizados na comparação de desempenho dos classificadores.

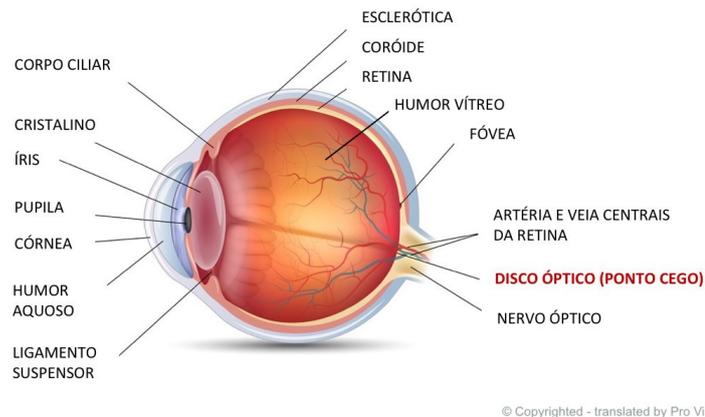
2.1 Estrutura do Olho

O olho humano é o aparelho sensorial responsável pela visão, possui um formato aproximadamente esférico com um diâmetro em torno de 24 milímetros (mm), 75 mm de circunferência, 6,5 centímetros cúbicos (cm³) de volume e peso de 7,5 gramas. Esse órgão é formado por três camadas principais: a camada externa que é responsável pela proteção do globo ocular, formada pela córnea e esclera; a camada intermediária que é a região transmissora, formada pela íris, coroide, cório e o corpo ciliar; e a camada interna que é a parte sensorial contendo células nervosas, formada pela retina (MAGALHAES, 2019). Essas estruturas podem ser visualizadas na Figura 1.

O processo de formação de imagens é complexo, porém rápido. O olho é responsável por captar a luz e transformá-la em impulsos nervosos, esses impulsos transmitem a informação sensorial para o cérebro que a transforma em imagens.

O globo ocular contém substâncias que permitem a projeção de luz sobre a retina, sejam transparentes ou translúcidos. Esses raios luminosos passam pela córnea, o cristalino, o humor aquoso e o humor vítreo e atingem a retina. Na retina, mais de cem milhões de células

Figura 1 – Estrutura do olho



Fonte: Adaptado de (SEGRE, 2019).

fotorreceptoras transformam sinais luminosos em impulsos eletroquímicos que são interpretados pelo cérebro. Nos seres humanos, existem dois principais tipos de células fotorreceptoras: os cones (permite enxergar na presença de luz, garantindo a visão em cores) e os bastonetes (permite a visão noturna em preto e branco). A lente do olho produz uma imagem invertida e o cérebro a converte para a posição correta. Em pessoas com vícios de refração, a imagem pode ser projetada mais para trás ou para frente, esses casos são corrigidos com auxílio de lentes refratoras (HELENE; HELENE, 2011).

Segundo dados da Organização Mundial de Saúde (OMS), atualmente cerca de 285 milhões de pessoas estão visualmente prejudicadas no mundo, dos quais, entre 60% a 80% dos casos podem ser evitados e tratados. No Brasil, o último Censo Demográfico divulgado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2010, identificou, por exame de fundo de olho, mais de 35 milhões de pessoas com a algum grau de dificuldade visual (FRASAO, 2019).

O exame de fundo de olho é realizado desde 1851, quando Von Helmholtz criou o primeiro oftalmoscópio, constitui o principal vínculo entre a oftalmologia e os demais ramos da medicina (MARTINS *et al.*, 2014). De acordo com informações de NETTO *et al.* (2004), este exame, conhecido por fundoscopia, é um exame não invasivo realizado por um médico especialista, no qual fornece informações visuais clínicas que avalia pacientes com doenças sistêmicas e oculares. As principais estruturas oculares avaliadas no referido exame de fundoscopia são: câmara anterior, pupila, cristalino, retina, nervo óptico, mácula, artérias e veias.

Segundo o Ministério da Saúde, no Brasil, as doenças oculares responsáveis pela maior parte dos atendimentos feitos pelos oftalmologistas são: catarata, glaucoma, conjuntivite,

retinopatia diabética, degeneração macular relacionada à idade e erros de refração (miopia, hipermetropia, astigmatismo e presbiopia ou vista cansada). Algumas doenças oculares, a longo prazo, levam a cegueira, como é o caso do glaucoma.

2.1.1 Glaucoma

De acordo com a Organização Mundial da Saúde, dentre as doenças oculares, a segunda doença que mais causa cegueira no mundo é o glaucoma. É um distúrbio ocular crônico caracterizado por um padrão particular de dano progressivo ao nervo óptico, geralmente associada ao aumento da pressão intraocular, pode causar danos irreversíveis ao disco óptico que atrofia de modo progressivo o campo visual, podendo levar à cegueira (KINGMAN, 2004).

Essa doença tem tratamento, porém não tem cura. O tratamento inicial é feito com colírios que diminuem a pressão intraocular, entretanto, se o estágio for muito avançado, o indivíduo precisa realizar cirurgia. É uma doença silenciosa que pode levar à cegueira, isso porque os sintomas somente são percebidos nos estágios mais avançados da doença, quando a visão já está comprometida. A incidência de glaucoma é mais comum em pessoas com diabetes, traumas oculares, idade superior a 35/40 anos e também pode ocorrer devido à hereditariedade. Os tipos de glaucoma podem ser categorizados em: ângulo fechado, ângulo aberto, congênito e secundário (BOYD, 2019).

2.1.1.1 Ângulo aberto

No glaucoma de ângulo aberto, os canais de drenagem dos olhos são obstruídos por partículas minúsculas e microscópicas ao longo de meses ou anos, e a pressão intraocular elevada causa um dano permanente no nervo óptico, causando perda do campo visual (BOYD, 2019).

Nos estágios iniciais da doença, o paciente não identifica os sintomas, pois é indolor. Os sintomas são desenvolvidos ao longo do tempo, a visão passa a ter pontos cegos, o paciente enxerga manchas geradas na visão, os pontos crescem e se fundem causando a perda de visão gradualmente. A primeira área afetada é a região periférica, a visão central é a última a ser perdida, desse modo, muitas pessoas enxergam o que está a sua frente, mas não enxergam as outras direções.

2.1.1.2 Ângulo Fechado

O glaucoma de ângulo fechado ocorre quando a saída do humor aquoso é fisicamente obstruída, que origina um aumento demasiado da pressão intraocular, sendo classificado em crônico ou agudo. No glaucoma de ângulo fechado crônico, a pressão ocular aumenta lentamente, geralmente os sintomas são vermelhidão nos olhos, desconforto, visão turva, ou uma dor de cabeça que diminui ao dormir (BOYD, 2019).

Já no glaucoma de ângulo fechado agudo, a pressão ocular aumenta rapidamente, e os sintomas são fortes dores nos olhos e de cabeça, vermelhidão nos olhos, visão turva, ver halos em arco-íris em volta das luzes, e tem perda súbita de visão. Nesse caso, o atendimento é emergencial, pois, a pessoa pode perder a visão total em poucas horas após a aparição dos sintomas. Em alguns casos, o paciente necessita de intervenção cirúrgica para reverter os impactos causados (REFOSCO *et al.*, 2014).

2.1.1.3 Congênito

O glaucoma congênito é, como o próprio nome diz, transmitido geneticamente. É observado, por exemplo, em uma criança que já nasce com a doença. Este tipo é considerado raro e se descoberto no início, o tratamento é imediato (COELHO, 2018).

2.1.1.4 Secundário

Com base nas informações do *National Eye Institute*, o glaucoma secundário pode se desenvolver como complicações de outras condições médicas, por exemplo, resultado de diabetes mal controlada ou pressão alta. Ocorrem com catarata, certos tumores oculares ou quando o olho está inflamado ou irritado por uma condição chamada uveíte. É possível desenvolver após outras cirurgias oculares ou lesões oculares graves. Bem como por drogas esteroides usados para tratar inflamações oculares e outras doenças (FEIBEL; BIGGER, 1972).

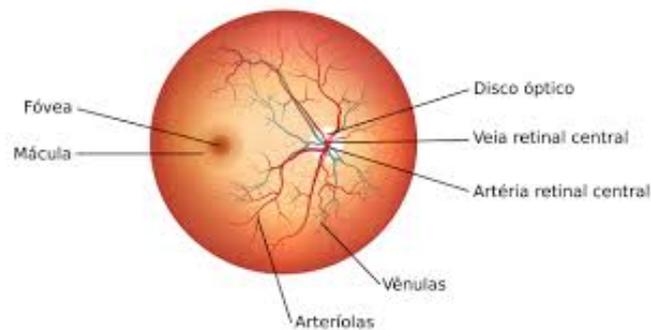
2.2 Imagens do Fundo de Olho

A região do nervo óptico é uma das regiões mais importantes para descrever anomalias oculares - apesar disso, não existe um padrão para todas as pessoas, pois, algumas características são normais para uma pessoa e essas mesmas características podem ser indícios

de anomalias para outras. Assim, essas regiões são importantes, pois servem para identificar parâmetros que ajudam a descrever a presença ou não de glaucoma. Contudo, estudar apenas um parâmetro não é confiável, é preferível estudar um conjunto de parâmetros geométricos para determinar a presença e estágio da doença.

Conforme o Conselho Internacional de Oftalmologia, os sinais característicos do glaucoma podem ser destacados a partir de imagens do nervo óptico. As lesões iniciais deste nervo podem incluir uma escavação em torno de 0,4, ou seja, quatro décimos do disco óptico é ocupado pela escavação, defeitos focais na camada de fibras nervosas da retina, afilamento focal do anel, escavação vertical, assimetria na escavação/disco, escavação focal e hemorragia no disco. As lesões médias a avançadas no nervo óptico podem apresentar uma escavação ampla entre 0,4 e 0,7, ou seja, de quatro a sete décimos do disco óptico é ocupado pela escavação, defeitos difusos nas fibras nervosas da retina, estreitamento difuso do anel, cavidade do nervo óptico, fosseta adquirida do nervo óptico e hemorragia no disco (GUPTA *et al.*, 2016). As principais estruturas avaliadas por um médico no exame de fundo de olho para identificar o glaucoma podem ser visualizadas na Figura 2.

Figura 2 – Imagem típica de fundo de olho e suas principais estruturas.



Fonte: (SILVA M. A. da, 2018).

O disco óptico, Optic Disc (OD), é a porção do nervo óptico vista no fundo dos olhos, formado pelo encontro de todos os axônios das células ganglionares da retina quando penetram no nervo óptico. O OD é uma estrutura circular que apresenta uma tonalidade rosa pálido ou esbranquiçada e bordas bem definidas, situa-se no lado nasal da retina. Em média, o disco óptico mede aproximadamente 1,92 mm verticalmente em 1,76 mm horizontalmente (HOOVER; GOLDBAUM, 2003).

A escavação é a região central no nervo óptico, formada pela ausência de fibras do

nervo que abrange cerca de 1/3 da área do disco. Seu tamanho depende do tamanho do orifício (quanto maior a escavação, mais espaço existe para as fibras passarem) e do número de fibras nervosas.

Os vasos retinianos projetam-se radialmente a partir do disco óptico, com a veia acompanhando a artéria na maioria das vezes. Os vasos retinianos transportam os produtos de excreção das células da retina, sua aparência varia com a forma do globo ocular e com as alterações refrativas. Desse modo, como há variação em diferentes pessoas, as informações de cores não são importantes para distinção de incidência de glaucoma (NETTO *et al.*, 2004).

Essas regiões descritas são consideradas as mais importantes nesta dissertação para o cálculo das características geométricas oculares, isto porque as características selecionadas são baseadas na região do nervo óptico. Para isso, é necessário identificar as bordas da região do disco óptico e da escavação. A partir dessas informações, é possível calcular os parâmetros estabelecidos na identificação do glaucoma.

2.2.1 Segmentação do Nervo Óptico

O processo de subdividir a imagens em objetos ou regiões de uma imagem digital é denominado de segmentação. Há dois tipos: baseada nas bordas e a segmentação baseada em região. A segmentação baseada nas bordas é a mais usada, já a segmentação baseada em regiões tem como fundamento a divisão de uma imagem em regiões que sejam semelhantes de acordo com critérios predefinidos (GONZALEZ; WOODS, 2017).

O processo mais simples de segmentação de uma imagem digital é chamado de limiarização que se baseia na diferença dos níveis de cinza que compõe diferentes objetos de uma imagem. Em geral, é mais simples quando a imagem apresenta um histograma bimodal, ou seja, há apenas dois objetos presentes que podem ser o fundo e um objeto ou região de interesse. Neste caso, com a informação presente no histograma da imagem, escolhe-se um limiar que separe os dois objetos. A partir desse limiar estabelecido, de acordo com as características dos objetos que se deseja separar, a imagem pode ser segmentada em dois grupos: o grupo de *pixels* com níveis de cinza abaixo do limiar e o grupo de *pixels* com níveis de cinza acima do limiar. Em uma imagem limiarizada, atribui-se um valor fixo para todos os *pixels* de um mesmo grupo (GONZALEZ; WOODS, 2017).

Entretanto, a escolha de um valor ótimo ou ideal de um limiar pode não ser fácil de se obter quando há superposição de níveis de cinza dos objetos a serem segmentados.

2.2.2 Limiarização de Otsu

O método de Otsu é um algoritmo de limiarização, proposto em 1975 por Nobuyuki Otsu. Seu objetivo é, a partir de uma imagem em tons de cinza, determinar o valor ideal de um limiar que separe os elementos do fundo e de um objeto ou região da imagem em dois grupos, atribuindo a cor diferente para cada um deles. Devido à essa característica, funciona bem para casos de imagens com histogramas bimodais, podendo ser divididas adequadamente com um único valor. A diferença na aplicação desse caso é que o algoritmo deve buscar maximizar a variância interclasse. O valor que atinge esse objetivo é o mesmo limiar que minimiza a variância intraclasse e, portanto, resulta na limiarização ideal (TOROK; CONCI, 2016).

2.2.3 Método de Contornos Ativos

De acordo com Rebouças REBOUÇAS FILHO P. P.; CORTEZ P. C. et al. (2013) o Método de Contornos Ativos (MCA) é um método que permite realizar a segmentação de objetos em imagens digitais por meio da detecção de bordas obtidas por informações da imagem, como exemplo, a intensidade dos níveis de cinza. Tal método tem superado as técnicas tradicionais, até então desenvolvidas, como limiarização, crescimento de regiões, detecção de bordas, *Watershed*, dentre outros. Um MCA consiste em traçar uma curva inicial, em torno ou dentro de um objeto de interesse, que se deforma, conforme algumas forças que atuam sobre o mesmo, deslocando-a até as bordas do objeto. Este processo é realizado por iterações sucessivas de minimização de uma dada energia, associada à curva.

O contorno ativo é um dos modelos ativos de técnicas de segmentação, que utiliza as restrições e forças de energia na imagem para a separação da região de interesse. Esta técnica define um limite ou curvatura separada para as regiões do objeto de destino para segmentação. O contorno depende de várias restrições com base nas quais são classificados em diferentes tipos, como fluxo de vetor gradiente e modelos geométricos. Modelos de contorno ativos são usados em várias aplicações de processamento de imagem, especificamente no processamento de imagens médicas (HEMALATHA *et al.*, 2018).

2.3 Descritores de Características

Os descritores de características são utilizados para extrair informações das imagens para que as torne mais representativa e adequadas para classificá-las. Estes buscam representar

as informações mais relevantes da imagem, segundo um aspecto específico como: cor, textura, forma dos objetos, dentre outras características. Tal procedimento é realizado para que seja possível aplicar técnicas de identificação, como algoritmos de busca ou reconhecimento de padrões. Neste contexto, cada imagem passa a ser representada por um vetor de características.

A escolha de descritores está associada a aplicação específica, após testes exaustivos com diferentes descritores de imagens, os extratores de características que mais adequaram-se a base estudo são: HOG, LBP, momentos de Zernike e informações estatísticas obtidas das imagens processadas pelo filtro de Gabor.

2.3.1 *Histograma de Gradientes Orientado*

Os descritores Histograma de Gradientes Orientados, *Histogram of Oriented Gradients* (HOG), generalizaram-se em 2005 quando Navneet Dalal e Bill Triggs, pesquisadores do Instituto Nacional Francês de Pesquisa em Ciência da Computação e Automação (INRIA), apresentaram seu trabalho intitulado *Histogram of Oriented Gradients for Human Detection*, no qual estudaram a detecção de pedestres em imagens estáticas (DALAL; TRIGGS, 2005).

O HOG é usado em processamento de imagens como descritor de características e também na detecção de objetos. O HOG é um descritor de características globais, que contém informações da aparência e forma do objeto em uma imagem, através da distribuição de gradientes de intensidade ou direções de borda (ALMEIDA *et al.*, 2018).

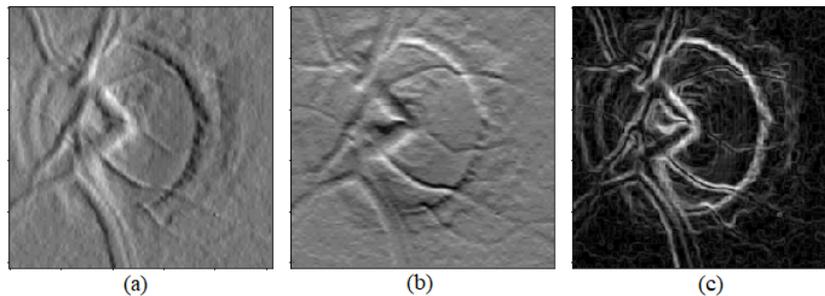
O descritor de características é importante por fornecer uma representação compacta da imagem. Por exemplo, uma imagem 8×8 contém $8 \times 8 \times 3 = 192$ *pixels*, já seu gradiente contém valores da magnitude e direção por *pixel* que somam $8 \times 8 \times 2 = 128$ números. Esses números são representados usando um histograma, além de ser mais compacto, o cálculo do histograma torna a representação mais robusta ao ruído (MALLICK, 2016).

A implementação desse descritor requer dividir a imagem em pequenas regiões conectadas chamadas células e, em seguida, para cada célula, computar um histograma de gradientes orientados para os *pixels* em cada célula. Estes histogramas são acumulados em várias células para formar o vetor de características. A implementação desse descritor é realizada da seguinte forma (DALAL; TRIGGS, 2005):

1. opcionalmente, as imagens são normalizadas para atenuar os efeitos de iluminação, reduzindo os efeitos das variações locais de sombreamento;
2. logo após, os gradientes de primeira ordem são calculados, que fornecem características

como contorno, borda e textura. A operação de convolução é aplicada para obter as imagens de gradiente $G_x = I \star D_x$ e $G_y = I \star D_y$, nos quais I é a imagem de entrada, D_x é o filtro na direção x e D_y é o filtro na direção y ; Em seguida, a magnitude e direção do gradiente são obtidas por $|G| = \sqrt{G_x^2 + G_y^2}$. E a orientação do gradiente para cada *pixel* na imagem de entrada pode ser calculada por $\theta = \arctan \frac{G_y}{G_x}$ (MALLICK, 2016). Na Figura 3 é possível visualizar a os gradientes nos eixos x e y , e a magnitude do gradiente;

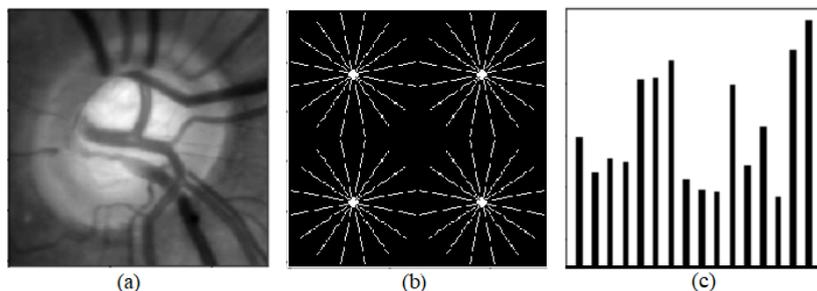
Figura 3 – Gradientes (a) no eixo x G_x , (b) no eixo y G_y e (c) magnitude do gradiente $|G|$.



Fonte: elaborado pelo autor.

3. a terceira etapa produz uma codificação que seja resistente a pequenas mudanças, mantendo a aparência da imagem. Em seguida, a imagem é dividida em células e blocos. E, para cada uma das células da imagem, é construído um histograma de gradientes orientados usando a magnitude $|G|$ e a orientação do gradiente θ , então, para cada uma dessas células, calcula-se um HOG usando 8 orientações por histograma, como mostrado na Figura 4;

Figura 4 – (a) Imagem do fundo do olho original, (b) cálculo do HOG e (c) histograma da imagem resultante.



Fonte: elaborado pelo autor.

4. a quarta etapa calcula a normalização do resultado, introduzindo uma invariância à iluminação, ao sombreamento e ao contraste e bordas;
5. a etapa final acumula os descritores do HOG de todos os blocos, de uma densa grade

sobreposta de blocos que cobrem a janela de detecção, em um vetor de características combinadas.

2.3.2 *Padrões binários locais*

Os Padrões Binários Locais, do inglês Local Binary Pattern (LBP), foram descritos por Ojala *et al.* (1996) e são um caso particular do modelo de Expectro de Textura, definido por Wang *et al.* (2009). Esse método apresenta uma abordagem simples para classificação de texturas em níveis de cinza e é invariante a rotação. O LBP é um descritor local, baseado na ideia de que a informação de uma textura pode ser descrita pelos padrões espaciais locais e pelo contraste dos níveis de cinza.

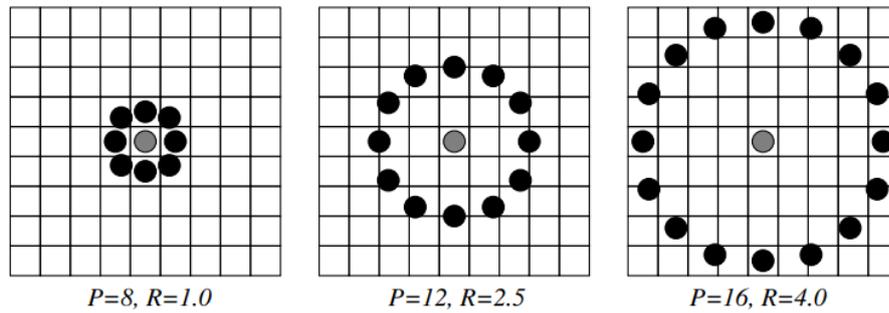
O LBP se baseia na suposição de que a informação de uma textura pode ser decomposta em conjuntos de pequenas unidades, chamadas de unidades texturais. É um descritor robusto a mudanças monótonas em escala de cinza, como variações de iluminação, além disso, sua simplicidade computacional permite analisar imagens em configurações desafiadoras em tempo real. Apesar das vantagens, esse método reduz seu desempenho na presença de ruídos aleatórios (PIETIKÄINEN, 2010).

O LBP é definido como um conjunto de características de texturas em escala de cinza, cujos valores são calculados através de um valor binário atribuído a cada *pixel* da imagem, formando uma vizinhança linear de raio R em torno do *pixel* principal posicionado em uma região central (AQUINO, 2017). Devido o seu poder discriminativo e simplicidade computacional, o LBP tornou-se uma abordagem popular em várias aplicações. Diversos autores vêm aumentando a abrangência do LBP, várias extensões e modificações foram propostas, algumas delas estão descritas por Wang *et al.* (2015).

A sua configuração original considera uma vizinhança 3×3 , em torno de cada *pixel*, compara-o com o *pixel* central e se atribui o valor igual a 1 para o *pixel* se for maior ou igual ao *pixel* central, e o valor igual a 0, no caso contrário. O código LBP é calculado como um número de 8 bits, em que o limiar de cada *pixel* recebe os pesos 2^n , com $n = 0, \dots, p - 1$. Por fim, é construído um histograma dos códigos LBP calculados (MONTAGNER, 2014).

Um exemplo pode ser visto na Figura 5 (a), na qual se tem a região 3×3 em análise, cujo pixel central está circulado. Na Figura 5 (b) mostra-se a região obtida pela limiarização, na Figura 5 (c) pode-se observar a distribuição dos pesos em potência de 2 no sentido horário, e por fim, a Figura 5 (d) mostra o resultado correspondente ao pixel central. Esse procedimento

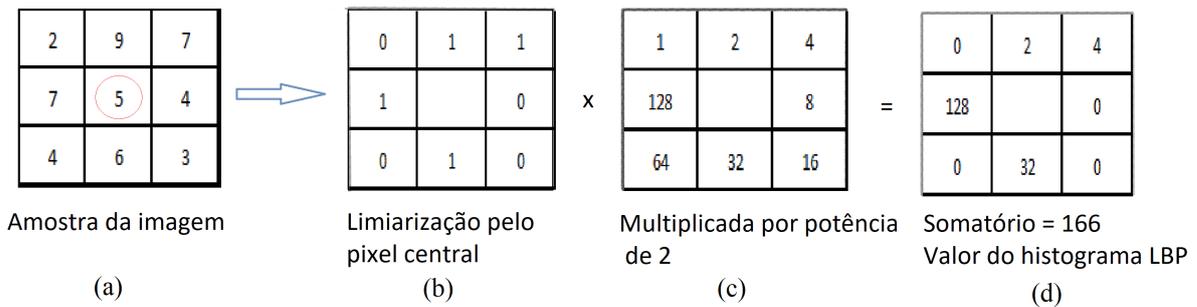
Figura 6 – Exemplos de vizinhança com p e r usados para construir LBP.



Fonte: (MÄENPÄÄ; PIETIKÄINEN, 2005) (2015)

abrange toda a imagem, e a imagem resultante do LBP possui as mesmas dimensões da imagem de entrada.

Figura 5 – Região em análise em (b), limiarização da imagem em (b), multiplicação em potências de 2 em (c) e resultado do pixel central em (d).



Fonte: elaborado pelo autor.

Em Ojala *et al.* (2002), foi proposta uma formulação mais geral baseada em vizinhanças circulares e a interpolação bilinear de valores de coordenadas de *pixel* não inteiro, que permitem qualquer raio e número de *pixels* na vizinhança.

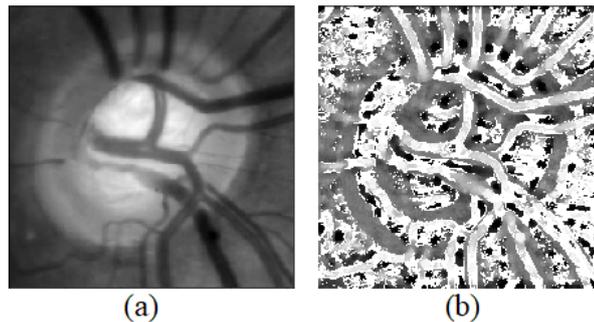
Para considerar os tamanhos de vizinhança variável, dois parâmetros são introduzidos: o número de pontos P em uma vizinhança circular simétrica a considerar (removendo assim a dependência de uma vizinhança quadrada); e o raio do círculo R , que permite contabilizar diferentes escalas. Exemplos de vizinhanças estão mostrados na Figura 6.

Na construção dos valores de LBP, inicialmente, uma imagem é convertida em escala de cinza, a qual é dividida em blocos de matrizes. Para cada matriz é utilizado um ponto central, e para cada *pixel* é selecionada uma vizinhança de tamanho R , em torno desse ponto central. A partir desse ponto central são comparados os *pixels* vizinhos: se o *pixel* vizinho for maior ou igual do que o valor do *pixel* central, então é atribuído o valor 1, se for menor é atribuído 0. Um valor LBP é calculado para esse *pixel* central e armazenado em uma matriz 2D nas mesmas

dimensões da imagem de entrada. Ao final, é gerado uma matriz de zeros e uns. Essa matriz é multiplicada pela matriz ortogonal de *pixels*.

Posteriormente, é realizada a soma dos valores da nova matriz, resultando no valor do histograma LBP correspondente ao *pixel* central. Este processo abrange toda a imagem, resultando assim em uma imagem com as mesmas dimensões de entrada, conforme Figura 7. Para utilizar o LBP como descritor de textura, obtém-se os diferentes níveis de cinza que são as informações de textura da imagem, a quantidade de características resultantes é igual a $P + 2$, em que P é a quantidade de níveis de cinza da imagem LBP resultante. A entrada LBP adicionada tabula todos os padrões que não são uniformes, adicionando um nível extra de rotação e invariância em escala de cinza. Com isso, é obtido o vetor de características com dimensão $P + 2$, que representa as texturas da imagem de entrada.

Figura 7 – Imagem original em (a) e representação do LBP em (b).



Fonte: elaborado pelo autor.

2.3.3 Momentos de Zernike

Em 1934, Frits Zernike desenvolveu um sistema de equações matemáticas que definem superfícies geométricas para descrever aberrações ópticas, chamados de polinômios de Zernike, em que cada polinômio representa uma determinada forma geométrica. Esses polinômios, são complexos e possuem um modo eficiente de representar as alterações de forma (JANKOV *et al.*, 2002).

Os momentos de Zernike foram introduzidos em análise de imagens na década de 1980 por Teague, sendo definidos como projeções da função de intensidade da imagem $f(x,y)$ sobre funções de bases ortogonais, no caso, os polinômios de Zernike, os quais formam um conjunto ortogonal complexo no interior de um círculo unitário $(x^2 + y^2) \leq 1$ (IMADA, 2014). As expressões para o desenvolvimento dos momentos de Zernike bidimensional A_{mn} são

desenvolvidas por (SHUTLER, 2015)

$$A_{mn} = \frac{m+1}{\pi} \int_x \int_y f(x,y) [V_{mn}(x,y)]^* dx dy$$

nos quais $m = 0, 1, 2, \dots, \infty$ define a ordem, e $*$ denota o complexo conjugado. Neste caso, n é um inteiro (positivo ou negativo) que representa a rotação, sujeito às condições

$$|n| \leq m, \quad A_{mn}^* = A_{m,-n}.$$

Considerando uma imagem digital, a expressão anterior é dada sob a forma discreta, se P_{xy} for o *pixel* atual, a equação se torna (IMADA, 2014):

$$A_{mn} = \frac{m+1}{\pi} \sum_x \sum_y P_{xy} [V_{mn}(x,y)]^* \quad \text{em que } x^2 + y^2 \leq 1.$$

Para calcular os momentos de Zernike, a imagem (ou região de interesse - ROI) é primeiro mapeada para um disco de raio R, cujo seu centro é a origem. Os *pixels* que estão fora do disco da unidade não são usados no cálculo. No mapeamento das coordenadas cartesianas, o valor de \tan^{-1} é definido ao longo do intervalo $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$, a invariância da escala e da translação pode ser obtida pela normalização da imagem usando os momentos cartesianos antes do cálculo dos momentos de Zernike. A invariância de escala é produzida alterando cada objeto para que sua área (ou contagem de *pixels* para uma imagem binária) seja $m_{00} = \beta$, no qual β é um valor predeterminado. Ambas as propriedades de invariância (para uma imagem binária) podem ser obtidas usando (SHUTLER, 2015)

$$h(x,y) = f\left(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}\right), \quad \text{em que } a = \sqrt{\frac{\beta}{m_{00}}},$$

e $h(x,y)$ é a nova função dimensionada. O erro envolvido na implementação discreta pode ser reduzido por interpolação. Se a coordenada calculada por essa equação não coincidir com uma localização real da grade, o valor de cada *pixel* associado é interpolado pelos quatro *pixels* adjacentes. Como resultado da normalização, os momentos de Zernike $|A_{00}|$ e $|A_{11}|$ são definidos com valores conhecidos, sendo o valor de $|A_{11}|$ igual a zero, devido à translação da forma para o centro do sistema de coordenadas.

Além disso, o valor absoluto de um momento de Zernike é invariante à rotação, conforme refletido no mapeamento da imagem para o disco. A rotação da forma deste disco

expressa como uma mudança de fase, se ϕ é o ângulo de rotação, A_{mn}^R é o momento de Zernike da imagem rotacionada e A_{mn} é o momento de Zernike da imagem original, então $A_{mn}^R = A_{mn} \exp(-jn\phi)$ (SHUTLER, 2015).

Os momentos são medidas estatísticas projetadas para obter informações globais e geométricas sobre uma imagem. Os momentos exploram informações em uma imagem inteira, em vez de fornecerem informações apenas em um único ponto limite, podem obter algumas das propriedades globais ausentes das representações baseadas em limites puros, como a orientação geral da imagem (RAO *et al.*, 2013).

Como os polinômios de Zernike são ortogonais entre si, os momentos de Zernike podem representar propriedades de uma imagem sem redundância ou sobreposição de informações entre os momentos. Embora os momentos de Zernike sejam significativamente dependentes do dimensionamento e da tradução do objeto em uma região de interesse (ROI), suas magnitudes são independentes do ângulo de rotação do objeto (TAHMASBI *et al.*, 2010).

2.3.4 Filtros de Gabor

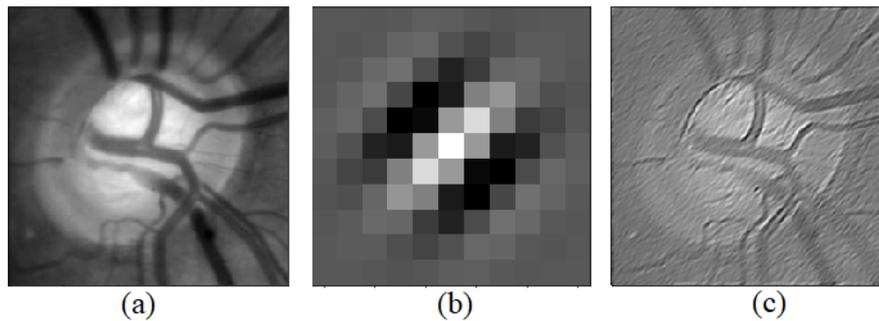
O filtro de Gabor foi desenvolvido em 1964 por Dennis Gabor, é um filtro linear usado no processamento de imagens para análise de textura. Este filtro analisa se há algum conteúdo de frequência específico na imagem em determinadas direções em uma região, e são considerados apropriados para representação e discriminação de texturas (DERPANIS, 2007).

Os filtros de Gabor são filtros passa-banda, ou seja, permite a passagem de frequência de certas faixas e rejeita (atenua) as frequências fora dessa faixa. São usados em análise de textura para detecção de bordas de objetos em uma imagem. Quando um filtro Gabor é aplicado a uma imagem, produz a maior resposta nas bordas e nos pontos em que a textura é alterada (MURRTHY, 2014).

A transformada de Fourier da resposta ao impulso de um filtro Gabor é a convolução da transformada de Fourier da função harmônica (função sinusoidal) com a transformada de Fourier da função Gaussiana. Este filtro tem uma componente real e uma imaginária, representando as direções ortogonais que podem ser compostos em um número complexo ou usado individualmente (HENRIKSEN, 2007).

Um exemplo de aplicação de um filtro de Gabor com ângulo igual a 3 está exposto na Figura 8 (b), e a imagem resultante é mostrada na Figura 8 (c).

Figura 8 – Imagem do fundo do olho em (a), filtro de Gabor na frequência em (b), e imagem após a aplicação do filtro com ângulo 3 em (c).



Fonte: elaborada pelo autor.

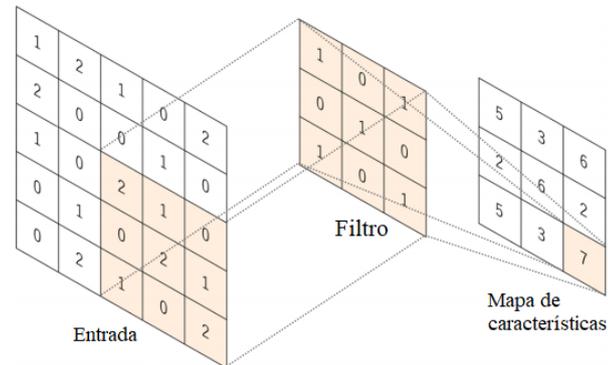
2.4 Redes Neurais Convolucionais

A Rede Neural Convolutiva (CNN) é um algoritmo de aprendizado profundo que pode captar uma imagem de entrada, atribuir pesos e vieses que podem ser aprendidos a vários aspectos da imagem e ser capaz de diferenciar um do outro. O pré-processamento exigido em uma CNN é muito menor em comparação com outros algoritmos de classificação. Enquanto nos métodos primitivos os filtros são feitos à mão, com treinamento suficiente, as CNNs têm a capacidade de aprender esses filtros (DLB, 2019).

Uma CNN é capaz de capturar com sucesso as dependências espaciais e temporais em uma imagem através da aplicação de filtros relevantes. A arquitetura executa um melhor ajuste ao conjunto de dados da imagem devido à redução no número de parâmetros envolvidos e à capacidade de reutilização dos pesos. Em outras palavras, a rede pode ser treinada para entender melhor a sofisticação da imagem.

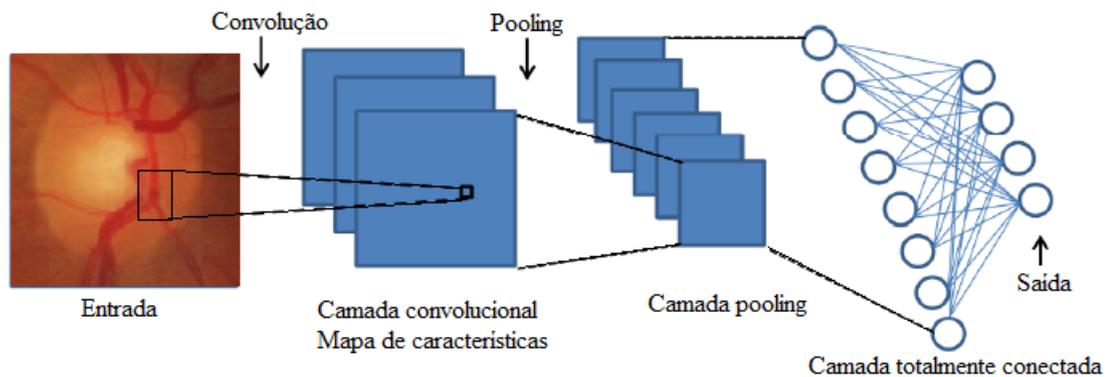
Em uma rede neural convencional os neurônios de entrada conectam-se com todos os neurônios das camadas ocultas. Ao final tem-se uma última camada totalmente conectada que gera o vetor de neurônios de saída. As CNNs funcionam de forma diferente. Como a informação de entrada é uma imagem, os neurônios dessa arquitetura são modificados para possuírem três dimensões: largura, comprimento e profundidade. Outra diferença é que os neurônios de uma camada não são totalmente conectados aos da camada seguinte, mas apenas à uma parte deles. Outra característica bastante importante das CNNs são os tipos de camadas que formam sua arquitetura, essas são: convolução, subamostragem e totalmente conectada (CRUZ A. C. da; SCHWAB, 2019). A Figura 9 é uma representação genérica de uma arquitetura CNN.

Figura 10 – Operação de convolução com filtro 3×3 .



Fonte: Adaptado de Yamashita *et al.* (2018).

Figura 9 – Arquitetura genérica de uma rede neural convolucional.



Fonte: Adaptado de Faria (2018).

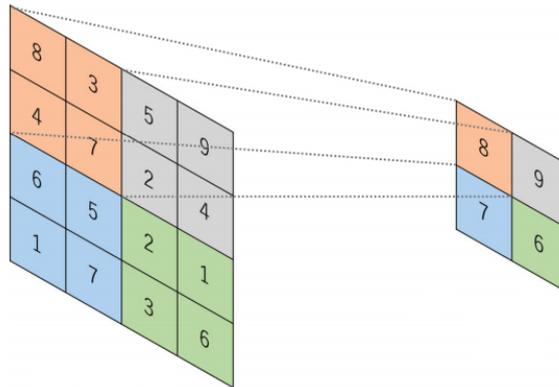
A camada convolucional usa filtro de aprendizado e cada filtro responde a um subconjunto ou uma pequena região do conjunto de entrada (um pequeno quadrado em uma imagem poderia ser considerado como uma pequena região) e as pequenas regiões coletivamente cobrem o conjunto de entrada como um todo. Ou seja, para cada um destes filtros, um neurônio está conectado a apenas um subconjunto dos neurônios na camada anterior. Os filtros são gerados a partir de uma pequena região denominada campo receptivo local e o resultado da aplicação destes filtros são denominados mapas de características (FARIA, 2018). Um exemplo de operação de convolução com filtro de tamanho 3×3 é visto na Figura 10.

As saídas de uma operação linear como convolução são passadas através de uma função de ativação não linear. Embora as funções não-lineares suaves, como a função sigmóide ou tangente hiperbólica, tenham sido usadas anteriormente porque são representações matemáticas de um comportamento biológico dos neurônios, a função de ativação não-linear mais comum usada atualmente é a unidade linear retificada (ReLU), que simplesmente calcula a função:

$f(x) = \max(0, x)$ (YAMASHITA *et al.*, 2018).

A camada de subamostragem tem o objetivo de reduzir dimensionalmente a imagem de saída, isto é utilizado tanto para reduzir o esforço computacional, quanto para marcar as características mais acentuadas da imagem de entrada. É comum que essa camada seja implementada após algumas convoluções dentro de uma rede profunda. Existem, basicamente, dois tipos de camada Subamostragem: *Max Pooling* e *Average Pooling*. O primeiro tipo preserva somente o maior valor da área de atuação do filtro, como mostrado na Figura 11, buscando extrair apenas a característica dominante deste quadrante. Já o segundo, calcula uma média dos valores da região de atuação do filtro (CRUZ A. C. da; SCHWAB, 2019).

Figura 11 – Operação de subamostragem com kernel 2×2



Fonte: Adaptado de Yamashita *et al.* (2018).

E por fim, a camada totalmente conectada, a qual usa as características (ou mapas de ativação) de mais alto nível que são geradas a partir das camadas anteriores (camadas convolucionais e de subamostragem) para realizar a classificação dos dados de entrada em várias classes baseada em um conjunto de treinamento (FARIA, 2018).

Contudo, construir redes convolucionais a partir de pesos iniciais aleatórios não é uma tarefa simples, além disso, dependendo da quantidade de dados ou profundidade da rede o treinamento apresenta um custo computacional elevado. Para contornar esse problema, modelos treinados anteriormente em um situação específica podem ser utilizados em outro problema, esse tipo de aprendizado é conhecido como aprendizado por transferência, *transfer learning*.

2.4.1 *Aprendizado por transferência*

O aprendizado por transferência é uma técnica que permite uma rede neural, que é usada em determinada tarefa, seja aplicada em outra tarefa. Com o aprendizado por transferência, uma CNN comumente usada para reconhecimento de objetos em imagens, é ajustada para treinar e classificar o objeto de interesse em uma outra imagem.

Em modelos de redes convolucionais, as redes aprendem filtros que identificam informações relevantes sobre a imagem de entrada. As camadas iniciais são capazes de detectar informações como bordas, contornos, texturas, informações de gradiente, entre outras. À medida que a rede se torna maior, as camadas mais profundas conseguem captar as informações mais específicas da imagem. Esses modelos pré-treinados podem ser eficientes porque as redes pré-treinadas aprenderam pesos relevantes anteriormente, isto por terem sido treinadas com outras bases de imagens.

As arquiteturas de modelos pré-treinados são obtidas e usadas como modelos de extração de atributos. A saída de uma camada anterior à camada de saída, é usada como entrada para um novo classificador. O Modelo CNN funciona como uma inicialização de pesos, em que o modelo pré-treinado, ou parte do mesmo, é integrado a um novo modelo (BROWNLEE, 2019b).

Utilizar a aprendizagem por transferência é uma alternativa viável, pois, é complexo obter um conjunto de dados grande o suficiente, portanto, treinar um conjunto de dados usando aprendizagem por transferência é atraente pelo seu custo computacional reduzido em comparação às CNNs comuns.

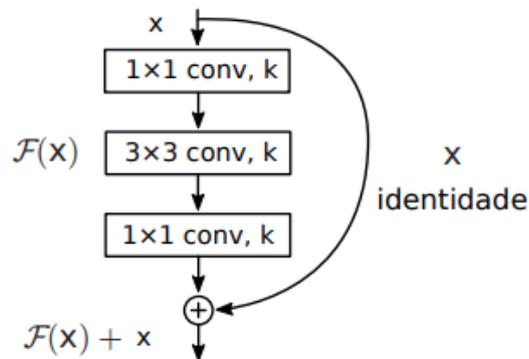
Os modelos CNNs pré-treinados que são utilizados para a análise das imagens desta dissertação são descritos a seguir.

2.4.1.1 *ResNet50*

As Redes neurais residuais (ResNet) foram desenvolvidas pelos pesquisadores da *Microsoft Research*. O principal aspecto que a torna diferente das demais arquiteturas de CNNs é devido à aplicação do conceito de bloco residual, o qual utiliza atalhos entre as camadas, adicionando os valores iniciais de entrada da camada a função ReLU de saída $y = F(x, W_i) + x$, em que a função $F(x, W_i)$ representa o mapa residual aprendido pelo bloco e x a imagem inicial (POMARI *et al.*, 2018).

As camadas em uma rede tradicional estão aprendendo a verdadeira saída ($H(x)$) enquanto as camadas em uma rede residual estão aprendendo a residual ($R(x)$). Essas conexões de atalho pulam uma ou mais camadas e estão dispostas em forma de blocos, chamados de bloco residual. Esses blocos possuem sempre 3 camadas de convolução com k filtros 1×1 , 3×3 e 1×1 respectivamente para cada camada. Dada uma entrada x , o mapeamento residual $F(x)$ é denotado por $F(x) + x$, ou seja, esse mapeamento soma a entrada de um bloco residual (x identidade) a saída resultante desse mesmo bloco (BEZERRA, 2019). O funcionamento de um bloco residual é representado pela Figura 12.

Figura 12 – Modelo de um bloco residual.



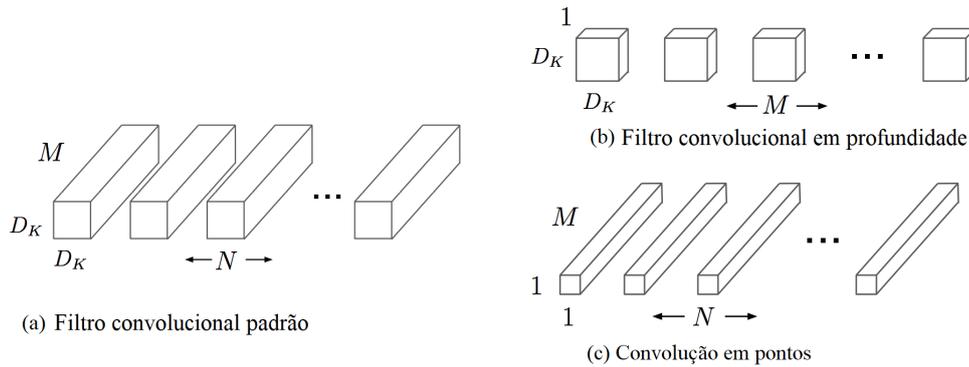
Fonte: Adaptado de He *et al.* (2016).

Essa soma é realizada enquanto os mapas de características resultantes das camadas de convolução possuem a mesma resolução. Entretanto, quando as resoluções dos mapas de características são diferentes, essa soma ocorre com preenchimento extra de zero para aumentar a resolução dos mapas menores. Essas conexões de atalho não adicionam nenhum parâmetro extra na complexidade computacional do modelo (HE *et al.*, 2016).

A ResNet propôs uma solução para o problema do gradiente de fuga. As redes neurais treinam via retropropagação, que depende da descida do gradiente para obter os pesos ideais que minimizam a função de perda. Quando mais camadas são adicionadas, a multiplicação repetida de suas derivadas eventualmente torna o gradiente infinitesimalmente pequeno, o que significa que camadas adicionais não melhoram o desempenho ou podem até reduzi-lo. Devido a essas conexões de salto, é possível propagar gradientes maiores para as camadas iniciais e essas camadas também podem aprender o mais rápido possível, permitindo assim treinar redes mais profundas (SAHOO, 2018).

Em vez de tratar o número de camadas de um hiperparâmetro importante para ajustar, adicionando conexões de salto à rede, permite que a rede salte o treinamento para as camadas

Figura 13 – Filtro convolucional padrão em (a), convolução em profundidade em (b) e convolução em ponto (c).



Fonte: Adaptado de Howard *et al.* (2017).

que não são úteis e não agregam valor com precisão geral. De certa forma, pular conexões torna nossas redes neurais dinâmicas, para que possa otimizar o número de camadas durante o treinamento (SAHOO, 2018).

Os tamanhos das saídas e núcleos convolucionais da rede ResNet para as camadas 18, 34, 50, 101 e 152 podem ser vistos na Figura 39. Para esta dissertação é utilizada a ResNet com 50 camadas, essa quantidade refere-se aos blocos residuais, portanto, 50 blocos residuais.

2.4.1.2 MobileNet

A arquitetura MobileNet é uma CNN que foi proposta por pesquisadores do Google, desenvolvida para estar presente em aplicações móveis. Também é baseada no conceito de *depthwise separable convolutions*, o que torna a arquitetura simples, reduzindo a necessidade de um alto poder computacional (POMARI *et al.*, 2018).

Um das características mais importantes da rede MobileNet são suas convoluções separáveis em profundidade, no qual dividem um núcleo em 2 núcleos separados que executam duas convoluções: a convolução em profundidade e a convolução em ponto. Com menos multiplicações, a complexidade computacional diminui e a rede é capaz de rodar mais rapidamente (WANG, 2018). A Figura 13 exemplifica o funcionamento de uma MobileNet, em que os filtros convolucionais padrão são substituídos por convolução em profundidade e convolução em ponto para criar um filtro separável em profundidade.

A arquitetura da rede convolucional MobileNet é descrita Tabela 13, Anexo A. A camada de Conv/ s2 informa que a rede inicia uma camada convolucional com preenchimentos de zeros, seguida de uma camada convolucional comum, normalização em lotes e função de

ativação ReLU. A camada Conv/s1 indica que tem apenas uma cada de convolução, normalização em lotes e função de ativação ReLU. A Conv dw informa que a camada é convolucional em profundidade.

A convolução com preenchimento de zeros contorna a matriz de entrada com zeros na borda. Essa técnica calcula e adiciona o preenchimento necessário à imagem de entrada (ou mapa de recursos) para garantir que a saída tenha o mesmo formato da entrada. Os valores de pixel adicionados podem ter o valor zero que não afeta a operação do produto escalar quando o filtro é aplicado (BROWNLEE, 2019a).

A normalização em lotes normaliza a saída de uma camada de ativação anterior, da qual subtrai a média do lote e divide por seu desvio padrão. Este processo aumenta a estabilidade de uma rede neural. Este método reduz a adaptação excessiva, porque apresenta alguns efeitos de regularização. Adiciona algum ruído às ativações de cada camada oculta, portanto, ocorrem menos desistências, o que é bom porque não se perdem muitas informações. No entanto, não se deve depender apenas da normalização de lotes para regularização (PECCIA, 2017).

2.4.1.3 VGG

A arquitetura VGG Net, criada pelo *Visual Geometry Group*, consiste em uma versão mais atual da arquitetura conhecida como AlexNet, suas principais variações são a arquitetura VGG16 (com um total de 16 camadas) e a VGG19 (com um total de 19 camadas). As duas são compostas por seis blocos com diversas camadas, sendo que os primeiros blocos são compostos por uma combinação de camadas de convolução e subamostragem, e o sexto, um bloco inteiramente conectado (POMARI *et al.*, 2018).

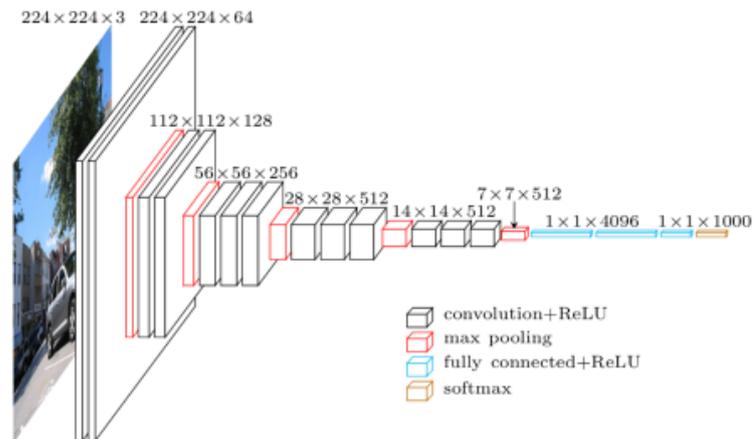
A VGG16 é uma arquitetura de uma CNN que foi usada para vencer a competição ILSVR (ImageNet) em 2014. Considerada uma das excelentes arquiteturas de modelos de visão até hoje. Esta é uma rede grande e possui aproximadamente 138 milhões de parâmetros (HASSAN, 2018).

Suas camadas são compostas por camadas convolucionais, camadas de pool máximo, camadas de ativação e camadas totalmente conectadas. Existem 13 camadas convolucionais, 5 camadas *Max Pooling* e 3 camadas densas, que somam 21 camadas, mas apenas 16 camadas de peso (TEWARI, 2019). Uma ilustração dessa arquitetura é visualizada na Figura 14.

Na arquitetura da VGG16, o número 16 refere-se às 16 camadas que têm pesos, o primeiro bloco de convolução possui duas camadas com 64 neurônios, os quais atuam como

filtros convolucionais de 3×3 , sempre seguidos por uma camada de *max pooling* 2×2 . Os demais blocos (2, 3, 4 e 5) seguem uma composição similar, mas aumentando o número de filtros por bloco (128, 256, 512, 512). O sexto bloco possui três camadas totalmente conectadas, em que as duas primeiras possuem 4096 neurônios e a terceira é uma camada de saída com 1000 neurônios com ativação *softmax* (TEWARI, 2019).

Figura 14 – Arquitetura VGG16.



Fonte: Adaptado de Tewari (2019).

A diferença entre as arquiteturas VGG16 e VGG19 se resume a três camadas adicionais existentes na VGG19, sendo uma camada a mais no quarto, uma no quinto e uma no sexto bloco, assim como mostrado na Tabela 14, Anexo A (POMARI *et al.*, 2018).

2.5 Métodos de Classificação

Classificação é um processo de divisão relacionado à categorização, no qual os grupos são separados de acordo com as características mais similares. Os métodos de classificação utilizados para distinguir as características das imagens e identificar o glaucoma são apresentados a seguir.

2.5.1 Regressão Logística

A Regressão Logística (RL) é um caso particular dos modelos lineares generalizados, sendo um dos mais empregados para a análise de dados categóricos. Esse modelo foi introduzido no âmbito estatístico em torno de 1970 pelo matemático David Cox. Uma das vantagens do uso de regressão logística é a possibilidade de incluir mais de uma variável de natureza qualitativa

e, também, pela facilidade de interpretação dos coeficientes estimados, os quais descrevem a relação das variáveis explicativas com a variável dependente (PAULA G. A. de, 2004).

Um modelo de regressão logística simples descreve apenas uma variável explicativa, na regressão logística múltipla é possível ajustar um modelo para a variável resposta, incluindo mais de uma variável explicativa (covariável).

A função logística, também conhecida como *logit* é a inversa da função sigmóide, é ajustada usando análise de regressão linear, o valor previsto é convertido em probabilidades estimadas através da função exponencial. Apesar da variável dependente ser categórica, a regressão logística estima as probabilidades. Desse modo, considerando um conjunto de p variáveis independentes X_1, X_2, \dots, X_p , a função de ligação *logit* é dada por:

$$g(X) = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 X_1 + \beta_1 X_2 + \dots + \beta_p X_p$$

e a probabilidade associada aos valores de Y_i , que é variável dependente, também conhecida como variável resposta, pode ser expressa como a função de distribuição acumulada de uma distribuição logística, representada pela expressão:

$$E(Y_i) = \pi_i = \frac{e^{g(X)}}{1 + e^{g(X)}}.$$

Os coeficientes $\beta_0, \beta_1, \dots, \beta_p$ são estimados pelo método da máxima verossimilhança, o qual encontra uma combinação de coeficientes que maximiza a probabilidade da amostra ter sido observada. Considerando uma certa combinação de coeficientes $\beta_0, \beta_1, \dots + \beta_p$ e variando os valores de X , observa-se que a curva logística tem um comportamento probabilístico no formato de uma sigmóide, o qual é uma característica da regressão logística (HOSMER *et al.*, 1989).

A regressão logística é um método de aprendizado de máquinas supervisionado. Os coeficientes do algoritmo de regressão logística devem ser estimados, a partir dos seus dados de treinamento. Isso é realizado usando a estimativa de probabilidade máxima. Esse modelo de regressão prever probabilidades que variam entre 0 e 1, portanto, também é empregado como método de classificação (BROWNLEE, 2016).

Alguns hiperparâmetros podem ser definidos para aprimorar um modelo, como a inclusão de um método de penalidade na estimação dos coeficientes. A regularização L1 é chamada de Regressão de Lasso, que estima os coeficientes pelo método dos mínimos absolutos. Neste caso, é adicionado um termo de regularização para impedir que os coeficientes se ajustem

tão perfeitamente, ocasionando o excesso de ajuste. A regularização adiciona o valor absoluto da magnitude do coeficiente à função de perda, que é basicamente a soma das diferenças absolutas entre o valor real e os valores estimados, como mostrado na expressão

$$SQE_{L1} = \sum_{i=1}^n (y_i - \sum_{j=1}^p X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|,$$

em que SQE_{L1} é a soma dos quadrados dos resíduos da regularização L1, e λ é o hiperparâmetro associado ao termo de regularização, se λ for zero não haverá regularização, a estimação retorna para o método de mínimos quadrados ordinários. Enquanto valores menores especificam uma regularização mais forte, à medida que a constante λ aumenta, mais coeficientes se tornam nulos. Essa técnica é bastante útil quando se tem uma grande quantidade de dados, pois, reduz as variáveis menos significativas para o modelo.

A regularização L2 é conhecida como regressão de *Ridge*, que adiciona a magnitude ao quadrado do coeficiente como termo de penalidade à função de perda (NAGPAL, 2017),

$$SQE_{L2} = \sum_{i=1}^n (y_i - \sum_{j=1}^p X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2,$$

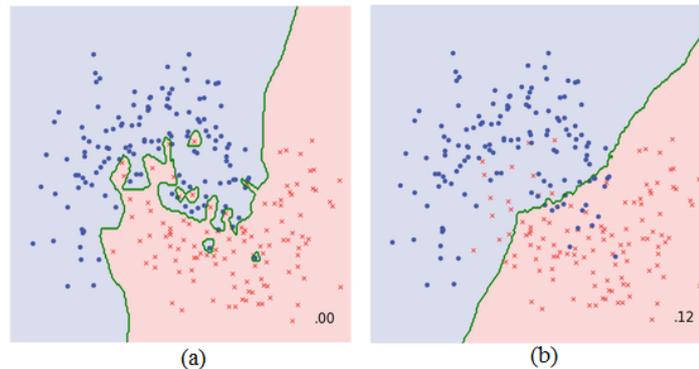
em que SQE_{L2} é a soma dos quadrados dos resíduos da regularização L2. Se o valor de λ for muito grande, aumenta seu peso e produz um desequilíbrio, ocasionando um ajuste insuficiente. Desse modo, é importante a escolha adequada para o valor de λ . Essa técnica funciona muito bem para evitar problemas de ajuste excessivo.

2.5.2 *k - Vizinhos Mais Próximos*

Em 1951, Fix e Hodges introduziram um método não paramétrico para classificação de padrões e, desde então, se tornou conhecida como a regra dos *k* vizinhos mais próximos, *k-Nearest Neighbors* (k-NN). Em 1967, Cover e Hart demonstraram que, para $k = 1$ e $n \rightarrow \infty$, o erro de classificação k-NN é limitado acima por duas vezes a taxa de erro de Bayes. O *k-Nearest Neighbors* (k-NN) é um algoritmo de aprendizado supervisionado, sendo uma das técnicas mais simples para classificação e regressão (PETERSON, 2009).

O k-NN é um tipo de aprendizado baseado em instância, ou aprendizado lento, em que a função armazena os exemplos de treinamento e a generalização é feita somente quando uma nova instância tiver que ser classificada. O classificador k-NN é comumente baseado na

Figura 15 – Exemplo de vizinhança para o algoritmo k-NN.



Fonte: (RAVARANI, 2016)

distância euclidiana entre uma amostra de teste e as amostras de treinamento especificadas (PETERSON, 2009).

Quando o k-NN é usado para classificação binária, a saída está associada a uma das classes. Um objeto é classificado com base nos seus vizinhos, em que lhe é atribuído a classe representada mais frequentemente dentre as k amostras mais próximas e utilizando um esquema de votação, sendo k o número inteiro positivo, geralmente menor que 10. Se $k = 1$, o objeto é simplesmente atribuído à classe do único vizinho mais próximo.

A escolha para o valor de k depende dos dados, normalmente é escolhido um valor ímpar para evitar empates na votação e escolha de classes. Se k for muito pequeno, a classificação fica sensível a pontos de ruído; e se k for muito grande, a vizinhança pode incluir elementos de outras classes. Um exemplo de separação entre os grupos de acordo com a escolha de k é visualizado na Figura 15 (a), na qual representa um exemplo de separação com $k=1$, sendo possível verificar que os dados estão superajustados aos dados de treinamento, e na Figura 15 (b) é visto um valor de $k=7$, que representa uma separação adequada entre os grupos para o exemplo em questão.

De acordo com Srivastava (2018), as etapas para implementar o algoritmo k-NN são descritas da seguinte forma:

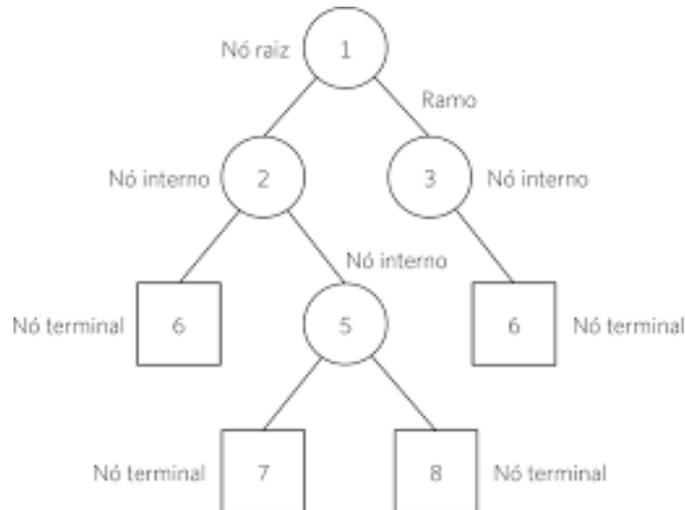
1. para cada novo vetor de características desconhecidas (ainda não classificado), realizar uma busca pelo índice do vetor de atributos mais próximo $i^* = \arg_{i=1, \dots, N} \min \|x - x_i\|^2$,
2. atribuir a x a mesma classe que x_i se $\|x - x_i^*\|^2 < \|x - x_i\|^2, \forall i \neq i^*$.

Este algoritmo é comumente utilizado por sua fácil interpretação e baixo tempo de cálculo. Outra vantagem é a habilidade para modelar funções complexas por uma coleção de aproximações locais menos complexas (SRIVASTAVA, 2018).

2.5.3 Árvore de Decisão com Aumento de Gradiente

A árvore de decisão, *Decision Tree* (DT), é um classificador que contém algoritmos com instruções de controle condicional, em que cada nó interno representa a probabilidade de ocorrência da característica de interesse em determinada classe. Os caminhos da raiz para a folha representam regras de classificação. O primeiro nó é chamado de nó raiz ou nó principal, do qual se originam os nós filhos ou internos. Os ramos são os valores do atributo de saída de um nó que se conectam ao nó filho, e as folhas ou nó terminais são as regiões associadas a um rótulo de valor (MEDEIROS *et al.*, 2014). A Figura 16 é uma representação genérica do funcionamento de uma árvore de decisão.

Figura 16 – Representação genérica de uma árvore de decisão.



Fonte: (MEDEIROS *et al.*, 2014)

A árvore de decisão é um método de classificação ou regressão que particiona o espaço de características em porções disjuntas e, então, classificam os dados ou ajustam um modelo simples (uma constante, por exemplo) de acordo com a partição à qual pertencem. São modelos simples, mas poderosos (SRIVASTAVA, 2019).

Estas árvores são caracterizadas por um viés baixo, no entanto, apresentam alta variação, mesmo com pequenas alterações na amostra de treinamento. Alguns métodos podem ser usados para reduzir a correlação entre árvores e evitar o ajuste excessivo. A estratégia mais simples é restringir que todas as árvores tenham a mesma profundidade, também é comum definir um número mínimo de observações em cada nó terminal da árvore, evitando assim particionamentos que contenham poucos dados (MAYRINK, 2016).

Crescer uma árvore envolve decidir quais recursos escolher e quais condições usar para dividi-las, além disso, saber quando parar. Uma árvore de decisão muito extensa tende a ser muito complexa e superajustada, ou seja, o erro do conjunto de treinamento pode ser baixo, mas alto no conjunto de validação. Uma árvore menor, com menos divisões, pode levar à menor variação e melhor interpretação ao custo de um pequeno viés. Portanto, este parâmetro deve ser definido de acordo com a complexidade do problema. Pode-se evitar o ajuste excessivo adicionando um critério de parada, que auxilia na decisão de continuar dividindo a árvore, ou parar e transformar o vértice atual em uma folha (CHEPENKO, 2013).

O *Gradient Boosting* (GB), ou aumento de gradiente, é uma implementação de máquinas de aumento de gradientes criada por Tianqi Chen, atualmente com contribuições de muitos desenvolvedores. Tianqi Chen fornece uma breve e interessante história sobre a criação do XGBoost no *post Story e Lessons Behind the Evolution of XGBoost* (BROWNLEE, 2016).

O princípio de qualquer algoritmo de *boosting* é a combinação do resultado de muitos classificadores fracos para formar uma espécie de comitê forte de decisão. O algoritmo *boosting* é uma forma de expansão, o qual ajusta dados em uma soma ponderada de funções elementares, dadas por $f(x) = \sum_{m=1}^M \beta_m b(x; \gamma_m)$, em que β_m são os coeficientes da expansão e $b(x; \gamma)$ são funções simples do argumento x , caracterizadas pelo parâmetro γ . O algoritmo de aumento de gradiente usa o método de descida de gradiente para otimizar a função de perda, normalmente, se minimiza o valor esperado (sobre o conjunto de treino) de alguma perda, ajustando o modelo para obter o melhor desempenho evitando mínimos locais $\min_{\beta_m, \gamma_m} \sum_{i=1}^N L(y_i, \sum_{m=1}^M \beta_m b(x_i; \gamma_m))$ (SRIVASTAVA, 2019).

O GB constrói árvores, uma de cada vez, em que cada nova árvore auxilia na correção dos erros cometidos por árvores previamente treinadas.

Em geral, existem três parâmetros que caracterizam uma árvore de decisão aprimorada por gradiente (GBDT): número de árvores, profundidade das árvores e taxa de aprendizado. Estes modelos são de fácil implementação e produzem resultados acima daqueles obtidos com técnicas tradicionais, melhoram o desempenho do algoritmo, reduzindo o *overfitting* (BROWNLEE, 2016).

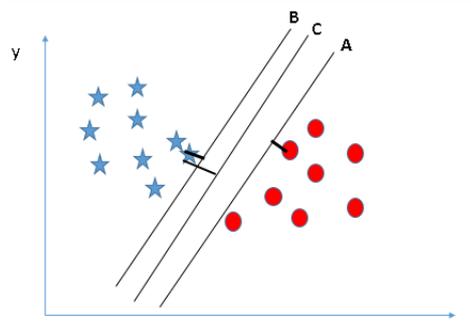
2.5.4 Máquina de Suporte Vetorial

A Máquina de Suporte Vetorial, *Support Vector Machine* (SVM), foi elaborada por Boser, Guyon e Vapnik em 1992. É um algoritmo de aprendizado supervisionado, o qual tem

por objetivo classificar determinado conjunto de dados, que são mapeados para um espaço de características multidimensional, usando uma função kernel (PORTELA *et al.*, 2017). O intuito desse classificador é encontrar um hiperplano em um espaço N-dimensional que tenha a margem máxima que classifica os pontos de dados, ou seja, uma fronteira que melhor segrega as duas classes (SOUZA, 2019).

É possível separar, por meio de uma linha, o conjunto de objetos em seu respectivo grupo, o que caracteriza um classificador linear, como mostrado na Figura 17, em que é possível separar linearmente o conjunto, de acordo com as duas classes. No entanto, é interessante escolher a reta que melhor maximiza as distâncias (margem) entre os pontos mais próximos, esta escolha melhora a robustez do modelo. Portanto, neste exemplo, o melhor hiperplano é o C, por apresentar margem máxima.

Figura 17 – Separação linear

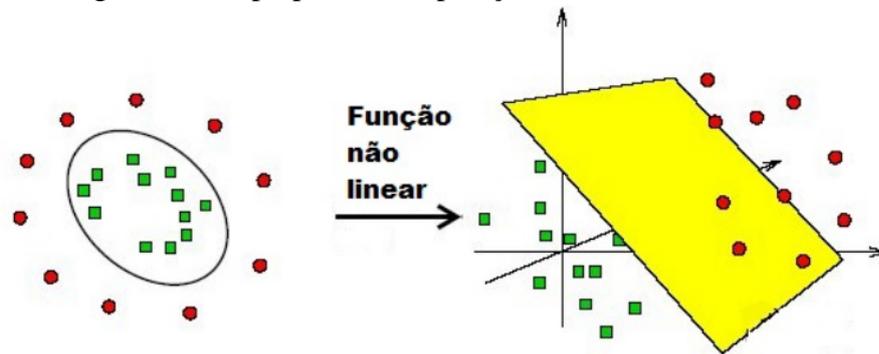


Fonte: (SOUZA, 2019)

Geralmente, aplicações reais envolvendo classificação costumam ser complexas, sendo necessária a utilização de separações complexas. O SVM propõe um método que converte um problema não separável em um problema separável, chamada de truque de kernel. Tal técnica mapeia implicitamente um espaço de entrada de baixa dimensão e o transforma em um espaço de alta dimensão, como mostrado na Figura 18.

No aprendizado de máquinas, a função Kernel mais comum é a *Radial Basis Function* (RBF), em que duas amostras x e x' são representadas por $K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma(\|\mathbf{x} - \mathbf{x}'\|^2))$, em que $\|\mathbf{x} - \mathbf{x}'\|^2$ é a distância ao quadrado entre os dois vetores de características, e $\gamma = \frac{1}{2\sigma^2}$.

Figura 18 – Hiperplano de separação.



Fonte: (PEIXOTO *et al.*, 2016)

Conforme se observa na Figura 18, ao invés de construir uma curva complexa, o SVM propõe separar os grupos por uma linha ótima. Para isso, constrói essa separação em um terceiro eixo z positivo, o qual é obtido através do resultado da soma quadrática dos eixos x e y . Consequentemente, classifica corretamente os grupos sem aumentar a complexidade.

Assim como para outros modelos de classificação, o SVM também possui parâmetros que, quando adicionados no modelo, melhoram seu desempenho. O parâmetro C é uma regularização do SVM, em que relaciona a classificação correta dos dados de treinamento e a maximização da margem da função de decisão, minimizando a norma dos pesos, obtendo assim um menor erro nos dados de treino e teste. Um valor pequeno para C torna a superfície da decisão suave com uma margem maior, desse modo, o classificador aceita os dados classificados incorretamente. Enquanto para valores altos de C , a função aceita uma margem menor de erro quando a função de decisão classifica corretamente os dados de treinamento. Neste caso, o classificador é fortemente penalizado por dados classificados incorretamente e, portanto, ocorre um superajuste nos dados de treino (BONESSO, 2013).

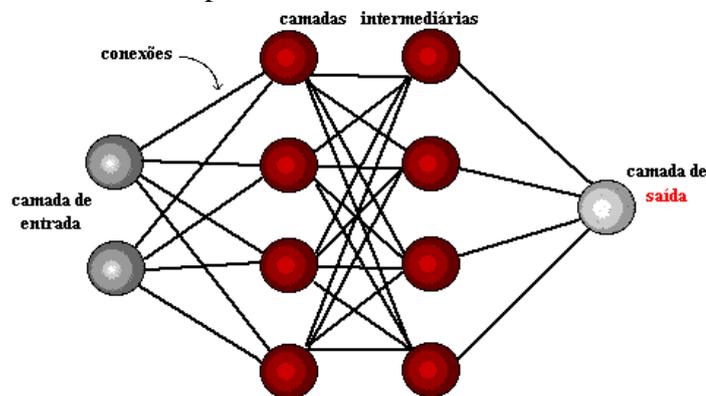
O parâmetro γ pode ser visto como o inverso do raio de influência das amostras selecionadas pelo modelo, se este for alto, o raio da área de influência dos vetores de suporte inclui apenas o próprio vetor de suporte, e nenhuma quantidade de regularização C pode impedir o ajuste excessivo. No caso contrário, se o valor de γ for muito pequeno, o modelo se torna restrito e não consegue identificar a complexidade ou a forma dos dados (BONESSO, 2013).

O SVM possui grande abrangência de aplicações em diversas áreas, como finanças, biologia, medicina, logística, entre outras. Isso ocorre devido às suas vantagens de aplicação, como o bom desempenho de generalização, tratabilidade matemática, interpretação geométrica e a utilização para a exploração de dados não rotulados (PORTELA *et al.*, 2017).

2.5.5 Perceptron Multicamadas

Um Perceptron Multicamadas, Multilayer Perceptron (MLP), é uma rede neural artificial composta por mais de um perceptron. São formados por uma camada de entrada do sinal, uma camada de saída que toma uma decisão ou previsão sobre a entrada e, entre essas duas, um número arbitrário de camadas ocultas que são o verdadeiro mecanismo de classificação do MLP. Modelos com apenas uma única camada oculta são capazes de aproximar qualquer função contínua (DLB, 2019). A Figura 19 ilustra uma arquitetura típica de uma MLP.

Figura 19 – Arquitetura do neurônio da rede neural Perceptron Multicamadas.



Fonte: <https://www.docsity.com/pt/redes-neurais-artificiais-4/4775214/>.

Os MLPs são frequentemente aplicados a problemas de aprendizagem supervisionados: treinam um conjunto de pares entrada-saída e aprendem a modelar a correlação (ou dependências) entre essas entradas e saídas. O treinamento envolve o ajuste dos parâmetros, ou os pesos e viés (*bias*), para minimizar o erro. A retropropagação é usada para fazer os ajustes dos pesos e de viés em relação ao erro, e o próprio erro pode ser medido de várias maneiras, inclusive pelo erro quadrático médio (DLB, 2019).

O algoritmo de aprendizado é baseado na correção do erro, pode ser visto como uma generalização do algoritmo de mínimos quadrados. Basicamente, a aprendizagem por retropropagação consiste em dois passos, um para frente e outro para trás. No passo à frente, o vetor de entrada x_1, x_2, \dots, x_n é aplicado aos nós sensoriais e seu efeito é propagado através da rede, camada por camada, um conjunto de saída é produzido como resposta, cujos pesos sinápticos são fixos. No passo para trás, os pesos sinápticos são ajustados de acordo com a regra de correção do erro, ou seja, a resposta real é subtraída da resposta desejada para produzir um sinal de erro. Isso faz com que os pesos ajustados produzam resultados mais próximos da

resposta desejada (HAYKIN, 2007).

A vantagem de utilizar a retropropagação é devido ao seu poder de aproximação universal, ou seja, é possível aproximar a retropropagação de uma função contínua, sendo um dos algoritmos mais utilizados. Apesar das vantagens, é um algoritmo com custo computacional elevado no treinamento.

Os passos para desenvolver um MLP com algoritmo de aprendizado retropropagação é resumido em: inicialização, ativação, treinamento de pesos e iteração. Na etapa de inicialização são atribuídos valores aleatórios aos pesos e ao limite, esses valores são pequenos e provenientes de uma distribuição uniforme. A função de ativação calcula os valores dos neurônios da camada oculta e da camada de saída. Os pesos dos erros são treinados pelo algoritmo de retropropagação, sendo que o processo usado para a função de ativação e o cálculo dos pesos é repetido até que o critério de erro seja satisfeito (HAYKIN, 2007).

Uma das principais funções de ativação para o MLP é a função de unidade linear retificada (ReLU), a qual retorna $f(x) = \max(0, x)$, com solucionador de otimização de peso ADAM, este refere-se a um otimizador estocástico baseado em gradiente proposto por Kingma, Diederik e Jimmy Ba.

2.6 Análise de Componentes Principais

A Análise de Componentes Principais (PCA) é uma técnica estatística multivariada que converte o conjunto de dados originais em um conjunto de combinações lineares não correlacionados, chamadas de componentes principais. Esta técnica tem o intuito de reduzir a dimensionalidade das variáveis, portanto, as componentes resultantes sempre são menores do que a quantidade de variáveis. Tal método identifica padrões nos dados, evidenciando as similares e eliminando as sobreposições (VASCONCELOS, 2008).

Os componentes principais apresentam propriedades importantes: cada componente principal é uma combinação linear de todas as variáveis originais, são independentes entre si e estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos da variação total contida nos dados. A análise de componentes principais é associada à ideia de redução de massa de dados, com menor perda possível da informação. Procura-se redistribuir a variação observada nos eixos originais de forma a se obter um conjunto de eixos ortogonais não correlacionados (VARELLA, 2008).

Para a construção da PCA, considera-se uma matriz de dados X composta por p

características e n indivíduos, a matriz possui dimensão $n \times p$. Os dados estão dispostos conforme a matriz a seguir,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Em seguida, calcula-se a matriz de covariâncias denotada por S das variáveis da matriz X , assim como mostrado a seguir,

$$S = \begin{bmatrix} \hat{V}ar(x_1) & \hat{C}ov(x_1, x_2) & \cdots & \hat{C}ov(x_1, x_p) \\ \hat{C}ov(x_2, x_1) & \hat{V}ar(x_2) & \cdots & \hat{C}ov(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}ov(x_p, x_1) & \hat{C}ov(x_p, x_2) & \cdots & \hat{V}ar(x_p) \end{bmatrix}$$

Geralmente, as variáveis possuem unidades diferentes, com isso é importante que haja uma padronização para que todas estejam na mesma escala. A padronização comumente utilizada é através da normalização dos dados, em que todos possuirão média zero e variância unitária, dado pela expressão $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s(x_j)}$, em que $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$. Em que \bar{X}_j é a média estimada da variável j e $S(x_j)$ o desvio padrão da variável j . A partir da normalização obtém-se a matriz Z , denotada por

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix}$$

A matriz Z das variáveis padronizadas z_j é igual a matriz de correlação da matriz de dados X . Para determinar os componentes principais normalmente partimos da matriz de correlação R . É importante observar que o resultado encontrado para a análise a partir da matriz S pode ser diferente do resultado encontrado a partir da matriz R (VARELLA, 2008).

Os componentes principais são calculados a partir da expressão a seguir,

$$\det[R - \lambda I] = 0$$

Se a matriz R for de posto completo igual a p , isto é, não apresentar nenhuma coluna que seja combinação linear de outra, a equação terá p raízes chamadas de autovalores ou raízes características da matriz R denotadas por λ , então $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_p$ (VARELLA, 2008).

Para cada autovetor, existe um autovalor correspondente α . Então o i -ésimo componente principal é dado por:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p.$$

Os componentes principais possuem propriedades como:

1. as componentes não são correlacionadas entre si;

$$\hat{Cov}(Y_i, Y_j) = 0$$

2. a variância da componente principal é igual ao autovalor;

$$\hat{Var}(Y_i) = \lambda_i$$

3. o primeiro componente é o que apresenta maior variância, e assim por diante;

$$\hat{Var}(Y_1) > \hat{Var}(Y_2) > \dots > \hat{Var}(Y_p)$$

4. o total de variância das variáveis originais é igual ao somatório dos autovalores que é igual ao total de variância dos componentes principais.

$$\sum_{i=1}^p \lambda_i = \sum_{i=1}^p \hat{Var}(Y_i)$$

A importância de uma componente principal é avaliada por meio de sua contribuição, isto é, pela proporção de variância total explicada pelo componente. A soma dos primeiros k autovalores representa a proporção de informação retida na redução de p para k dimensões (VARELLA, 2008).

A contribuição C_i de cada componente principal Y_i é expressa em porcentagem. Essa contribuição é calculada dividindo a variância da componente Y_i pela variância total, como descrito na expressão a seguir:

$$C_i = \frac{\hat{Var}(Y_i)}{\sum_{i=1}^p \hat{Var}(Y_p)} \times 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100$$

De forma matemática, a PCA é definida como uma transformação linear ortogonal que transforma o conjunto de dados em um novo conjunto, de tal maneira que a maior variância

por qualquer projeção dos dados ocupada a primeira coluna, e assim sucessivamente. Desse modo, as primeiras componentes contêm informações suficientes sobre os dados, portanto, é possível descartar, sem perda de informação, parte das componentes (WOLD *et al.*, 1987).

2.7 Otimização dos Hiperparâmetros

Os hiperparâmetros são variáveis que configuram um modelo, geralmente, permanecem constante durante o treinamento dos dados. Os valores dos hiperparâmetros são adicionados ao modelo antes de iniciar o processo de aprendizagem, enquanto os valores dos parâmetros são otimizados pelo processo de treinamento.

Conforme ALVARENGA JÚNIOR, W. J. de (2018), o uso de hiperparâmetros com valores ótimos possibilita um modelo com melhor desempenho, evitando assim, problemas de superajuste (*overfitting*), modelos simplistas com baixa capacidade, dentre outros. Os hiperparâmetros influenciam diretamente nos resultados obtidos nos modelos de aprendizado de máquinas.

Como as configurações dos modelos mudam de acordo com cada conjunto de dados, é importante utilizar técnicas que otimizem os hiperparâmetros, pois, a escolha desses valores afeta diretamente os resultados. Uma das técnicas tradicionalmente utilizada para otimização dos hiperparâmetros é a pesquisa em grade.

Este método faz uma busca exaustiva no intervalo de hiperparâmetros especificado, ou seja, para cada hiperparâmetro é especificado um conjunto de valores candidatos que devem ser postos em um modelo e avaliados para obter a melhor configuração para os dados em questão. O problema da utilização do método de pesquisa em grade reside no crescimento exponencial de treinamentos necessários, à medida que o número de hiperparâmetros a serem estimados aumenta. Isto resulta em um tempo relativamente longo de execução de treinamento (ALVARENGA JÚNIOR, W. J. de, 2018).

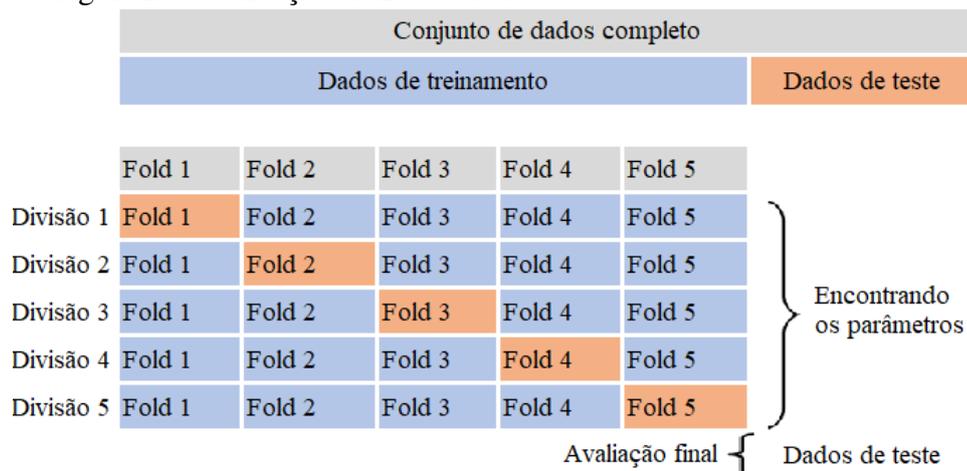
2.7.1 Validação Cruzada

A técnica de Validação Cruzada, Cross-Validation (CV), tem por objetivo avaliar o ajuste de um modelo em diferentes subconjuntos de dados, verificando se o modelo está adequado com o nível de erro aceitável. Inicialmente, o conjunto de dados é particionado em dados de treino e dados de teste. O conjunto de treino, então, é subdividido em conjunto de

estimativa e outro de validação, sendo que o primeiro, usado para treinar o modelo e obter os parâmetros que são aplicados ao conjunto de validação, já os dados de teste avaliam como deve ser o comportamento desse modelo em um conjunto desconhecido.

Dentre as técnicas de validação cruzada, a *k-fold* é a mais conhecida e utilizada, geralmente, a quantidade de particionamentos é menor ou igual a 10. Essa técnica consiste em particionar o conjunto de treino em k partes iguais, em que $k - 1$ dessas partes são usadas para treinar o modelo e a parte restante serve como um conjunto teste para avaliar o modelo obtido. Esse procedimento é repetido k vezes e, a cada uma das vezes, seleciona um subgrupo de $k-1$ para treino e um diferente para teste, como mostrado na Figura 20. Após as k vezes que o modelo treina e classifica os dados testes, tem-se k medidas de avaliação escolhida. Em geral, alguma medida de posição e dispersão desses resultados é calculada e se tem um valor final, isto é realizado para comparar o desempenho entre as técnicas empregadas (BROWNLEE, 2018).

Figura 20 – Validação cruzada 5-fold



Fonte: Adaptado de Brownlee (2018).

2.8 Métricas de Avaliação

As técnicas utilizadas nesta dissertação para avaliar a qualidade do modelo a fim de comparar o desempenho dos classificadores são: a acurácia, sensibilidade, especificidade e área sob a curva (AUC). Contudo, antes de adentrar a essas técnicas, é importante entender o funcionamento da matriz de confusão, a qual serve como base para obter essas métricas.

2.8.1 Matriz de Confusão

A matriz de confusão é uma tabela que mostra os valores reais e valores preditos de cada classe, permitindo visualizar o desempenho do classificador, dessa forma, sendo possível obter o número de classificações corretas. Um exemplo típico de uma matriz de confusão para um problema envolvendo duas classes, é ilustrado na Tabela 1.

Tabela 1 – Matriz de confusão.

Teste	Doença (padrão ouro)	
	Presente	Ausente
Positivo	VP	FP
Negativo	FN	VN

Fonte: elaborado pelo autor

O Verdadeiro Positivo (VP) indica que o teste previu corretamente os indivíduos com a doença. O Verdadeiro Negativo (VN) indica que o teste previu corretamente os indivíduos saudáveis. O Falso Positivo (FP) aponta que o resultado do teste indica presença da doença quando na realidade o indivíduo é saudável. O Falso Negativo (FN) indica que o resultado do teste mostra que o indivíduo é saudável quando na realidade tem a doença.

Para avaliar a proporção de predições corretas é utilizada a acurácia (A), ou seja, verifica se o modelo classifica corretamente indivíduos doentes e saudáveis, em um caso médico, por exemplo. A sensibilidade (S) é a capacidade do modelo predizer corretamente os casos em que o indivíduo tem a doença. A especificidade (E) é a capacidade do modelo predizer corretamente os indivíduos saudáveis. A fórmula da acurácia, sensibilidade e especificidade são dadas pelas seguintes equações:

$$A = \frac{VP + VN}{N}, \quad S = \frac{VP}{VP + FN}, \quad E = \frac{VN}{VN + FP},$$

em que N é a quantidade total de observações. Desse modo, é importante que haja um ponto de corte que possa conciliar os valores de sensibilidade e especificidade para que as informações sejam balanceadas entre o diagnóstico correto da doença e evitar falsos positivos.

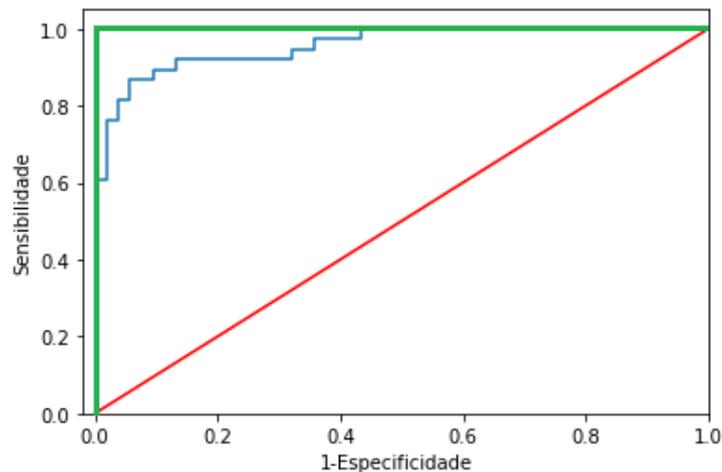
2.8.2 Área sob a curva

A Curva Receiver Operating Characteristic (ROC) é uma representação gráfica que verifica a qualidade de um modelo binário, o qual determina a probabilidade de classificar um indivíduo doente (sensibilidade), contra a probabilidade de classificar incorretamente este indivíduo com a doença (1-especificidade).

De acordo com Robin *et al.* (2014), o ponto de corte ideal para classificação é baseado em um valor que maximize tanto a sensibilidade quanto 1-especificidade. Graficamente, pela análise da curva, o ponto de corte é aproximadamente aquele em que a curva começa a flexionar. A área sob a curva (Area Under a Curve (AUC)) é uma medida de separabilidade, indicando o quanto o modelo é capaz de distinguir as classes. Assim, quanto mais próximo de 1 o valor do AUC, melhor é o modelo para distinguir as classes.

A representação da curva ROC pode ser vista através da Figura 21, em que a linha vermelha indica que o modelo não teve alteração, não distingue as duas classes. A linha azul indica que o modelo apresenta um valor intermediário em relação a distinção entre as duas classes avaliadas. E a linha verde indica um modelo ótimo, em que a AUC atinge o valor constante igual a 1.

Figura 21 – Representação da curva ROC



Fonte: elaborado pelo autor.

3 TRABALHOS RELACIONADOS

Neste Capítulo é apresentado um levantamento bibliográfico de trabalhos considerados relevantes que foram publicados recentemente, todos têm por objetivo melhorar as técnicas de detecção e classificação do glaucoma em imagens digitais.

Os sistemas de diagnóstico empregam abordagens baseadas em segmentação e extração de características para detecção de glaucoma. As abordagens baseadas em segmentação usam o tamanho do disco óptico e da escavação para o cálculo de métricas estruturais, enquanto as abordagens baseadas em extração de características se concentram na intensidade geral, bem como nas propriedades texturais e de forma da imagem do fundo para classificação (GOUR; KHANNA, 2019). Os trabalhos descritos a seguir são referentes a classificação de imagens do fundo do olho em grupo de pessoas saudáveis e grupo de pessoas com glaucoma (ou estágios do glaucoma). A separação por cada tópico é realizada com base na extração das características das imagens: somente características não geométricas; adição de métricas geométricas obtidas por segmentação; e extração de características por meio de CNNs.

3.1 Características não geométricas

As características não geométricas das imagens comumente utilizadas são informações estatísticas, textura e entropias. No trabalho proposto por Acharya *et al.* (2015), foram extraídos dos coeficientes da transformada de Gabor: média, variância, assimetria, curtose, energia, entropias de Shannon e Rényi. Essas variáveis foram submetidas à análise de componentes principais (PCA) para reduzir a dimensionalidade, as 23 componentes resultantes foram aplicadas ao classificador SVM e Naive Bayes (NB), ambos utilizando validação cruzada 10-*fold*. O melhor desempenho ocorreu com o classificador SVM, produzindo uma precisão média de 93,10%, sensibilidade de 89,75% e especificidade de 96,20%.

Outro trabalho que também utilizou informações estatísticas foi em Septiarini *et al.* (2018), em que extraiu das imagens do nervo óptico características estatísticas como média, suavidade e terceiro momento. Em seguida, selecionou as variáveis através de técnicas de correlação. Essas características foram aplicadas nos classificadores MLP, NB, SVM e k-NN. O melhor resultado foi produzido pelo algoritmo K-NN, o qual alcançou uma acurácia de 95,24%, sensibilidade de 95,12%, especificidade de 95,35% e AUC igual a 0,93.

No trabalho de Ghosh *et al.* (2015) foram extraídas do disco óptico características

de cor não-morfológica baseada no método *Grid Color Moment*. Essas informações foram aplicadas ao classificador de rede neural de propagação reversa, do inglês Back-Propagation Neural Network (BPNN), com 80% do conjunto de dados para treino, 10% para validação e 10% para teste. Os resultados alcançaram uma acurácia de 87,47%, sensibilidade de 87,50% e especificidade de 87,45%.

Outro trabalho que utilizou BPNN para classificação foi em Samanta *et al.* (2015), no qual foram extraídas características de Haralick, que é uma matriz de co-ocorrência baseadas em quatorze características estatísticas definidas sobre uma imagem. A base de imagens utilizada foi a RIM-ONE versão 2, das 455 imagens foram escolhidas de forma aleatória 321, dessas 90% foram usadas para treino, 5% para validação e 5% para teste. O resultado produziu uma acurácia de 96,26%, sensibilidade de 99,51% e especificidade de 90,43%.

No estudo realizado por Balasubramanian *et al.* (2016) foram extraídos os vasos sanguíneos através dos filtros de Gabor. Em seguida, a imagem pré-processada foi fornecida ao extrator HOG, a partir das características extraídas, foram calculadas as propriedades estatísticas: média, variância, assimetria, curtose, entropia e energia. As características mais relevantes foram fornecidas ao classificador SVM, alcançando uma precisão de 83,3% e a sensibilidade de 75,0%.

No artigo de Singh *et al.* (2016) foi proposto um método de extração de características empregando *wavelts* a partir de imagens do disco óptico segmentado e vaso sanguíneo removido. O PCA foi utilizado para redução de dimensionalidade, as componentes obtidas foram classificadas com o SVM e o k-NN ($k = 3$) com validação *leave one out*, ambos os classificadores apresentam resultados com acurácia de 94,75% e sensibilidade de 100%. A diferença foi na especificidade que resultou em 87,50% para o SVM e 90,91% para o K-NN.

No trabalho desenvolvido por Araújo *et al.* (2017) foram extraídas, das imagens pré-processadas, características dos índices de diversidade de Shannon e McIntosh como descritores de textura. Essas características foram aplicadas ao classificador SVM. A melhor separação de treino e teste foi para proporção 80/20, o resultado médio para acurácia, sensibilidade e especificidade das 5 execuções foi de 88,35%, 84,50% e 91,37%, respectivamente.

No artigo desenvolvido por Sousa *et al.* (2017) foi proposto um método de análise de textura através do LBP para representar a região do disco óptico, e funções geoestatísticas (semivariograma, emimadograma, covariograma e correlograma) para descrever quantitativamente a textura da representação LBP. Baseado no modelo gerado por um algoritmo genético (GA), o SVM foi utilizado para a classificação final de imagens, resultando em uma sensibilidade de

95,0%, acurácia de 91,2% e especificidade de 88,2%.

No trabalho elaborado por Koh *et al.* (2017) foram extraídas características de energia e entropias (Yager, Renyi, Kapoor, Shannon e Fuzzy) das imagens decompostas por *2D-Continuous Wavelet Transform* (CWT). Essas características foram selecionadas através do *Particle Swarm Optimization* (PSO), os atributos escolhidos foram usados para treinar o classificador *Random Forest* com validação cruzada estratificada *10-fold*. Este método apresentou uma taxa de precisão de 92,48%, para a sensibilidade e especificidade de 89,37% e 95,58%, respectivamente.

Maheshwari *et al.* (2016) utilizou transformada wavelet empírica (EWT) para decompor as imagens do fundo de olho, a partir desses componentes, calcularam-se as informações de correntropia. As características foram classificadas com o SVM de mínimos quadrados (LS-SVM) utilizando validação cruzada de *3-fold*. O melhor resultado deu-se com a base pública RIM-ONE v2, alcançando uma acurácia de 98,33%, sensibilidade de 100% e especificidade de 96,67%.

O método para obter as características das imagens pode ser um processo exaustivo e não intuitivo, necessitando de conhecimento prévio das imagens e dos descritores a serem utilizados, além disso é importante verificar a correlação entre essas características, pois a adição de alguma outra variável pode atrapalhar a classificação, causando problemas de multicolinearidade. A partir desse fundamento, é importante selecionar características relevantes para a distinção entre as classes de imagens. A Tabela 2 apresenta um resumo dos resultados obtidos nos trabalhos relacionados, em que os autores utilizaram características previamente indicadas.

Tabela 2 – Resumo dos trabalhos relacionados utilizando extração de características previamente indicadas.

Autores	Dados	Qtd	Método	A (%)	S (%)	E (%)
Acharya <i>et al.</i> (2015)	Kasturba Medical College	510	SVM	93,10	89,75	96,20
Septiarini <i>et al.</i> (2018)	Dr. YAP Eye and Dr. Sardjito Hospital	84	K-NN	95,24	95,12	95,35
Ghosh <i>et al.</i> (2015)	RIM-ONE	169	BPNN	87,47	87,50	87,45
Samanta <i>et al.</i> (2015)	RIM-ONE v2	455	BPNN	96,16	99,51	90,44
Balasubramanian <i>et al.</i> (2016)	Não informado	30	SVM	83,30	75,00	-
Singh <i>et al.</i> (2016)	Venu Eye Research Center	63	K-NN	94,75	100,00	90,91
Araújo <i>et al.</i> (2017)	RIM-ONE v2	455	SVM	88,35	84,50	91,37
Sousa <i>et al.</i> (2017)	RIM-ONE v2	455	SVM	91,20	95,00	88,20
Koh <i>et al.</i> (2017)	Kasturba Medical College	1486	RF	92,48	89,37	95,58
Maheshwari <i>et al.</i> (2016)	Kasturba Medical College	455	LS-SVM	98,33	100,00	96,67

Fonte: elaborado pelo autor.

Esses trabalhos apresentaram bons resultados, no entanto, a técnica de avaliação não demonstra generalização do método com outras bases de imagens. Apesar de alguns autores

utilizarem as técnicas de validação cruzada *k-fold* e outros autores utilizarem algumas repetições, ainda há o problema na generalização para diferentes bases de imagens, pois, esses trabalhos não consideram diferentes qualidades das imagens, isto é importante porque se forem englobadas diferentes imagens, eventuais vieses, causados por imagens de baixa qualidade, são reduzidos.

3.2 Características geométricas e não-geométricas

Alguns trabalhos utilizam somente as informações geométricas, obtidas a partir da segmentação, para identificação do glaucoma. A característica geométrica comumente utilizada por diversos autores é a CDR, sendo calculada a partir da regra ISNT. Como é o caso do trabalho proposto por Agarwal *et al.* (2015), no qual foi utilizado um método baseado no limiar adaptativo para segmentação da região de interesse do disco óptico, escavação e área neuro-retiniana. A partir desses resultados extraíram-se as características oculares da relação CDR e RDR (aro/disco) que foram aplicadas ao classificador SVM para categorizar as imagens como glaucomatosas ou não glaucomatosas. Os resultados apresentam uma precisão de 90%, sensibilidade 100% e especificidade 80%.

Entretanto, utilizar somente essas características para classificar o glaucoma não é confiável, pois nem todas as pessoas possuem um mesmo padrão, em algumas o valor de CDR igual a 0,5 pode ser considerado normal enquanto para outras esse valor é indício de glaucoma. O CDR se torna mais eficiente quando acompanhado de outras informações geométricas ou outras características da imagem. Em outros trabalhos, os autores optam por utilizar informações geométricas juntamente com as características não geométricas como entropias, textura e intensidade de cores. Isto ocorre no trabalho desenvolvido por Haleem *et al.* (2016), em que é proposto uma abordagem para detecção automática de glaucoma baseada no *Regional Image Features Model* (RIFM), em que são extraídas propriedades geométricas (ex: CDR, NRR) e propriedades não geométricas (ex: intensidade de *pixels*, textura). Os métodos de classificação são realizados por meio do SVM com diferenças nos kernel, o melhor classificador foi com o SVM *Twin*. Os resultados experimentais apresentaram uma acurácia de 94,4%, sensibilidade 92,2% e especificidade 95,3% para o conjunto RIM-ONE.

O método elaborado por Salam *et al.* (2016) foi semelhante a técnica do trabalho mencionado anteriormente. Nesse caso foram combinadas propriedades geométricas (CDR e NRR) e não geométricas (textura e intensidade) para melhorar a precisão do diagnóstico automatizado do glaucoma. Essas características foram aplicadas ao PCA para redução de dimensionalidade,

as componentes principais foram utilizadas para treinar o classificador SVM, os resultados alcançaram uma acurácia de 91%, sensibilidade de 100% e especificidade de 87%.

No trabalho elaborado por Kausu *et al.* (2018), foi proposto um método de identificação de glaucoma baseado nas características invariantes no tempo e transformada de *wavelets*. A segmentação do disco óptico foi realizada usando o método de agrupamento *Fuzzy C-Means* e o limiar de Otsu foi usado para a segmentação da escavação, as características de CDR, NRR e BVR foram calculadas com base na segmentação. A classificação foi realizada com diferentes combinações do conjunto de características usando quatro tipos de classificadores: MLP, SVM com função de núcleo polinomial e linear, *random forest* e *Adaboost*. O melhor resultado foi através do MLP com validação cruzada *10-fold*, alcançando uma acurácia de 97,67%, sensibilidade de 98,0% e especificidade de 97,1%.

No trabalho formulado por CARVALHO JUNIOR *et al.* (2017) foi realizado uma comparação entre quatro métodos de segmentação utilizando os algoritmos de Otsu e k-means. Na primeira etapa foi realizada a segmentação do disco óptico, na segunda etapa foi utilizada a extração de características baseada nos índices de diversidade filogenética para caracterizar a textura. O melhor resultado foi aplicando o Otsu com o classificador Decision Table, a sensibilidade alcançou 97,5%, especificidade de 100,0% e uma acurácia de 97,9%.

Assim como no caso em que os autores utilizaram apenas as características não-geométricas para identificar o glaucoma, no caso da adição de características geométricas os resultados também estão satisfatórios, no entanto, a forma de avaliação dos resultados não abrange diferentes tipos de imagens. A maioria dos trabalhos utilizaram apenas uma única execução de validação cruzada com *k-fold*, o que não é suficiente para avaliar diferentes configurações e qualidades de imagens. Ademais, em determinadas situações, dependendo da qualidade das imagens e algoritmos utilizados, a segmentação se torna imprecisa, resultando em cálculo incorretos dos parâmetros de identificação do glaucoma, ocasionando em taxas de erros e classificação incorreta das imagens. A Tabela 3 apresenta um resumo dos resultados obtidos nos trabalhos relacionados, em que os autores utilizaram métodos de segmentação para obter os valores das características geométricas das imagens.

Tabela 3 – Resumo dos trabalhos relacionados utilizando extração de características geométricas.

Autores	Dados	Qtd	Método	A (%)	S (%)	E (%)
Agarwal <i>et al.</i> (2015)	Não informado	60	SVM	90,00	100,00	80,00
Haleem <i>et al.</i> (2016)	RIM-ONE	124	SVM Twin	94,40	92,20	95,30
Salam <i>et al.</i> (2016)	-	100	SVM	91,00	100,00	87,00
Kausu <i>et al.</i> (2018)	Venu Eye Research Centre	86	MLP	97,67	98,00	97,10
Carvalho Júnior <i>et al.</i> (2017)	RIM-ONE v1	169	Decision Table	97,50	97,90	100,00

Fonte: elaborado pelo autor.

3.3 Redes Neurais Convolucionais

As técnicas mais comuns para classificação de imagens utilizando redes convolucionais são aquelas que extraem e classificam a imagem em um único processo. Os seguintes trabalhos fazem uso dessa técnica, além disso, alguns autores usam métodos para selecionar as características mais importantes e otimizar as camadas do modelo proposto.

No trabalho de Abbas (2017), a arquitetura não-supervisionada de uma CNN foi usada para extrair características da imagem do fundo do olho. Posteriormente, o modelo de *deep-belief network* (DBN) foi usado para selecionar as características mais discriminantes. Por fim, a decisão final foi realizada por um classificador linear *softmax* para diferenciar entre glaucoma e não-glaucoma. As bases utilizadas foram: DRIONS-DB, HRF-dataset, sjchoi86-HRF e PVR. O resultado médio da validação cruzada com 10-*fold* apresentou uma sensibilidade de 84,50%, especificidade de 98,01% e acurácia de 99,00%.

Lima *et al.* (2018) usaram um Algoritmo Genético (GA) para otimizar arquiteturas de CNN para auxiliar no diagnóstico de glaucoma usando a imagem de fundo do olho. O trabalho foi dividido em três etapas: pré-processamento das imagens redimensionando-as, execução do GA para encontrar a melhor arquitetura da CNN, e a CNN para classificar as imagens em saudáveis ou não. Os resultados, após o treinamento do melhor modelo escolhido pelo GA, apresentaram uma precisão de 91%.

Outros autores optam por utilizar a CNN somente para extração das características e classificam com métodos comuns de aprendizado de máquinas. Tal como o método elaborado por Al-Bander *et al.* (2017), em que um sistema automatizado baseado em uma CNN foi desenvolvido para extrair características que distinguem padrões normais e glaucomatosos, essas características serviram de entrada para o SVM classificar as imagens. Os resultados da acurácia, sensibilidade e especificidade foram 88,2%, 85,0% e 90,8%, respectivamente.

Um outro método utilizando CNN é através de aprendizado por transferência (*learning transfer*), no qual a rede armazena o conhecimento adquirido ao resolver um problema e

aplica a um problema diferente, mas que possua alguma relação com aquele. Recentemente, alguns trabalhos tem optado por utilizar tais técnicas, pois, construir uma rede com inicialização aleatória é complexo e requer muito conhecimento. Como o modelo de aprendizado por transferência já foi treinado, seu resultado consegue ser mais rápido do que CNNs comuns. A seguir alguns trabalhos que fizeram uso de tal técnica.

O método proposto por Cerentini *et al.* (2018) foi baseado no treinamento de uma rede neural GoogLeNet adaptada. A metodologia foi dividida em detecção da região de interesse (ROI) e classificação de imagens. A rede foi treinada usando ROIs manualmente extraídos e outras características da imagem. Posteriormente, outro modelo da GoogLeNet foi treinado usando as imagens resultantes anteriores. Os bancos de dados utilizados foram HRF, RIM-ONE (versão 1, 2 e 3) e a junção deles. O melhor desempenho se deu com a avaliação individual do RIM-ONE versão 1, alcançando uma acurácia de 94,2%.

No trabalho proposto por Ahn *et al.* (2018) foram comparados três modelos: classificação logística simples, uma CNN usando o Tensorflow, e o modelo GoogleNet Inception v3 pré-treinado. O melhor desempenho para o conjunto teste foi com a CNN proposta pelo trabalho, alcançando uma precisão de 87,9%, enquanto a classificação logística simples e a Inception v3 alcançaram somente 77,2% e 84,5%, respectivamente.

No trabalho de Zhen *et al.* (2018), oito CNNs (VGG16, VGG19, ResNet, DenseNet, InceptionV3, InceptionResNet, Xception e NASNetMobile) foram treinadas separadamente para avaliar a gravidade do glaucoma. As CNNs, exceto VGG16 e VGG19, obtiveram escores kappa médios de 80,36% e 78,22% quando treinados com pesos iniciais iguais a zero em ROIs globais e locais, e 85,29% e 82,72% usando os pesos pré-treinados, respectivamente. Entre essas CNNs, a DenseNet teve a maior precisão de classificação (75,50%) com base em pesos pré-treinados ao usar ROIs globais, em comparação com 65,50% ao usar ROIs locais.

Gómez-Valverde *et al.* (2019) exploraram a aplicação de diferentes arquiteturas de CNN para demonstrar a influência no desempenho de fatores relevantes, como o tamanho do conjunto de dados, a arquitetura e o uso de aprendizado de transferência versus arquiteturas recém-definidas. A base utilizada foi uma junção da ESPERANZA, RIM-ONE (v1, v2 e v3) e DRISHTI-GS. O melhor desempenho foi com aprendizagem por transferência com o modelo VGG19, produzindo uma AUC de 0,94, acurácia de 88,05%, sensibilidade de 87,01% e especificidade de 89,01%.

Trabalhos mais recentes utilizaram técnicas que combinam características não geo-

métricas manualmente indicadas com as características extraídas por uma rede convolucional. Esse foi o caso do trabalho elaborado por Claro *et al.* (2018), no qual extraíram características através do uso de descritores de textura e Redes Neurais Convolucionais (CNNs). Os resultados mostraram que a junção dos descritores GLCM e CNNs e a utilização do classificador *Random Forest* foram promissoras na detecção dessa doença, obtendo uma acurácia de 91,06%.

No método elaborado por Ghamdi *et al.* (2019), foi utilizada uma CNN pré-treinada VGG16 com pesos obtidos pela base ImageNet, em seguida foi ajustada e treinada de maneira supervisionada usando os dados rotulados de imagens do fundo de olho. A abordagem de autoaprendizagem foi então usada para prever os dados não rotulados. Os resultados com a base RIM-ONE v2 utilizando a técnica semi-supervisionada atingiu acurácia de 92,4%, sensibilidade de 91,7% e especificidade de 93,3%.

Empregar redes convolucionais é uma ótima alternativa para classificação e segmentação de imagens. No entanto, apesar da praticidade na extração de características, sem a necessidade de descritores previamente estabelecidos, esse procedimento ainda é demorado. Com base nos resultados obtidos nos trabalhos relacionados, é importante utilizar uma arquitetura de redes pré-treinadas que sejam específicas para obter informações relevantes na identificação de glaucoma em imagens do fundo de olho. Uma possibilidade para melhorar o desempenho é testar outras arquiteturas de aprendizado por transferência e identificar qual produz resultados com maiores taxas de acurácia. Desse modo, a extração de características seria realizada de maneira mais rápida do que CNN comuns treinadas a partir de pesos iniciais iguais a zero. Além disso, utilizar métodos para reduzir a quantidade de atributos é opção para acelerar o processo de classificação das classes de imagens. A Tabela 4 apresenta um resumo dos resultados obtidos nos trabalhos relacionados, em que os autores utilizaram redes neurais convolucionais para extrair as características das imagens, alguns classificaram no mesmo procedimento, já em outros trabalhos foram utilizados CNNs somente para extrair as características.

Tabela 4 – Resumo dos trabalhos relacionados utilizando extração de características por redes neurais convolucionais.

Autores	Dados	Qtd	Método	A (%)	S (%)	E (%)
Abbas (2017)	DRIONS-DB, HRF e PVR	1200	CNN	84,50	98,01	99,00
Lima <i>et al.</i> (2018)	RIM-ONE v2	455	CNN	91,00	-	-
Al-Bander <i>et al.</i> (2017)	RIM-ONE v2	455	SVM	88,20	85,00	90,80
Cerentini <i>et al.</i> (2018)	RIM-ONE v1	158	CNN	94,20	-	-
Ahn <i>et al.</i> (2018)	Kim's Eye Hospital	1542	CNN	87,90	-	-
Zhen <i>et al.</i> (2018)	Beijing Tongren Hospital	5978	CNN	75,50	-	-
Gómez-Valverde <i>et al.</i> (2019)		2313	VGG19	88,05	87,01	89,01
Claro <i>et al.</i> (2018)		873	Random Forest	91,06	-	-
Ghamdi <i>et al.</i> (2019)			VGG16	92,40	91,70	93,30

Fonte: elaborado pelo autor.

4 MATERIAIS E MÉTODOS

Neste Capítulo são apresentados os materiais e métodos que auxiliam na construção desta dissertação. São explicados os três modelos propostos baseados nos diferentes tipos de extração de características, juntamente com os métodos de avaliação de desempenho dos classificadores em estudo para a base RIM-ONE v2.

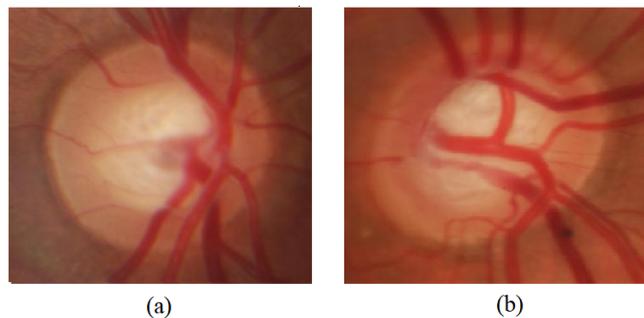
4.1 Ambiente de desenvolvimento e teste

As análises foram realizadas em uma máquina com processador Intel(R) Core(TM) i3-4330 CPU 3,50 GHz, com memória RAM de 12,0 GB e o sistema operacional de 64 bits. Utiliza-se a linguagem de programação Python para a análise computacional.

4.2 Base de imagens

As imagens utilizadas são de uma base pública *An Open Retinal Image Database for Optic Nerve Evaluation* (RIM-ONE) versão 2, composta por imagens centralizadas no disco óptico, das quais 200 são referentes a pessoas com glaucoma/suspeita de glaucoma e 255 a pessoas saudáveis, totalizando 455 imagens (FUMERO *et al.*, 2011). As imagens foram adquiridas com diferentes dimensões. Todas estão no formato JPEG, no sistema RGB (vermelho, verde e azul, do inglês *R-Red, G-Green e B-Blue*). Exemplos de imagens do banco de dados RIM-ONE versão 2, exibindo olhos saudáveis e com glaucoma, respectivamente, encontram-se na Figura 22.

Figura 22 – Imagem de fundo do olho de paciente, (a) saudável e (b) com glaucoma.



Fonte: RIM-ONE-v2

Esta base é formada por bancos de dados oftalmológicos projetados, servindo de

referência para um projeto de algoritmos de segmentação do nervo óptico e no diagnóstico de glaucoma auxiliado por computador. Foi criado pela colaboração de três hospitais espanhóis: Hospital Universitario de Canarias, Hospital Clínico San Carlos e Hospital Universitario Miguel Servet. O conjunto de imagens foi organizado pela colaboração de 4 especialistas em glaucoma, cujo equipamento usado para capturar as imagens foi uma câmera de fundo Nidek AFC-210 com uma Canon EOS 5D Mark II de 21,1 megapixels (GÓMEZ-VALVERDE *et al.*, 2019).

4.3 Métodos

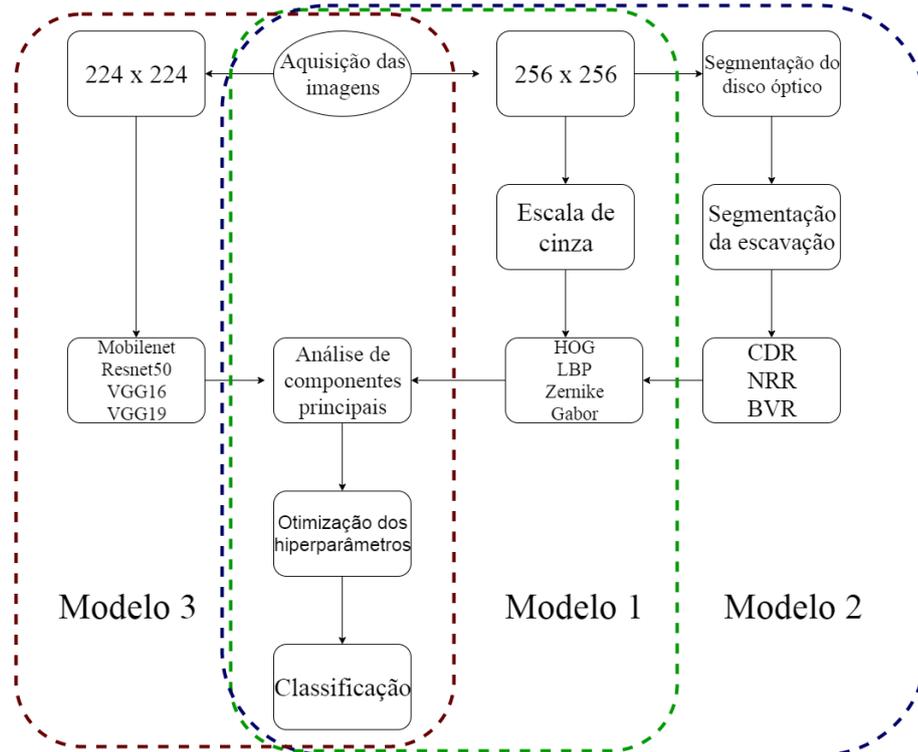
São avaliados três modelos, os quais estão separados de acordo com a seleção de características. No modelo 1 são utilizadas informações não geométricas indicadas previamente; o modelo 2 é similar ao modelo 1 com a adição de características geométricas obtidas a partir da segmentação do nervo óptico; e no modelo 3 são extraídas características das imagens com base nas redes convolucionais pré-treinadas. O diagrama de blocos representando os passos dos modelos propostos para avaliação de métodos de extração de características e classificação é mostrado na Figura 23.

No modelo 1, as imagens originais são redimensionadas para o tamanho 256×256 , para que todas as imagens estejam com as mesmas dimensões. Em seguida, todas são convertidas em escala de cinzas. Após esse pré-processamento, são extraídas as características não geométricas indicadas pelos descritores LBP, HOG, momentos de Zernike e informações estatísticas da imagem após o filtro de Gabor. No modelo 2, após o redimensionamento das imagens, são realizadas as etapas de segmentação do disco óptico e escavação, para que as informações geométricas BVR, CDR e NRR sejam calculados. Estas três informações são adicionadas aos descritores indicados no modelo 1. E, para o modelo 3, as imagens são redimensionadas para 224×224 *pixels*, este procedimento é realizado para que as dimensões de entrada sejam consistente com a camada de entrada dos modelos pré-treinados.

Após as imagens estarem representadas por um vetor de características, o PCA é aplicado, em cada modelo, para reduzir a dimensionalidade dos dados, são testadas configurações de porcentagem de variância explicada, o percentual de variância explicada utilizado é 95%, este valor é selecionado devido 95% das componentes representarem melhor a imagem do que com outro percentual, 90% e 99%, por exemplo. As componentes resultantes são utilizadas como entrada no algoritmo de classificação. Com o intuito de melhorar o desempenho dos classificadores para o conjunto de treino, e conseqüentemente, obter melhores resultados nos

dados de teste, é realizada a otimização dos hiperparâmetros pelo método de pesquisa em grade com validação cruzada *5-fold*. São realizadas 100 execuções e escolhidos os hiperparâmetros que apresentam a maior taxa de acurácia no conjunto de treino de validação *5-fold*. Ao final de cada algoritmo de classificação, são escolhidos os hiperparâmetros com maiores frequências nas diferentes execuções.

Figura 23 – Diagrama de Blocos do modelo proposto.



Fonte: elaborado pelo autor.

Para cada um dos modelos, após a escolha dos hiperparâmetros pelo método de pesquisa em grade, são realizadas 100 execuções. Em cada uma é realizada uma separação aleatória de treino e teste com amostras diferentes. Contudo, todas as execuções têm a configuração na proporção de 80/20, ou seja, 364 imagens de treino e 91 imagens de teste. Como forma de comparação, são calculadas as taxas médias das 100 execuções: acurácia, sensibilidade, especificidade e área sob a curva.

Tais técnicas são utilizadas a fim de comparar o desempenho dos classificadores nos três modelos propostos. Como o interesse é identificar indivíduos com glaucoma, então é importante que o classificador distinga as diferentes classes, além de apresentar a maior taxa de acurácia, é desejado que o método alcance uma taxa de sensibilidade alta.

4.3.1 Modelo 1

Após o pré-processamento, cada descritor extrai informações das imagens, e essas informações são postas em um vetor que representará a imagem de forma reduzida. Os parâmetros para cada descritor são escolhidos através de testes exaustivos, e assim os melhores valores são selecionados conforme os resultados e a quantidade de características extraídas. Os descritores foram testados individualmente e combinados a outros. Os descritores indicados a seguir estão dispostos conforme o acréscimo no modelo. Portanto, os parâmetros selecionados são baseados na junção de todos esses descritores no modelo 1.

Para o descritor HOG é utilizado a normalização pela raiz quadrada, as imagens são normalizadas com 8 orientações, 128×128 pixels por célula, e 1×1 células por blocos, o total de características obtidas para esse descritor é igual a 32. Quando o LBP é combinado com histogramas de descritores de gradientes orientados (HOG), melhora consideravelmente o desempenho de detecção em alguns conjuntos de dados (WANG *et al.*, 2009).

Para o conjunto de imagens RIM-ONE v2, o cálculo do LBP é tratado usando o raio igual a 8 e o número de vizinhos igual a 24, totalizando 26 atributos, isto porque o descritor LBP é baseado nos níveis de cinza para descrever a textura da imagem, portanto, as intensidades de níveis de cinza é igual ao número de pontos somado dois, que referem-se a um nível extra de rotação e invariância em escala de cinza. O valor de 24 foi escolhido por apresentar maior acurácia, pois um número maior resultaria em uma quantidade de características maior, e a diferença no desempenho da classificação não seria significativa, visto que o valor da acurácia aumentaria somente na terceira casa decimal.

Considerando os momentos de Zernike, o raio utilizado é igual a 128, esse valor é referente ao raio da imagem de entrada. A quantidade de características para este descritor é igual a 25.

E por fim, os filtros de Gabor são aplicados nas imagens, e a partir das imagens filtradas são calculadas as informações estatísticas: média, assimetria, curtose e energia. O valor de σ que se refere ao ângulo do filtro que varia de 1 a 3, totalizando 12 informações estatísticas, ou seja, 4 características para cada ângulo, calculadas a partir da imagem filtrada.

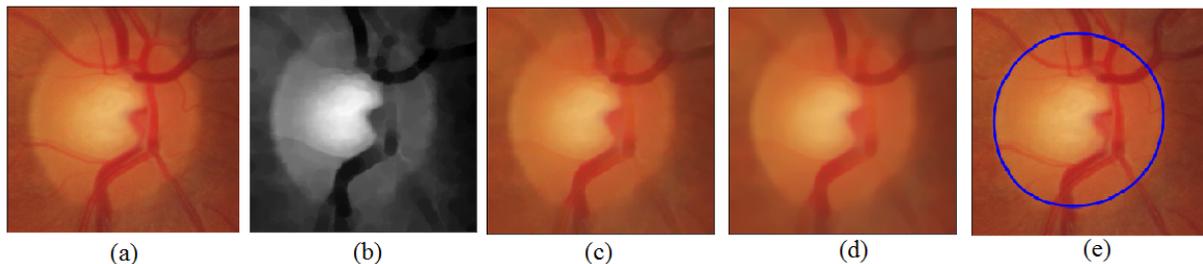
O total de características do modelo 1 resulta em 95. Após a aplicação do PCA, a quantidade de componentes se reduz para 42 componentes, ou seja, 42 componentes contém as informações mais importantes das imagens, enquanto o restante das 43 componentes não contém informações tão relevantes que representam as imagens.

4.3.2 Modelo 2

Para calcular os parâmetros, é necessário segmentar o disco óptico e escavação. Inicialmente, para segmentar o disco óptico, aplica-se a operação morfológica *black top hat* (h) no componente verde (G) para remoção dos vasos sanguíneos, $h = (f \bullet b) - f$, em que f é a imagem de entrada e b um elemento estruturante definido como um disco de dimensão igual a 6, estas são funções que atribuem um nível de cinza a cada par distinto de coordenadas (x, y) , e \bullet é a operação de fechamento, ou seja, dilatação de f por b , seguida pela erosão do resultado por b . A imagem resultante é visualizada na Figura 24 (b).

Posteriormente, as componentes vermelho e azul são acrescentadas na imagem resultante após o procedimento realizado na componente verde, assim como exibido na Figura 24 (c). Aplica-se na imagem resultante um filtro mediana para redução do ruído, conforme Figura 24 (d). Para segmentar o disco óptico é utilizado o Método de Contorno Ativo (MCA). Este método traça automaticamente uma curva inicial externa, que se deforma por iterações sucessivas, deslocando-a e reduzindo-a até as bordas do objeto, conforme Figura 24 (e).

Figura 24 – Imagens original em (a), operação de *black top hat* na componente verde em (b), junção das componentes RGB após o processamento da componente G em (c), filtragem mediana em (d) e Método do Contorno Ativo em (e).

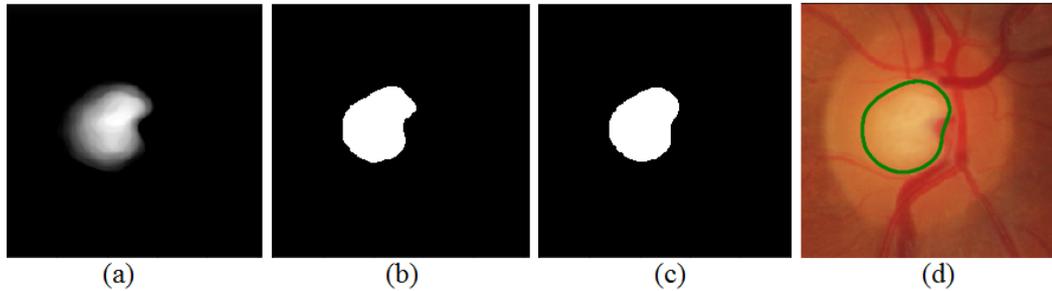


Fonte: elaborada pelo autor.

Para segmentar a escavação foi utilizada somente a componente verde do processamento realizado para segmentação do disco óptico, conforme Figura 24 (d). Portanto, a componente verde G_{i+1} da imagem é recalculada por $G_i - (\mu_i + \sigma_i)$, a imagem resultante é visualizada na Figura 25 (a). Neste caso, G_i é a componente verde original, μ_i é a média da imagem e σ_i o seu desvio padrão. Em seguida, aplica-se a Limiarização de Otsu para discriminar a região da escavação, como visto na Figura 25 (b). Para suavizar o contorno das bordas aplica-se um filtro da mediana, a imagem resultante é vista na Figura 25 (c). E finalmente, para contornar a escavação é utilizado uma função para localização das bordas, à medida que é identificado uma

variação de intensidade um ponto é aplicado na região, esse processo é realizado até que todo o objeto seja contornado, no caso a região da escavação, conforme Figura 25 (d).

Figura 25 – Imagem pré-processada em (a), limiarização de Otsu em (b), filtragem mediana em (c) e contorno da escavação em (d).



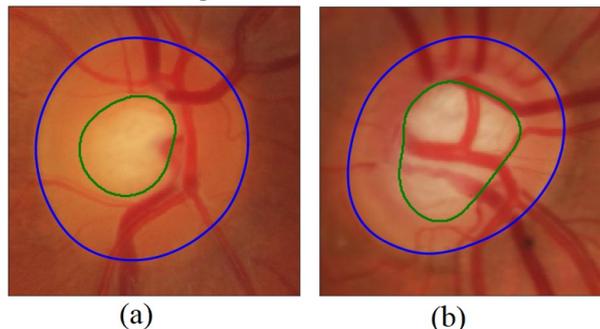
Fonte: elaborada pelo autor.

Após a segmentação do disco óptico e escavação, são calculados os descritores geométricos da imagem do nervo óptico. Um dos parâmetros mais utilizados na literatura para descrever a progressão do glaucoma é o CDR. O cálculo é obtido a partir da expressão dada por (KAUSU *et al.*, 2018),

$$CDR = \frac{\text{Área da escavação}}{\text{Área do disco óptico}}$$

Em pacientes saudáveis, a proporção da escavação em relação ao disco está em torno de 0,4, ou seja, quatro décimos do disco é ocupado pela escavação, um valor maior pode ser indício de glaucoma ou outra doença. Na Figura 26 é possível visualizar o contorno referente a segmentação do disco óptico e escavação para o cálculo CDR, na Figura 26 (a) é um exemplo de um paciente com disco óptico saudável, enquanto na Figura 26 (b) um paciente com glaucoma.

Figura 26 – Contorno do disco óptico e escavação de um paciente saudável em (a) e com glaucoma em (b).

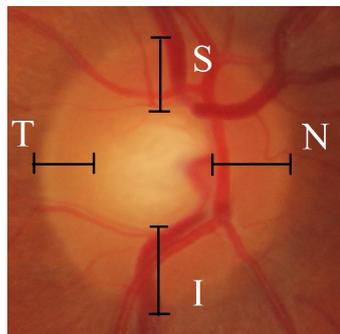


Fonte: elaborado pelo autor.

No entanto, é importante verificar outros parâmetros, pois, este parâmetro avaliado isoladamente não é indicativo da doença. Isto porque outros fatores como a idade, podem ser indicativos do aumento da escavação. Esse parâmetro possui limitações como: não refletir o tamanho do disco óptico e a posição da escavação, não considerar a quantidade de perda de fibras nervosas na borda neuroretiniana, que determina diretamente a perda do campo visual. Devido a isso, é importante a adição de outros parâmetros, juntamente com a CDR, para a identificação do glaucoma.

Com base nisso, são avaliados mais dois parâmetros, são eles: NRR e BVR. Estes parâmetros consideram a área do disco óptico e o deslocamento dos vasos, conforme a separação pela regra ISNT. Essas características, quando irregulares, são indicativos de glaucoma. É possível verificar na Figura 27 a divisão ISNT em que, para um disco saudável, tem-se que a espessura da borda neuroretiniana, segue a ordem para a área dos quadrantes Inferior (I) > Superior (S) > Nasal (N) > Temporal (T) (ABRÀMOFF *et al.*, 2010).

Figura 27 – Quadrante ISNT.



Fonte: elaborada pelo autor.

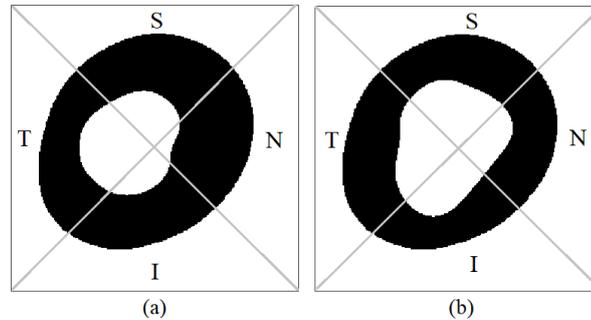
Outra referência usada para detecção de glaucoma é a espessura da borda neuroretiniana (NRR). De acordo com a regra ISNT, uma espessura normal do olho NRR é tal que a região inferior é maior que a superior, a qual é maior que a nasal e esta é maior que a temporal. Para calcular o NRR, remove-se a escavação de dentro do disco óptico, conforme Figura 28, e divide-se a área restante em quadrantes. Em seguida, encontra-se a razão entre as áreas dos quadrantes verticais e horizontais dada por (KAUSU *et al.*, 2018),

$$NRR = \frac{\text{Área I} + \text{Área S}}{\text{Área N} + \text{Área T}} \quad (4.1)$$

A Figura 28 é referente um paciente saudável e a Figura 28 (b) referente a um

paciente com glaucoma, ambas as imagens estão divididas conforme o quadrante ISNT para o cálculo das áreas das bordas do disco óptico.

Figura 28 – Borda neuroretiniana e quadrantes ISNT para o cálculo NRR de um fundo de olho saudável em (a) e com glaucoma em (b).



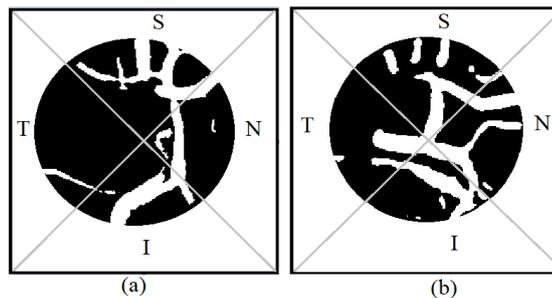
Fonte: elaborada pelo autor.

Outro parâmetro importante no diagnóstico do glaucoma é o *Blood Vessel Ratio* (BVR), em que é baseado na área dos vasos sanguíneos do fundo de olho. Se um olho é saudável, seus vasos sanguíneos concentram-se nos quadrantes superior e inferior. A fórmula do BVR é dado por (KAUSU *et al.*, 2018),

$$BVR = \frac{\text{Área dos vasos I} + \text{Área dos vasos S}}{\text{Área dos vasos N} + \text{Área dos vasos T}} \quad (4.2)$$

Para valores maiores que 1, é considerado que o paciente está saudável, enquanto valores menores que 1, é um indício do glaucoma. Na Figura 29 (a) observa-se os vasos extraídos de um paciente saudável e na Figura na Figura 29 (b) um paciente com glaucoma, além dos quadrantes que separam as áreas para o cálculo da característica geométrica ocular BVR.

Figura 29 – Segmentação dos vasos sanguíneos no disco óptico de um paciente saudável em (a) e com glaucoma em (b).



Fonte: elaborada pelo autor.

4.3.3 Modelo 3

No aprendizado por transferência é possível aproveitar informações de modelos treinados anteriormente, como características, pesos e etc. Os métodos de aprendizagem por transferência têm a arquitetura em camadas que permite utilizar uma rede pré-treinada, sem sua camada final, como um extrator de características de imagens. As arquiteturas utilizadas são: ResNet50, MobileNet, VGG16 e VGG19.

Todas as imagens são redimensionadas para 224×224 *pixels*, isto para que as imagens tenham as mesmas dimensões de entrada das arquiteturas de CNN. Após a extração de características, a PCA é aplicada para reduzir a dimensionalidade dos dados, a quantidade de componentes resultantes é igual a 186, 111, 113 e 114 para as arquiteturas MobileNet, ResNet50, VGG16 e VGG19, respectivamente.

5 RESULTADOS E DISCUSSÕES

Este Capítulo dedica-se a apresentação e discussões dos resultados obtidos nos experimentos realizados. Estão divididos conforme cada modelo. Inicialmente, são discutidos os resultados do modelo 1, o qual refere-se a aplicação dos métodos de classificação em que as características não geométricas das imagens são indicadas previamente. Em seguida, são apresentados os resultados do modelo 2, em que são adicionadas ao modelo 1 as características geométricas, calculadas a partir da segmentação de imagens. Por fim, descrevem-se os resultados do modelo 3, obtidos a partir da extração de características por meio de aprendizado por transferência com arquiteturas pré-treinadas. Todas as classificações são realizadas com os métodos de aprendizado de máquinas supervisionado.

5.1 Modelo 1

Com base nos trabalhos relacionados, testam-se várias configurações de descritores de características, os quais são avaliados individualmente e combinados com os demais para classificar as imagens. Com base em exaustivos experimentos combinando diferentes extratores de características, os descritores que apresentam melhores desempenho de acurácia, concomitantemente, para a base RIM-ONE v2 são: LBP, HOG, momentos de Zernike e informações estatísticas da imagem após a aplicação do filtro de Gabor.

Avaliando a influência de cada descritor individual para cada algoritmo de classificação, o melhor desempenho é obtido com o HOG, o qual apresenta uma acurácia média entre 79% e 83% em relação aos classificadores estudados. Apesar disso, a sensibilidade permanece abaixo da média de acurácia, ou seja, o algoritmo não classifica adequadamente os indivíduos com glaucoma. A sensibilidade média permanece entre 68% a 77%. Já os indivíduos saudáveis são classificados corretamente, o que é mostrado através do valor da especificidade do teste, em que a menor taxa avaliada é igual a 84% para a rede MLP. Em seguida, o descritor LBP é adicionado, os resultados apresentam um aumento significativo na acurácia e sensibilidade para todos os classificadores em relação aos dados de teste.

Outro descritor incluído no modelo 1 são os momentos de Zernike. Com a adição desse descritor, os resultados produzem aumento na sensibilidade, exceto com o classificador GBDT que reduziu em 2% a taxa de acurácia. Apesar disso, como houve aumento nas taxas de sensibilidades nos demais classificadores, decide-se que este descritor permaneça no modelo.

Adicionando as informações do filtro de Gabor para descrever a textura das imagens, a taxa de sensibilidade e especificidade aumentam levemente, conseqüentemente a acurácia média também. A regressão logística e MLP não apresentam alterações significativas, porém o SVM alcança os maiores valores observados até o momento. Desse modo, fora decidido que as informações estatísticas continuam no modelo.

Através de testes exaustivos com vários descritores, os descritores mencionados anteriormente são os melhores testados para comporem os modelos, pois, produzem os melhores resultados para os classificadores avaliados.

Com base nos testes, decide-se selecionar os descritores: HOG, LBP, momentos de Zernike e informações do filtro de Gabor. No total, são 95 características extraídas, compondo um vetor com 32 características obtidas pelo descritor HOG, 26 do LBP, 25 dos momentos de Zernike e 12 informações provenientes da imagem filtradas com Gabor.

Esses descritores são classificados sem a redução de dimensionalidade, no entanto, os resultados são inferiores aos obtidos com o uso do PCA, ademais o custo computacional é maior por ter que tratar uma quantidade maior de informações. Portanto, todos os resultados apresentados são obtidos com a aplicação do PCA para redução de dimensionalidade. A quantidade de componentes resultante deste procedimento é igual a 42, que são as informações de entrada para os classificadores.

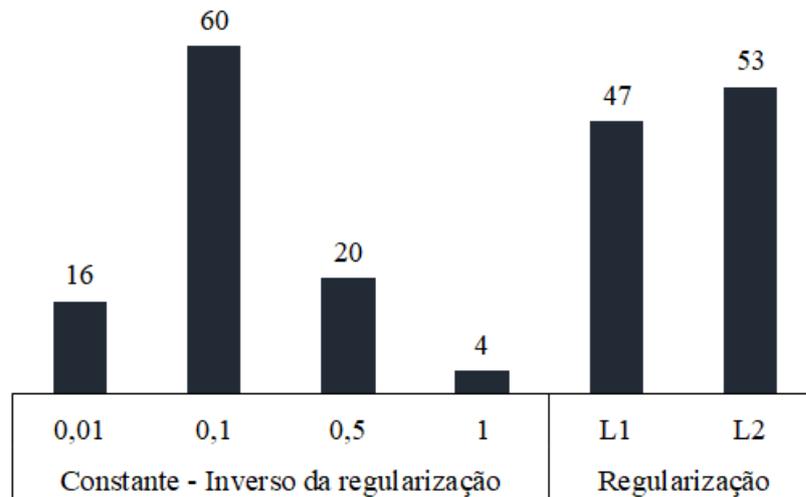
A seguir são apresentadas as avaliações de cada classificador com seus respectivos hiperparâmetros nos referidos intervalos, este procedimento tem o intuito de aprimorar os classificadores. Como a pesquisa em grade é um processo demorado, realiza-se somente para obter os melhores hiperparâmetros do conjunto de dados.

5.1.1 Regressão Logística

Considerando a frequência de hiperparâmetros das 100 execuções, para a regressão logística, a regularização L2 sobressai a quantidade de execuções em relação a regularização L1, conforme está apresentado na Figura 30. Apesar das quantidades estarem bastante próximas, é usada a regularização L2, conhecida como Regressão de Ridge. Para esse classificador, usa-se o valor de $C = 0,1$ (inverso da regularização - $\frac{1}{\lambda}$) que também é a maioria dos casos.

As taxas médias de sensibilidade são inferiores às taxas de especificidade, mostrando que, independente das execuções, o algoritmo tem dificuldade de identificar indivíduos com glaucoma, resultando assim em falsos negativos.

Figura 30 – Frequência dos hiperparâmetros do modelo de regressão logística.



Fonte: elaborado pelo autor.

Em algumas configurações de treino e teste, os resultados podem ser considerados bons, como nas configurações em que a taxa máxima resultante é igual a 91,21%, isso ocorre devido a algumas imagens estarem bem definidas e com tonalidades e contornos de fácil distinção. Enquanto em outras configurações, a identificação entre as classes apresenta valores menores, 75,82%. Nesses casos, as imagens de treinos estão mais borradas e difícil distinção, dificultando assim a classificação. As taxas médias mostram valores adequados por ser um processo rápido e simples, não apresentando complexidade na execução do algoritmo e com classificações desejáveis.

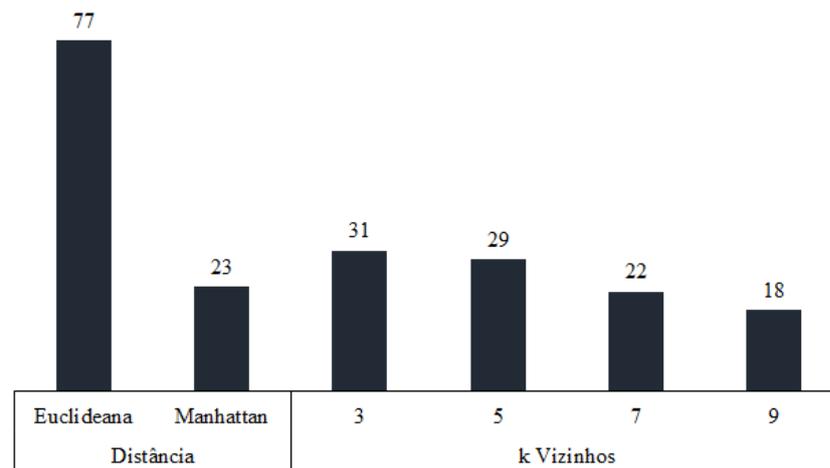
Os valores médios de sensibilidade e especificidade não se mantêm distantes, nas 100 execuções não houve uma grande variação em cada configuração de teste, sendo observado pela média do desvio padrão de 3,1%. As taxas de sensibilidade apresentam a maior variação, assim como observado anteriormente: o algoritmo de classificação não consegue distinguir bem os indivíduos com a doença.

A área sob a curva ROC representa uma forma de descrever a performance do conjunto teste, quanto maior o poder do teste em discriminar os indivíduos doentes e não doentes mais próximo de 1 é o valor de AUC. Portanto, como é uma medida de acurácia do teste, os valores de AUC e acurácia são próximos.

5.1.2 *k* Vizinhos Mais Próximos

Considerado o classificador k-NN, a métrica de distância euclidiana é escolhida em 77 execuções, enquanto a medida de distância Manhattan é selecionada em 23, porém em nenhuma execução a distância de Minkowski apresenta maior acurácia nos dados de teste. O peso uniforme é escolhido em 60 execuções, enquanto o peso de distância, para 40 execuções. O peso uniforme indica que todos os indivíduos possuem o mesmo peso, enquanto pela distância indica que o peso de cada ponto é medido pelo inverso da distância, ou seja, os vizinhos mais próximos devem ter pesos maiores do que vizinhos mais distantes. São testadas a quantidade de vizinhos k iguais a 3, 5, 7 e 9. A maior frequência de k , 31 vezes, com maiores taxas de acurácia são obtidas para o $k = 3$, conforme ilustrado na Figura 31.

Figura 31 – Frequência dos hiperparâmetros do modelo k-NN.



Fonte: elaborado pelo autor.

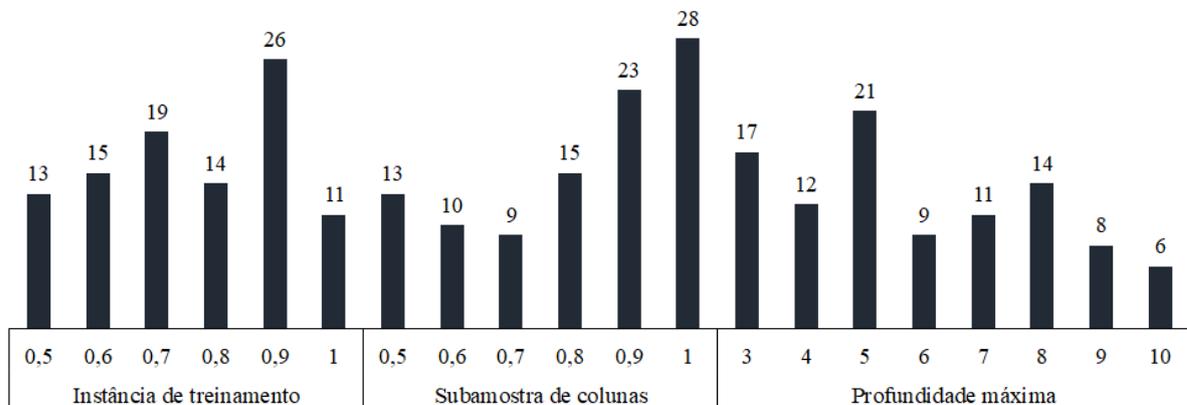
Avaliando os resultados de teste do algoritmo k-NN, a acurácia média se manteve em 81%, porém, este algoritmo não mostra ser bom para descrever pacientes com glaucoma, pois, a taxa média de sensibilidade manteve-se baixa com 69%, enquanto a taxa de especificidade é alta, 91%. As taxas mínimas e máximas têm uma variação grande, entre 73% e 89%, respectivamente. A diferença entre sensibilidade e especificidade mostra-se grande, indicando que o k-NN não classifica corretamente as classes analisadas, em especial, a classe de indivíduos portadores do glaucoma. O desvio padrão da acurácia e especificidade estão próximos, apresentando valores menores que 4%, enquanto o desvio padrão da sensibilidade manteve-se próximo a 7%. O k-NN classifica as imagens, em sua maioria, como sendo provenientes de indivíduos saudáveis, apresentando muitos falsos negativos.

Portanto, o classificador k-NN não se mostra um classificador eficiente para a base de dados estudada. Além disso, os resultados de cada execução indicam uma dispersão alta, com valores muito baixos e outros altos. No melhor dos casos, a acurácia máxima obtida é de 89% e sensibilidade 85%, não sendo um bom resultado, pois, esses valores são obtidos com as melhores configurações de imagens que o classificador conseguiu distinguir.

5.1.3 *Árvore de Decisão com Gradiente Aumentado*

Considerando o classificador GBDT, os hiperparâmetros escolhidos para o modelo final são: proporção de subamostras de colunas igual 1 ao construir cada árvore; o número máximo de folhas em cada árvore igual a 5; o número de árvores igual a 100; e a proporção da subamostra da instância de treinamento igual a 0,9. Esses hiperparâmetros são selecionados por apresentarem maiores frequências nas 100 execuções, como mostrado na Figura 35. O tempo total para o treinamento, pesquisa em grade com *5-fold* juntamente com a classificação dos dados de teste, é de 1 hora 14 minutos e 50 segundos. Por ser um procedimento mais complexo, o custo computacional é bem maior do que uma classificação usando árvore de decisão comum.

Figura 32 – Frequência dos hiperparâmetros do modelo GBDT.



Fonte: elaborado pelo autor.

A acurácia média resulta em 84%, enquanto a sensibilidade alcança 80%, que não é um valor alto, visto que o objetivo é distinguir as diferentes classes, em especial, identificar o indivíduo com glaucoma. A acurácia máxima obtida é de 91% e sensibilidade 94% nos dados de teste. Apesar desse resultado ser considerado bom, não representa o conjunto completo de dados, identifica somente uma pequena parte dos dados estudados.

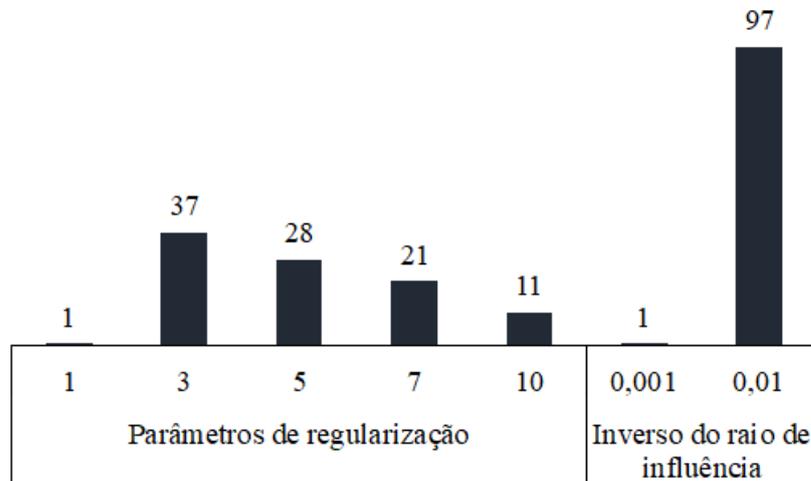
Os resultados obtidos com o GBDT são melhores do que os resultados utilizando o classificador k-NN. Contudo, são inferiores se comparados com a regressão logística. Conside-

rando o custo computacional, o GBDT não é tão eficaz para o problema em questão, já que a regressão logística retorna resultados mais precisos com um custo computacional menor. Porém, o GBDT é melhor do que o k-NN que, apesar de ser mais rápido, não produz bons resultados para a sensibilidade, ou seja, não consegue identificar pacientes com glaucoma. Mesmo o GBDT apresentando bons resultados, é preferível utilizar a RL que obtém resultados levemente mais altos, em um tempo menor.

5.1.4 Máquinas de Vetores de Suporte

Avaliando o classificador SVM, são comparados os hiperparâmetros γ igual a 0,001, 0,01, 0,1 e 1, que em 99 execuções é retornado $\gamma = 0,01$. O parâmetro de regularização é testado para diferentes valores, o que apresenta maior frequência é para $C = 3$. Para a função de ativação, testam-se os kernels: kernel Radial Basis Function (RBF), linear, logística, polinomial e sigmóide. Em todas as execuções, o kernel que apresenta melhor taxa de acurácia é o RBF.

Figura 33 – Frequência dos hiperparâmetros do modelo SVM.



Fonte: elaborado pelo autor.

As taxas médias de acurácia, sensibilidade, especificidade e AUC são as melhores encontradas até o momento, em comparação aos classificadores estudados. A acurácia média atinge 89%, sensibilidade 85% e especificidade 92%, indicando que, na maioria dos casos, as classes estão sendo distintas corretamente. Apesar disso, a taxa média de sensibilidade ainda é menor do que a taxa de especificidade. A menor acurácia obtida é 80% e a maior 97%, se estivesse sendo analisada somente uma execução há configurações de dados teste que resultariam

em 80% de acerto, enquanto outras configurações atingem 97% de acerto.

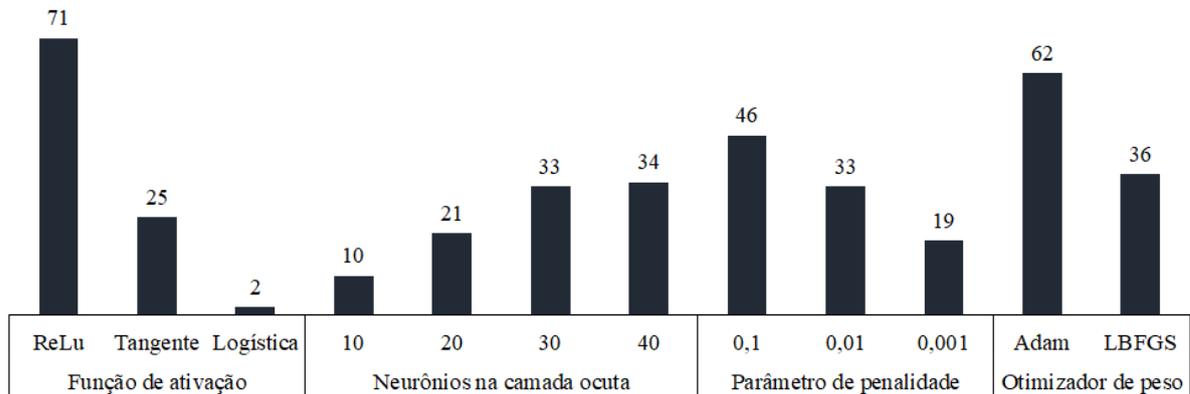
O SVM mostra ser um classificador bom para distinguir as diferentes classes do conjunto RIM-ONE v2, com as melhores taxas média de sensibilidade e especificidade. A variação de resultados em cada execução é baixa, isso pode ser visto pelo desvio padrão da acurácia que é igual a 3,33%, bem como desvio da sensibilidade também é o menor entre os classificadores.

Esse procedimento é considerado rápido e apresenta bom desempenho, pois até o momento é o melhor classificador para a base de imagens estudada. É importante ressaltar que, diversos trabalhos propõem métodos para identificação automática do glaucoma, no entanto, as bases de imagens são diferentes, e isso influencia nos resultados. Em algumas bases, o método abordado pode apresentar valores considerados bons, enquanto o mesmo método em outras bases pode não classificar corretamente. Desse modo, vale frisar que os resultados obtidos são considerados bons, como são os resultados médio de 100 execuções, essa metodologia se torna relevante também com outras bases de imagens, isto porque em cada execução são analisadas diferentes configurações de imagens da região de interesse, englobando imagens nítidas e também borradas.

5.1.5 *Perceptron Multicamadas*

Para o MLP, o algoritmo é mais complexo e demorado por ser uma rede neural. Para o parâmetro de regularização, a penalidade L1 sobressai-se a quantidade de L2, o valor $\alpha = 0,1$ destaca-se em 47 execuções enquanto $\alpha = 0,01$ e $\alpha = 0,001$ em 34 e 19, respectivamente. São testados diferentes neurônios para a camada oculta, a configuração que melhor se ajusta é utilizando apenas uma camada oculta com 40 neurônios. O total de épocas de treinamento é de 1000 épocas. Para a função de ativação da camada oculta, escolhe-se a ReLU, com frequência igual 71, a função de ativação logística apresenta a menor frequência, em apenas 2 execuções.

Figura 34 – Frequência dos hiperparâmetros do modelo MLP.



Fonte: elaborado pelo autor.

Os resultados são semelhantes aos encontrados pela regressão logística. A acurácia média resulta em 85% e sensibilidade 82%, as taxas são próximas comparando-as ao classificador k-NN. Como nos casos anteriores, as taxas de sensibilidade são menores do que a especificidade, nas taxas máximas é visto que a sensibilidade e especificidade estão altas e muito próximas. Isto sugere que algumas configurações dos dados de teste apresentam imagens mais nítidas, tornando o classificador mais preciso para identificação do glaucoma, reduzindo o problema de falso negativo. Ressaltando que as taxas máximas de sensibilidade e especificidade não são obtidas na mesma execução. Apesar desse classificador produzir resultados considerados bons, seu custo computacional é elevado, comparado ao SVM que produziu resultados melhores, e a regressão logística que apresenta desempenho semelhante.

5.1.6 Resultados do Modelo 1

Após o ajuste da pesquisa em grade, são selecionados os hiperparâmetros com maiores frequências para compor o modelo final 1, uma configuração resumida das escolhas dos hiperparâmetros são apresentados na Tabela 5. A forma de avaliação, para cada uma dessas métricas, utilizando 100 execuções é determinada para que o modelo proposto identifique o glaucoma em diferentes configurações de dados, obtendo resultados não tendenciosos e que represente o conjunto de dados considerando diferentes qualidades das imagens.

O tempo decorrido para extrair as características é de 2 minutos, 16 segundos e 61 milissegundos, e o tempo para o procedimento do PCA para redução de dimensionalidade dos dados consome 30 milissegundos. E para a execução de treino e teste das 100 execuções, o tempo decorre em 1,13 segundos, 1,53 segundos, 38,2 segundos, 1,49 segundos, e 23,9 segundos,

Tabela 5 – Resumo dos hiperparâmetros selecionados para o modelo 1.

Clas.	Hiperparâmetros
RL:	C = 0,1 ; Regularização L2
k-NN:	Distância euclidiana; k=3
GBDT:	proporção de subamostras de colunas = 1; n máximo de folhas = 5; número de árvores = 100; instância de treinamento = 0,9
SVM:	Kernel RBF; $\gamma = 0,01$; C=3
MLP:	Ativação ReLu, 40 neurônios na camada oculta; Otimizador Adam; $\alpha = 0,1$; 1000 épocas

Fonte: elaborado pelo autor.

para os algoritmos de classificação RL, k-NN, GBDT, SVM e MLP, respectivamente. Todos apresentam um processo que pode ser considerado rápido, cuja principal diferença entre os algoritmos de classificação está nos resultados de acurácia e sensibilidade, e ainda assim estão próximos, apresentando um desempenho considerado satisfatório para a base de imagens testada.

Para todos os classificadores, as taxas de treino permanecem maiores do que as taxas de teste, sendo comum isto ocorrer, visto que o classificador identificou anteriormente, na etapa treinamento, padrões dessas imagens. Todavia, uma diferença nas taxas de treino e teste pode ser indício de falta de ajuste (*underfitting*) ou ajuste excessivo. No entanto, esse pressuposto não pode ser tido como premissa.

Os resultados de teste dos classificadores regressão logística e MLP apresentam valores próximos. Mas o MLP obtém taxas levemente maiores na acurácia e especificidade, enquanto a regressão logística é maior na sensibilidade. O GBDT também permanece próximo dos valores da regressão logística, a principal diferença está na sensibilidade, menor para o GBDT, já a especificidade do GBDT se encontra levemente maior do que a regressão logística. O k-NN é o classificador que apresenta desempenho menor, com mais falsos negativos, mostrado através dos valores de taxa média de sensibilidade, enquanto a especificidade está com valores maiores. Esse padrão é semelhante ao GBDT, no entanto, o GBDT não apresenta essa diferença tão grande, em relação as taxas de sensibilidade e especificidade. Enquanto a diferença entre a sensibilidade e especificidade varia em 6% para o GBDT, para o k-NN varia 20%.

O SVM destaca-se entre os classificadores estudados, por apresentar maiores taxas média. Sua taxa média de acurácia atinge 89%, a sensibilidade é a maior obtida com 86% e a especificidade entre as maiores com 91%. O desvio padrão entre os classificadores não apresenta diferença estatisticamente significativa, também obtém menor desvio padrão em quase todas as métricas avaliadas, exceto para a especificidade, ainda assim mostrando ser um modelo equilibrado na distinção das classes estudadas.

O classificador que obtém maior desvio padrão para as métricas avaliadas é o k-NN, exceto para a especificidade. Esse algoritmo classifica, na maioria dos casos, os pacientes como saudáveis, produzindo muitos falsos negativos. Os resultados para esse algoritmo mostram taxas menores de acurácia e sensibilidade em comparação aos demais classificadores. A Tabela 6 indica as taxas médias das 100 execuções de acurácia, sensibilidade, especificidade e AUC \pm o desvio padrão para cada classificador, destacando-se em negrito o SVM que obteve maiores taxas em relação aos demais.

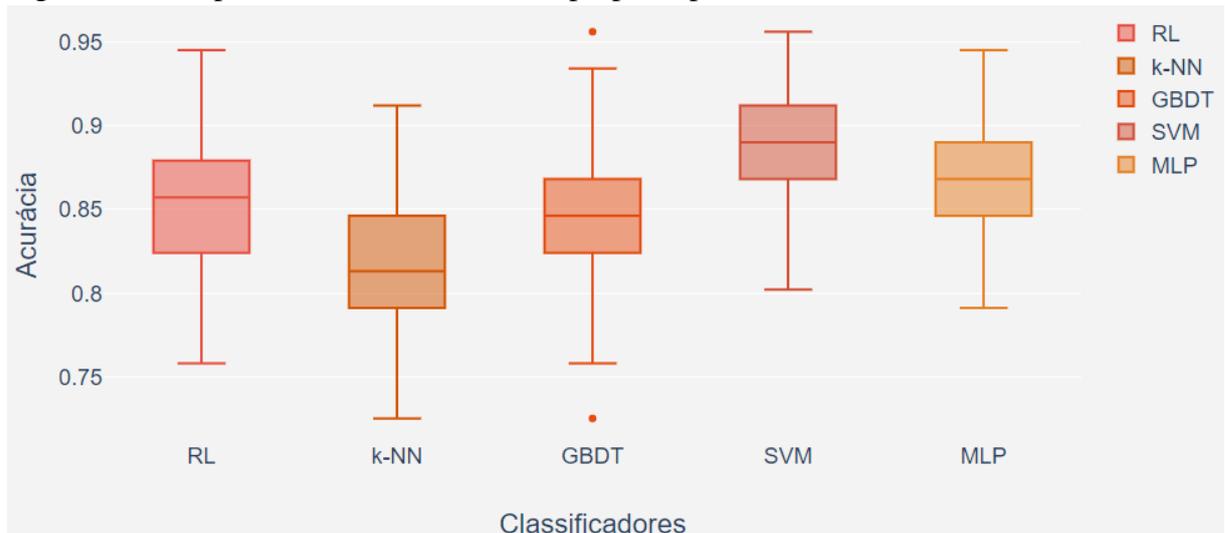
Tabela 6 – Desempenho médio dos dados de teste referente ao modelo 1.

Classificadores	Acurácia (%)	Sensibilidade (%)	Especificidade (%)	AUC (%)
RL	85,35 \pm 3,39	84,25 \pm 5,38	86,29 \pm 4,89	85,13 \pm 3,55
kNN	81,52 \pm 3,81	70,25 \pm 7,27	90,60 \pm 3,45	82,42 \pm 3,40
GBDT	84,60 \pm 3,26	81,51 \pm 5,67	87,16 \pm 4,83	84,49 \pm 3,41
SVM	89,03 \pm 2,88	86,59 \pm 4,98	91,06 \pm 4,20	88,95 \pm 3,01
MLP	86,35 \pm 3,52	83,33 \pm 5,73	88,77 \pm 4,54	86,17 \pm 3,67

Fonte: elaborado pelo autor.

É possível visualizar a dispersão dos resultados de acurácia através do boxplot mostrado na Figura 35. A regressão logística apresenta comportamento é simétrico, com a mediana próximo da taxa média. Já o k-NN mostra uma dispersão de acurácia maior, porém, não apresenta nenhum valor discrepante, sua mediana está mais distante do valor médio. E, pelo comportamento dos quartis, observa-se uma assimetria negativa, indicando valores concentrados à esquerda, acima da mediana. Para o GBDT que obtém a maior variação, apresentando um valor discrepante acima do limite superior e um abaixo do limite inferior, seu comportamento é simétrico e apresenta dispersão maior. Já o SVM é o classificador com a menor dispersão, apresentando limite superior mais alto dentre os demais, os dados são assimétricos positivo com valores concentrados abaixo da mediana. E por último, o MLP que apresenta comportamento semelhante ao SVM, no entanto seus valores de limites, quartis e mediana são menores.

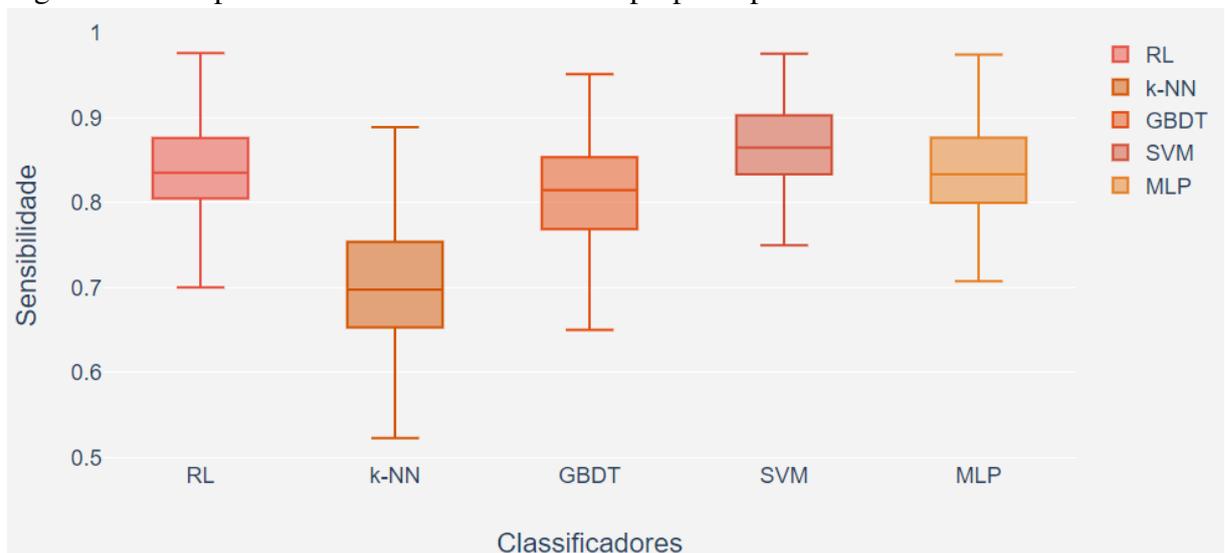
Figura 35 – Boxplot da acurácia do modelo proposto para cada classificador avaliado.



Fonte: elaborado pelo autor.

O *boxplot* da sensibilidade para cada classificador é mostrado na Figura 36. O k-NN apresenta valores mais baixos de sensibilidade, enquanto os demais permanecem com valores altos próximos uns aos outros. Em geral, a dispersão é pequena e apenas o SVM e MLP não apresentam dados discrepantes. A RL e SVM são simétricos, enquanto o k-NN e GBDT são levemente assimétricos a direita, com valores mais concentrados antes da mediana. E o MLP é assimétrico negativo, com valores concentrados além da mediana. Portanto, o SVM foi o melhor classificador para sensibilidade, ou seja, identificar o glaucoma, pois este apresenta uma distribuição simétrica e a dispersão baixa entre as diferentes qualidades de imagens.

Figura 36 – Boxplot da sensibilidade do modelo proposto para cada classificador avaliado.



Fonte: elaborado pelo autor.

A especificidade é mostrada na Figura 37, que é possível verificar que o k-NN apresenta a menor dispersão, com valores concentrados. A RL é o único classificador com valor discrepante abaixo no limite inferior. O comportamento do SVM não apresenta nenhum valor discrepante e os limites inferiores e superiores são um pouco maiores do que o k-NN. O MLP foi o classificador com mais valores discrepantes, tanto abaixo do limite inferior como acima do limite superior. E por fim, o GBDT apresenta valores abaixo do SVM com a mediana próximo ao primeiro quartil do SVM, não é observado nenhum valor discrepante.

Figura 37 – Boxplot da especificidade do modelo proposto para cada classificador avaliado.



Fonte: elaborado pelo autor.

5.2 Modelo 2

As análises são realizadas utilizando somente as informações geométricas, no entanto, os resultados obtidos são baixos, com a média de acurácia em torno de 60% a 70%. Devido a isso, é preferível adicionar essas 3 características aos descritores do modelo 1, pois, os resultados da classificação tendem a melhorar. Portanto, o modelo 2 consiste na adição das características geométricas aos descritores do modelo 1. Como o banco de dados em estudo possui o padrão ouro para a segmentação do disco óptico, é possível verificar que, para 90% das imagens, o método de segmentação do disco óptico contornou a região de forma correta. Já para a escavação não é possível avaliar, pois, o seu padrão ouro não está disponível.

Posteriormente, aplica-se a PCA para redução de dimensionalidade, pois este método permite separar as informações importantes das imagens, das informações redundante e aleatória. No modelo 1, a quantidade de característica é igual a 95, no modelo 2, com a adição das 3

informações geométricas, o total de característica passa a ser igual a 98. Após a aplicação do PCA, a quantidade de componentes torna-se igual a 44, ou seja, 44 componentes representam 95% da variância explicada.

O procedimento para otimização dos hiperparâmetros é realizado novamente para verificar se os hiperparâmetros continuam o mesmo ou se houve modificação devido o acréscimo dessas 3 características. A seguir são mostrados os hiperparametros obtidos pelo método de pesquisa em grade com validação cruzada *5-fold*.

5.2.1 Regressão logística

Para a regressão logística, os hiperparâmetros selecionados por apresentarem maiores frequências nas execuções pelo método da pesquisa em grade são, regularização L2 e o parâmetro C, inverso da regularização, igual a 0,1.

5.2.2 k Vizinhos mais próximos

Considerando o k-NN, a distância escolhida foi a euclidiana com peso uniforme entre as observações. A quantidade de vizinhos escolhido é igual 5.

5.2.3 Árvore de decisão com aumento de gradiente

Para o classificador GBDT, a proporção de subamostras de colunas ao construir cada árvore é igual a 1. O número máximo de folhas em cada árvore igual a 5, e os valores de proporção da subamostra da instância de treinamento igual a 0,6.

5.2.4 Máquinas de vetores de suporte

Para o classificador SVM, o parâmetro de penalidade do termo de erro é definido igual a 3, com kernel RBF e coeficiente γ igual a 0,01.

5.2.5 Perceptron Multicamadas

Considerando o MLP, a estrutura da rede é construída com uma única camada oculta com 40 neurônios, função de ativação é a ReLU, e regularização L2 com parâmetro α de penalidade igual a 0,01. O total de épocas de treinamento é igual a 1000.

Tabela 7 – Resumo dos hiperparâmetros selecionados para o modelo 2.

Clas.	Hiperparâmetros
RL:	C = 0,1 ; Regularização L2
k-NN:	Distância euclidiana; k=5
GBDT:	proporção de subamostras de colunas = 1; n máximo de folhas = 5; número de árvores = 100; instância de treinamento = 0,6
SVM:	Kernel RBF; $\gamma = 0,01$; C=3
MLP:	Ativação ReLu, 40 neurônios na camada oculta; Otimizador Adam; $\alpha = 0,1$; 1000 épocas

Fonte: elaborado pelo autor.

5.2.6 Resultados do modelo 2

Uma forma resumida dos hiperparâmetros selecionados no modelo 2 pode ser visualizada na Tabela 7.

As taxas médias obtidas para o modelo 2, nos diferentes classificadores, podem ser visualizadas na Tabela 8, assim como o modelo 1, no modelo 2 o SVM também foi o classificador que se destaca por apresentar as melhores taxas.

Tabela 8 – Desempenho médio dos dados de teste do modelo 2 referente a 100 execuções.

Classificadores	Acurácia (%)	Sensibilidade (%)	Especificidade (%)	AUC (%)
RL	84,51 ± 3,49	82,75 ± 6,02	85,91 ± 4,28	84,22 ± 3,61
KNN	80,74 ± 4,18	68,89 ± 7,94	90,22 ± 4,03	81,66 ± 3,96
GBDT	84,27 ± 3,39	81,17 ± 5,84	86,87 ± 5,17	84,18 ± 3,45
SVM	88,63 ± 2,79	84,81 ± 4,89	91,65 ± 3,99	88,66 ± 2,96
MLP	86,77 ± 3,58	82,70 ± 7,03	89,95 ± 4,80	86,88 ± 3,63

Fonte: elaborado pelo autor.

5.3 Modelo 3

As arquiteturas de aprendizado por transferência são utilizadas somente para extração das informações das imagens. As características obtidas pelos modelos CNN pré-treinados são classificadas diretamente sem a redução de dimensionalidade, porém os resultados apresentam valores médios de acurácia em torno de 40% a 60%. Devido a isso, opta-se por utilizar a redução de dimensionalidade para evitar multicolinearidade e otimizar o tempo - isto porque a quantidade de características se torna consideravelmente menor. A quantidade de componentes resultantes são de 186, 111, 113 e 114 para os modelos MobileNet, ResNet50, VGG16 e VGG19, respectivamente.

Novamente, são realizados os procedimentos de pesquisa de grade para obter os

melhores hiperparâmetros, foram avaliadas diversas configurações. A seguir são descritos, para cada arquitetura, os hiperparâmetros selecionados que apresentam maiores frequências nas 100 execuções, com base na acurácia dos dados de teste *k-fold*.

5.3.1 Regressão logística

Considerando a regressão logística, os hiperparâmetros de regularização selecionados são iguais em todas as estruturas de aprendizagem profundo: MobileNet, ResNet50, VGG16 e VGG19. A regularização é a L1, com inverso da força de regularização igual a 0,1.

5.3.2 *k* Vizinhos mais próximos

Para o *k*-NN, em todas as arquiteturas, a maior frequência é para a distância euclidiana com peso uniforme entre as observações. Em nenhum caso é escolhida a distância de Minkowski, apenas as distâncias euclidiana e Manhattan. A quantidade de vizinhos é igual a 5 para as arquiteturas MobileNet, VGG16 e VGG19, já para o ResNet50 a quantidade de vizinhos *k* é igual a 3.

5.3.3 Árvore de decisão com aumento de gradiente

Para o classificador GBDT, a proporção de subamostras de colunas ao construir cada árvore é de 1, 0,6, 0,6 e 0,9 para os classificadores MobileNet, ResNet50, VGG16 e VGG19, respectivamente. O número máximo de folhas em cada árvore igual 6 para o ResNet50, VGG16 e VGG19, já para o MobileNet é igual a 7. Os valores de proporção da subamostra da instância de treinamento igual a 0,7 para o MobileNet e VGG16, e igual a 0,6 para as arquiteturas VGG19 e ResNet50.

5.3.4 Máquinas de vetores de suporte

Para o classificador SVM, todas as arquiteturas apresentam, com maiores frequências, o parâmetro de penalidade do termo de erro definido igual a 3. E em todas as execuções, o kernel RBF foi o mais frequente, com o coeficiente γ igual a 0,001.

5.3.5 *Perceptron Multicamadas*

Considerando o MLP, todas as arquiteturas de aprendizado profundo apresentam apenas uma camada oculta. Para o VGG16 e VGG19 a função de ativação é a tangente hiperbólica, com 20 neurônios na camada oculta para o VGG19 e 30 para o VGG16. Para o MobileNet, a função de ativação é a ReLU com 20 neurônios na camada oculta, já para o ResNet50 a função de ativação é a logística com 10 neurônios na camada oculta. A regularização utilizada é a L2 com parâmetro de penalidade α igual a 0,01 para todas as arquiteturas.

5.3.6 *Resultados do modelo 3*

Esses procedimentos têm um custo computacional maior do que o método de extração de características previamente indicadas, contudo, as CNNs pré-treinadas não necessitam de especificar as características a serem extraídas. Os resultados são avaliados tanto com os classificadores quanto pelo método de descrição das características. Em média, o tempo para carregar as arquiteturas de CNNs pré-treinadas são descritas na Tabela 9.

Tabela 9 – Tempo para carregamentos de arquitetura e extração de características.

Arquitetura	Tempo
ResNet50	13 min, 58 s e 141 ms
MobileNet	5 min, 56 s e 114 ms
VGG16	17 min, 57 s e 212 ms
VGG19	19 min, 13 s e 170 ms

Fonte: elaborada pelo autor

Considerando a arquitetura ResNet50, nas classificações, exceto o k-NN, são produzidas acurácias médias maiores do que 80%, e as taxas de sensibilidade e especificidade são semelhantes entre as arquiteturas VGG16 e VGG19. O SVM destaca-se positivamente, para esta arquitetura, por apresentar o maior desempenho de acurácia.

De acordo com as taxas médias obtidas, a arquitetura MobileNet não é apropriada para identificar características que distinguem indivíduos com glaucoma dos indivíduos normais. Isto porque, para este modelo, nenhum classificador atinge resultado médio de 80%. O maior valor é com o SVM, no entanto, a acurácia média é igual a 77%, valor abaixo dos resultados obtidos em outras arquiteturas com o mesmo classificador. Já o algoritmo k-NN apresenta a menor acurácia, todavia, a taxa de especificidade é consideravelmente maior do que a taxa de

sensibilidade. Em todos os classificadores, a sensibilidade apresenta valor médio pequeno, ou seja, o classificador não consegue identificar com precisão os pacientes doentes. A RL e MLP obtêm desempenhos semelhantes, com valores de métricas próximos, ainda assim, os resultados não podem ser considerados bons.

A arquitetura VGG16 apresenta bons resultados, em destaque para os classificadores SVM e MLP, os quais apresentaram as melhores taxas médias de acurácia. Dentre os classificadores, a maior sensibilidade média é para o MLP e maior especificidade média para o SVM, ambos considerando a arquitetura VGG16. Já a arquitetura VGG19 obtém valores próximos à VGG16, no entanto, são levemente menores, assemelhando-se ao modelo ResNet50. O maior valor médio de acurácia é para o MLP atingindo 81,88%, seguido do RL e SVM, ambos com acurácia igual a 81%. A taxa de especificidade entre esses classificadores também permanece próximas.

Todas as arquiteturas, exceto MobileNet, apresentam taxas médias próximas, entretanto, a que apresenta maior taxa de acurácia e sensibilidade é a VGG16 com classificador SVM. Esta arquitetura também apresenta a maior especificidade, neste caso, com o classificador MLP. Desse modo, dentre as arquiteturas estudadas, a VGG16 é a que apresenta melhor desempenho. Ainda assim, esses resultados não são melhores do que os valores obtidos através do modelo 1.

Logo, a extração de características pelo método de aprendizado por transferência não é tão eficiente quanto pelo método de extração usado pelo modelo 1, para distinguir as classes de 'glaucoma/suspeito de glaucoma' e saudáveis. Isto ocorre porque a semelhança entre as imagens confunde a classificação, e as características obtidas por estes modelos não são favoráveis para a separação das classes.

Como as imagens são semelhantes, torna-se complexo especificar as características importantes para o banco de dados em questão, pois, não é possível basear-se nas cores das imagens e nem na identificação de determinados objetos na imagem, por exemplo. A Tabela 10 apresenta os resultados médios finais para cada classificador nas diferentes estruturas. Estão contidos os valores médios de acurácia, sensibilidade, especificidade e AUC, e seus respectivos desvios padrão.

Tabela 10 – Desempenho médio dos dados de teste referente ao modelo 3.

CNN	Classificador	Acurácia (%)	Sensibilidade (%)	Especificidade (%)	AUC (%)
ResNet50	RL	81,76 ± 3,90	75,48 ± 6,39	86,81 ± 5,08	81,77 ± 4,12
	KNN	72,84 ± 4,61	55,19 ± 7,52	86,98 ± 5,29	74,07 ± 4,93
	GBDT	80,07 ± 4,38	72,97 ± 6,90	85,92 ± 5,21	80,17 ± 4,42
	SVM	83,04 ± 3,80	74,02 ± 7,10	90,23 ± 4,33	83,58 ± 3,80
	MLP	81,10 ± 4,02	75,42 ± 7,21	85,43 ± 5,67	81,07 ± 4,17
MobileNet	RL	75,74 ± 4,37	66,89 ± 8,05	82,79 ± 5,82	75,77 ± 4,39
	KNN	68,49 ± 4,92	42,28 ± 8,22	89,47 ± 4,23	71,24 ± 4,74
	GBDT	72,38 ± 4,88	60,38 ± 8,10	82,21 ± 6,23	72,68 ± 5,12
	SVM	77,21 ± 4,49	61,83 ± 8,12	89,50 ± 4,67	78,59 ± 4,32
	MLP	75,04 ± 4,56	65,45 ± 7,09	82,72 ± 5,55	75,13 ± 4,76
VGG16	RL	81,35 ± 3,64	73,47 ± 6,41	87,42 ± 4,79	81,47 ± 3,85
	KNN	72,80 ± 4,39	51,60 ± 8,36	89,86 ± 4,30	75,04 ± 4,40
	GBDT	78,80 ± 4,04	67,72 ± 7,83	87,91 ± 4,86	79,45 ± 4,12
	SVM	83,51 ± 3,69	74,20 ± 6,98	90,89 ± 3,90	84,11 ± 3,59
	MLP	83,42 ± 3,48	78,16 ± 5,39	87,78 ± 4,23	83,42 ± 3,58
VGG19	RL	81,36 ± 4,38	73,26 ± 6,94	87,73 ± 4,93	81,57 ± 4,51
	KNN	72,97 ± 4,26	57,23 ± 7,27	85,62 ± 4,87	73,71 ± 4,53
	GBDT	78,73 ± 4,26	68,80 ± 8,24	87,06 ± 5,47	79,29 ± 4,41
	SVM	81,24 ± 4,10	71,46 ± 7,34	89,00 ± 4,27	81,71 ± 4,08
	MLP	81,88 ± 4,29	74,76 ± 7,68	87,34 ± 5,15	81,96 ± 4,28

Fonte: elaborado pelo autor.

5.4 Discussões dos resultados

O modelo 1 com a extração de características previamente indicadas apresentam melhor desempenho em comparação aos modelos 2 e 3. Os resultados do modelo 2 são levemente menores do que o modelo 1, ou seja, o acréscimo das características geométricas não auxilia na classificação das classes de imagens, a adição dessas características acaba confundindo o classificador na etapa de classificação de imagens, dificultando a identificação correta das classes. Apesar das taxas médias estarem bastante próximas, é preferível escolher o modelo 1, pois, este não necessita do processo de segmentação da região do disco óptico e escavação, tornando assim o processo mais rápido computacionalmente. Na etapa de segmentação do disco óptico, o algoritmo apresentou 90% de acurácia para detecção das bordas, e para a escavação não foi possível computar, pois o banco de imagens não fornece o padrão ouro.

Já no modelo 3, apesar de apresentar resultados considerados satisfatórios, as taxas médias são menores do que os dois modelos anteriores, e o processo é mais lento. Além disso, a classificação atinge taxas menores do que o modelo 1. Portanto, de acordo com os objetivos

propostos, é possível verificar que, dentre os modelos estudados, o melhor procedimento para extração de características foi através do modelo 1, em que se utilizam de descritores previamente informados para extrair características das imagens. Apesar do modelo apresentar resultados próximos aos obtidos pelo modelo 1, é evidente que a inclusão de características geométricas, obtidas a partir da segmentação de imagens, não auxiliou na identificação do glaucoma, no entanto, este resultado pode estar relacionado com o desempenho da segmentação das imagens, que acarretaria em valores errôneos das informações geométricas.

De forma geral, o modelo apresenta vantagens como os maiores valores médios de acurácia e também pelo tempo de extração de características das imagens, menor do que o tempo para extração de características utilizando redes convolucionais, como no modelo 3. Apesar disso, a desvantagem do modelo 1 é devido a escolha dos extratores de características, como é um procedimento manual é importante escolher extratores que possam descrever informações relevantes das imagens, e por isso, é necessário uma escolha apropriada desses extratores, além de testes exaustivos para escolher os parâmetros dos mesmos.

Já o modelo 2, a vantagem é pela inclusão das características geométricas, que contém informação da progressão do glaucoma. No entanto, a desvantagem é pela segmentação das imagens, como as imagens possuem qualidades diferentes, o algoritmo de segmentação não se torna tão preciso e isso pode ocasionar em valores inexatos, o que implica em uma classificação errônea.

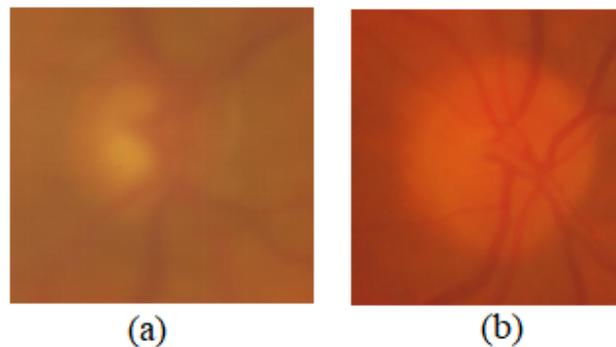
E o modelo 3, a principal vantagem do uso de redes neurais convolucionais é porque o algoritmo não necessita da indicação de extratores para obter as características das imagens. E o uso de arquiteturas pré-treinadas auxilia em relação ao tempo de processamento, sendo este menor do que o tempo de CNNs com pesos iniciais aleatórios. A desvantagem de tal técnica é que, apesar de se trabalhar com imagens, nem todas as arquiteturas pré-treinadas apresentarão um bom desempenho para a base estudada, isto porque quanto mais profunda a rede, mais informações específicas das imagens o algoritmo terá, e como as imagens são diferentes das imagens do banco ImageNet, isto pode ocasionar em muitos falsos positivos ou falsos negativos, confundindo a classificação correta entre as classes. Ainda assim, há algumas técnicas que podem reverter tal problema como utilizar somente as configurações iniciais da rede.

Na etapa de classificação, os hiperparâmetros foram fundamentais para obter um resultado mais preciso no conjunto de dados, isto porque se não houvesse essa escolha dos valores, os dados teriam uma modelagem menos precisa. Portanto, a escolha dos hiperparâmetros

para configurar o modelo auxiliou a classificação dos dados, distinguindo assim as classes. Com isso, os classificadores apresentaram um desempenho satisfatório para a distinção entre as classes, todos possuem um custo computacional baixo, já que a execução de treinamento e teste, após a otimização dos hiperparâmetros, pode ser realizada em menos de 5 minutos. O classificador que se destaca de forma positiva é o SVM, apresentando um desempenho considerado bom, em todos os modelos estudados. Enfatizando também a importância do PCA para redução de dimensionalidade, em que após a aplicação do PCA os resultados apresentaram uma melhora considerável em todos os classificadores, aumentando as taxas médias em pelo menos 10%, aperfeiçoando os três modelos analisados.

A qualidade das imagens prejudica o classificador na separação de classes, ocorrendo assim o erro no momento de identificar a qual classe a imagem pertence. Algumas características são responsáveis pela classificação errada, tais como: imagens bastante borradas, intensidades de cores próximas e dificuldade de distinção entre disco e escavação. Desse modo, quanto melhor a qualidade e o ângulo de obtenção da imagem, mais eficiente deve ser a extração de características relevantes para a região do nervo óptico, como bordas e contornos, que caracterizam as diferenças entre as imagens. A Figura 38 (a) mostra um disco óptico com glaucoma classificado erroneamente - a imagem está borrada e os contornos e bordas de difícil distinção - um falso negativo, o teste indica que o paciente está saudável, quando na realidade está com glaucoma. Assim como na Figura 38 (b), um falso positivo, ou seja, o teste indica que o indivíduo está doente, mas na realidade é saudável.

Figura 38 – Classificação errônea de um falso negativo (a) e um falso positivo (b).



Fonte: RIM-ONE-r2

Doravante, são comparados os resultados obtidos pelo modelo 1 - método proposto com os trabalhos relacionados na fundamentação teórica, conforme apresentado na Tabela 11.

Os resultados obtidos nesta dissertação apresentam taxas menores do que muitos trabalhos mencionados, ainda assim, esses valores estão próximos das taxas alcançadas nos trabalhos que utilizam a mesma base de imagens. A dificuldade em comparar os resultados com os resultados de outros trabalhos é devido a diferença entre as bases de imagens, principalmente nos trabalhos que utilizam imagens privadas, bem como em relação as diferentes técnicas de avaliação dos resultados.

Os trabalhos que utilizam a mesma base recorrem a diferentes técnicas de avaliação, como Araújo *et al.* (2017), que utiliza a média de apenas 5 execuções. Isto não é suficiente para generalizar a classificação das imagens, pois, se na presente dissertação fossem utilizadas somente 5 execuções a taxa média de acurácia resultaria em 92%. Sousa *et al.* (2017) utilizam uma única execução com a técnica pesquisa em grade de validação cruzada *5-fold*. Já em Samanta *et al.* (2015), a quantidade de treino é de 90%, 5% para validação e 5% para teste, ainda assim, não utilizam todas as imagens do banco de dados, selecionam apenas 321 imagens de forma aleatória para proceder com a classificação.

Na literatura pesquisada, outras bases de imagens são utilizadas, em sua maioria são bases de imagens privadas, por esta razão, não é possível comparar esses resultados com a base pública empregada nesta dissertação, pois, a qualidade das imagens difere. Na maioria desses trabalhos, a avaliação dos resultados é realizada com a técnica de validação cruzada *k-fold*. Alguns trabalhos alcançam acurácia mais alta, como é o caso de Maheshwari *et al.* (2016), que utiliza informações de correntropia e classifica com SVM de mínimos quadrados; e o trabalho desenvolvido por CARVALHO JUNIOR *et al.* (2017) que utiliza aprendizado por transferência de uma VGG6 para extrair e classifica as imagens em um único processo.

O método avaliado, representado pelo modelo 1, nesta dissertação alcança resultados promissores para o conjunto RIM-ONE v2, apesar de muitos resultados dos trabalhos relacionados serem maiores, as taxas de acertos são semelhantes e até maiores do que alguns trabalhos que utilizam a mesma base de imagens. Além disso, a metodologia desta dissertação pode ser considerada rápida, o que possibilita sua aplicação numa situação real. Frisando ainda que a forma de avaliação utilizada generaliza a base de imagens, englobando diferentes qualidades de imagens que a base apresenta.

Tabela 11 – Comparação do desempenho do método proposto com outros métodos.

Autores	Dados	Qtd	Método	A (%)	S (%)	E (%)
Acharya <i>et al.</i> (2015)	Kasturba Medical College	510	SVM	93,10	89,75	96,20
Septiarini <i>et al.</i> (2018)	Dr. YAP Eye and Dr. Sardjito Hospital	84	K-NN	95,24	95,12	95,35
Ghosh <i>et al.</i> (2015)	RIM-ONE	169	BPNN	87,47	87,50	87,45
Samanta <i>et al.</i> (2015)	RIM-ONE v2	455	BPNN	96,16	99,51	90,44
Balsubramanian <i>et al.</i> (2016)	Não informado	30	SVM	83,30	75,00	-
Singh <i>et al.</i> (2016)	Venu Eye Research Center	63	K-NN	94,75	100,00	90,91
Araújo <i>et al.</i> (2017)	RIM-ONE v2	455	SVM	88,35	84,50	91,37
Sousa <i>et al.</i> (2017)	RIM-ONE v2	455	SVM	91,20	95,00	88,20
Koh <i>et al.</i> (2017)	Kasturba Medical College	1486	RF	92,48	89,37	95,58
Maheshwari <i>et al.</i> (2016)	RM-ONE v2 + Kasturba Medical College	515	SVM	98,33	100,00	96,67
Agarwal <i>et al.</i> (2015)	Não informado	60	SVM	90,00	100,00	80,00
Haleem <i>et al.</i> (2016)	RIM-ONE	124	SVM Twin	94,40	92,20	95,30
Salam <i>et al.</i> (2016)	-	100	SVM	91,00	100,00	87,00
Kausu <i>et al.</i> (2018)	Venu Eye Research Centre	86	MLP	97,67	98,00	97,10
Carvalho Júnior <i>et al.</i> (2017)	RIM-ONE v1	169	Decision Table	97,50	97,90	100,00
Al-Bander <i>et al.</i> (2017)	RIM-ONE v2	455	SVM	88,20	85,00	90,80
Abbas (2017)	DRIONS-DB, HRF e PVR	1200	CNN	84,50	98,01	99,00
Lima <i>et al.</i> (2018)	RIM-ONE v2	455	CNN	91,00	-	-
Cerentini <i>et al.</i> (2018)	RIM-ONE v1	158	CNN	94,20	-	-
Ahn <i>et al.</i> (2018)	Kim's Eye Hospital	1542	CNN	87,90	-	-
Zhen <i>et al.</i> (2018)	Beijing Tongren Hospital	5978	CNN	75,50	-	-
Gómez-Valverde <i>et al.</i> (2019)		2313	VGG19	88,05	87,01	89,01
Claro <i>et al.</i> (2018)		873	Random Forest	91,06	-	-
Ghamdi <i>et al.</i> (2019)			VGG16	92,40	91,70	93,30
Método proposto	RIM-ONE v2	455	SVM	89,03	86,59	91,06

Fonte: elaborado pelo autor.

6 CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS

De acordo com os resultados, o modelo mais eficiente, por seu tempo computacional reduzido e melhores taxas médias, é o modelo 1 que utiliza os descritores LBP, HOG, momentos de Zernike e informações dos filtros de Gabor. Este modelo revela ser eficiente para descrever as imagens do conjunto RIM-ONE v2. Ademais, como classificador, o SVM destaca-se por apresentar os maiores valores, alcançando acurácia média igual a 89,33%, sensibilidade igual a 85,74% e especificidade igual a 92,21%. Apesar dos falsos negativos, o SVM mostra ser um classificador adequado para identificação de pacientes de acordo com sua classe, 'glaucoma/suspeita de glaucoma' ou saudáveis.

Esta dissertação tem como principal contribuição a avaliação de técnicas de classificação automática de glaucoma em imagens de fundoscopia, na qual analisam-se descritores apropriados para identificação do glaucoma, apresentam técnicas de segmentação do fundo de olho, e mostra alternativas, modelos de aprendizado por transferência, para extração de informações das imagens. Além disso, a avaliação é eficiente por englobar diferentes qualidades de imagens, indicando que a aplicação do método pode ser apropriada em outras bases de imagens de fundoscopia.

Como produto do método proposto, foram desenvolvidos dois trabalhos completos submetidos e aceitos para anais de eventos:

- "Avaliação de Métodos de Classificação de Glaucoma em Imagens do Fundo do Olho", no Simpósio Brasileiro de Automação Inteligente - SBAI (2019); e
- "Métodos de Classificação de Glaucoma em Imagens do Disco Óptico", no Congresso Brasileiro de Inteligência Computacional - CBIC (2019).

Para trabalhos futuros devem ser investigados:

- estudar outras técnicas de extração de características;
- aprimorar a segmentação a fim de obter resultados mais precisos das características geométricas;
- desenvolver um pré-processamento adequado para melhorar o contraste das imagens;
- estudar e avaliar diferentes arquiteturas de CNNs para a classificação automática do glaucoma;
- estudar e avaliar outras técnicas de redução de dimensionalidade, tais como Análise Linear Discriminante; e
- utilizar outras bases públicas de imagens.

REFERÊNCIAS

- ABBAS, Q. Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning. **International Journal of Advanced Computer Science and Applications (IJACSA)**, v. 8, n. 6, p. 41–5, 2017.
- ABRÀMOFF, M. D.; GARVIN, M. K.; SONKA, M. Retinal imaging and image analysis. **IEEE reviews in biomedical engineering**, IEEE, v. 3, p. 169–208, 2010.
- ACHARYA, U. R.; NG, E.; EUGENE, L. W. J.; NORONHA, K. P.; MIN, L. C.; NAYAK, K. P.; BHANDARY, S. V. Decision support system for the glaucoma using gabor transformation. **Biomedical Signal Processing and Control**, Elsevier, v. 15, p. 18–26, 2015.
- AGARWAL, A.; GULIA, S.; CHAUDHARY, S.; DUTTA, M. K.; TRAVIESO, C. M.; ALONSO-HERNÁNDEZ, J. B. A novel approach to detect glaucoma in retinal fundus images using cup-disk and rim-disk ratio. In: 2015 4TH INTERNATIONAL WORK CONFERENCE ON BIOINSPIRED INTELLIGENCE (IWOB). **Proceedings [...]**. San Sebastian, Spain: IEEE, 2015. p. 139–144.
- AHN, J. M.; KIM, S.; AHN, K.-S.; CHO, S.-H.; LEE, K. B.; KIM, U. S. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. **PloS one**, Public Library of Science, v. 13, n. 11, p. e0207982, 2018.
- AL-BANDER, B.; AL-NUAIMY, W.; AL-TAEE, M. A.; ZHENG, Y. Automated glaucoma diagnosis using deep learning approach. In: 14TH INTERNATIONAL MULTI-CONFERENCE ON SYSTEMS, SIGNALS & DEVICES (SSD). **Proceedings [...]**. Marrakech, Morocco: IEEE, 2017. p. 207–210.
- ALMEIDA, A.; BORGES, C.; PAULA, I. Aplicação do descritor hog e classificador svm no reconhecimento de poses humanas em imagens de profundidade. In: ANAIS DA IV ESCOLA REGIONAL DE INFORMATICA DO PIAUI. **Anais [...]**. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação (SBC), 2018. p. 119–124. Disponível em: <<https://sol.sbc.org.br/index.php/eripi/article/view/5154>>.
- ALVARENGA JÚNIOR, W. J. de. **Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária**. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 2018.
- AQUINO, J. H. N. D. **Extração de características de imagens para classificação da qualidade de couro caprino usando padrão binário local**. 63 p. Dissertação (Mestrado acadêmico em ciências da computação) — Universidade Federal do Ceará, Fortaleza, 2017.
- ARAÚJO, J. D. L.; PAIVA, A. C. de; ALMEIDA, J. D. S.; NETO, O. P. S.; SOUSA, J. A.; SILVA, A. C.; BRAZ JÚNIOR, G. Diagnóstico de glaucoma em imagens de fundo de olho utilizando os Índices de diversidade de shannon e mcintosh. In: ANAIS PRINCIPAIS DO XVII WORKSHOP DE INFORMATICA MEDICA. **Anais [...]**. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação (SBC), 2017. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/3698>>.
- BALASUBRAMANIAN, T.; KRISHNAN, S.; MOHANAKRISHNAN, M.; RAO, K. R.; KUMAR, C. V.; NIRMALA, K. Hog feature based svm classification of glaucomatous fundus

image with extraction of blood vessels. In: ANNUAL INDIA CONFERENCE (INDICON). **Anais [...]**. Bangalore, India: IEEE, 2016. p. 1–4.

BEZERRA, C. S. **Uma abordagem de segmentação de íris para fins biométricos usando aprendizagem profunda**. Dissertação (Mestrado em Informática) — Programa de Pós-Graduação em Informática, Universidade Federal do Paraná (UFPR), Curitiba, 2019.

BONESSO, D. **Estimação dos parâmetros do kernel em um classificador SVM na classificação de imagens hiperespectrais em uma abordagem multiclasse**. Dissertação (Mestrado em Sensoriamento Remoto) — Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia. Universidade Federal do Rio Grande do Sul (UFRS), Porto Alegre, 2013.

BOYD, K. **Who Is Glaucoma?** American Academy of Ophthalmology, 2019. Disponível em: <<https://www.aao.org/eye-health/diseases/what-is-glaucoma>>. Acesso em: 09 abr. 2020.

BROWNLEE, J. **Logistic Regression for Machine Learning**. Machine Learning Mastery, 2016. Disponível em: <<https://machinelearningmastery.com/logistic-regression-for-machine-learning/>>. Acesso em: 03 out. 2019.

BROWNLEE, J. **A Gentle Introduction to k-fold Cross-Validation**. Machine Learning Mastery, 2018. Disponível em: <<https://machinelearningmastery.com/k-fold-cross-validation/>>. Acesso em: 03 out. 2019.

BROWNLEE, J. **A Gentle Introduction to Padding and Stride for Convolutional Neural Networks**. Machine Learning Mastery, 2019. Disponível em: <<https://machinelearningmastery.com/padding-and-stride-for-convolutional-neural-networks/>>. Acesso em: 29 dez. 2019.

BROWNLEE, J. **Transfer Learning in Keras with Computer Vision Models**. Deep Learning for Computer Vision, 2019. Disponível em: <<https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/>>. Acesso em: 07 out. 2019.

CARVALHO JUNIOR, A. S. V.; CARVALHO FILHO, A. O.; SOUSA, A.; BARROS, P. S. Desenvolvimento de métodos para detecção automática do glaucoma. In: ANAIS PRINCIPAIS DO XVII WORKSHOP DE INFORMATICA MEDICA. **Anais [...]**. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação (SBC), 2017. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/3709>>.

CERENTINI, A.; WELFER, D.; D'ORNELLAS, M. C.; HAYGERTB, C. J. P.; DOTTOB, G. N. Automatic identification of glaucoma sing deep learning methods u. In: MEDINFO 2017: PRECISION HEALTHCARE THROUGH INFORMATICS: PROCEEDINGS OF THE 16TH WORLD CONGRESS ON MEDICAL AND HEALTH INFORMATICS. **Proceedings [...]**. Hangzhou, China: IOS Press, 2018. v. 245, p. 318.

CHEPENKO, D. **Introduction to gradient boosting on decision trees with Catboost**. Toward Data Science, 2013. Disponível em: <<https://towardsdatascience.com/introduction-to-gradient-boosting-on-decision-trees-with-catboost-d511a9ccbd14>>. Acesso em: 03 out. 2019.

CLARO, M.; VERAS, R.; SANTANA, A. Metodologia para identificação de glaucoma em imagens de retina. In: ANAIS ESTENDIDOS DO XIX SIMPOSIO BRASILEIRO

DE COMPUTACAO APLICADA A SAUDE. **Anais [...]**. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação (SBC), 2019. p. 103–108. Disponível em: <https://sol.sbc.org.br/index.php/sbcas_estendido/article/view/6292>.

CLARO, M. L.; VERAS, R. d. M. S.; SANTANA, A. M.; VOGADO, L. H. S.; SOUSA, L. P. Diagnóstico de glaucoma utilizando atributos de textura e cnn's pré-treinadas. **Revista de Informática Teórica e Aplicada**, v. 25, n. 1, p. 82–89, 2018.

COELHO, L. **Glaucoma: sintomas, causas, tratamentos, tipos e prevenção**. Ativo Saúde, 2018. Disponível em: <<https://www.ativosaude.com/saude-dos-olhos/glaucoma/>>. Acesso em: 02 out. 2019.

CRUZ A. C. da; SCHWAB, F. A. **Caracterização de micro vestígios entomológicos por imagem: aplicação de redes neurais profundas**. Monografia (Graduação em Engenharia Eletrônica) — Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, 2019.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: **PROCEEDINGS [...]. Computer society conference on computer vision and pattern recognition (CVPR'05)**. San Diego, CA, USA: IEEE, 2005. v. 1, p. 886–893.

DERPANIS, K. G. Gabor filters. CiteSeer, 2007.

DLB. **Deep Learning Book**. Data Science Academy, 2019. Disponível em: <<http://deeplearningbook.com.br/>>.

FARIA, E. L. de. **Redes neurais convolucionais e máquinas de aprendizado extremo aplicadas ao mercado financeiro brasileiro**. Tese (Doutorado em Engenharia Civil) — Programa de Pós-Graduação em Engenharia Civil. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

FEIBEL, R. M.; BIGGER, J. F. Rubeosis iridis and neovascular glaucoma: evaluation of cyclocryotherapy. **American journal of ophthalmology**, Elsevier, v. 74, n. 5, p. 862–867, 1972.

FRASAO, G. **Doenças oculares: quais são, tratamento, diagnóstico e prevenção**. Ministério da Saúde, 2019. Disponível em: <<http://www.saude.gov.br/saude-de-a-z/doencas-oculares/746-saude-de-a-a-z>>. Acesso em: 20 dez. 2019.

FUMERO, F.; ALAYN, S.; SANCHEZ, J. L.; SIGUT, J.; GONZALEZ-HERNANDEZ, M. Rim-one: An open retinal image database for optic nerve evaluation. In: 2011 24TH INTERNATIONAL SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS (CBMS). **Proceedings [...]**. Bristol, UK: IEEE, 2011. p. 1–6.

GHAMDI, M. A.; LI, M.; ABDEL-MOTTALEB, M.; SHOUSA, M. A. Semi-supervised transfer learning for convolutional neural networks for glaucoma detection. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP). **Proceedings [...]**. Brighton, United Kingdom: IEEE, 2019. p. 3812–3816.

GHOSH, A.; SARKAR, A.; ASHOUR, A. S.; BALAS-TIMAR, D.; DEY, N.; BALAS, V. E. Grid color moment features in glaucoma classification. **Int J Adv Comput Sci Appl**, v. 6, n. 9, p. 1–14, 2015.

GÓMEZ-VALVERDE, J. J.; ANTÓN, A.; FATTI, G.; LIEFERS, B.; HERRANZ, A.; SANTOS, A.; SÁNCHEZ, C. I.; LEDESMA-CARBAYO, M. J. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. **Biomedical optics express**, Optical Society of America, v. 10, n. 2, p. 892–913, 2019.

GONZALEZ, R. C.; WOODS, R. C. **Processamento digital de imagens**. [S.l.]: Pearson, 2017. 1192 p. ISBN 978-0133356724.

GOUR, N.; KHANNA, P. Automated glaucoma detection using gist and pyramid histogram of oriented gradients (phog) descriptors. **Pattern Recognition Letters**, Elsevier, 2019.

GUPTA, N.; AUNG, T.; CONGDON, N.; DADA, T.; LERNER, F.; OLAWOYE, S.; RESNIKOFF, S.; WANG, N.; WORMALD, R. Diretrizes da ico para glaucoma: Diretrizes para tratamento ocular do glaucoma. tradução: Homero gusmão de almeida. **Conselho Internacional de Oftalmologia**, Arco International Languages, 2016.

HALEEM, M. S.; HAN, L.; HEMERT, J. V.; FLEMING, A.; PASQUALE, L. R.; SILVA, P. S.; SONG, B. J.; AIELLO, L. P. Regional image features model for automatic classification between normal and glaucoma in fundus and scanning laser ophthalmoscopy (slo) images. **Journal of medical systems**, Springer, v. 40, n. 6, p. 132, 2016.

HASSAN, M. ul. **VGG16 – Convolutional Network for Classification and Detection**. Popular networks, 2018. Disponível em: <<https://neurohive.io/en/popular-networks/vgg16/>>. Acesso em: 07 out. 2019.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. São Paulo, Brasil: Bookman Editora, 2007.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. **Proceedings [...]**. Las Vegas, NV, USA, 2016. p. 770–778.

HELENE, O.; HELENE, A. F. Alguns aspectos da óptica do olho humano. **Revista Brasileira de Ensino de Física**, v. 33, n. 3, p. 3312, 2011.

HEMALATHA, R.; THAMIZHVANI, T.; DHIVYA, A. J. A.; JOSEPH, J. E.; BABU, B.; CHANDRASEKARAN, R. Active contour based segmentation techniques for medical image analysis. **Medical and Biological Image Analysis**, BoD–Books on Demand, p. 17, 2018.

HENRIKSEN, J. J. 3d surface tracking and approximation using gabor filters. **South Denmark University**, v. 28, 2007.

HOOVER, A.; GOLDBAUM, M. Localização do nervo óptico em uma imagem da retina usando a convergência difusa dos vasos sanguíneos. **transações do IEEE em imagens médicas**, IEEE, v. 22, p. 951–958, 2003.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. The multiple logistic regression model. **Applied logistic regression**, John Wiley & Sons, New York, NY, v. 1, p. 25–37, 1989.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **CoRR**, abs/1704.04861, 2017. Disponível em: <<http://arxiv.org/abs/1704.04861>>. Acesso em: 06 de janeiro de 2020.

IMADA, R. N. **Reconhecimento de contorno de edifício em imagens de alta resolução usando os momentos complexos de Zernike**. Dissertação (Mestrado em Matemática Aplicada e Computacional) — Universidade Estadual Paulista Júlio de Mesquita Filho, São Paulo, 2014.

JANKOV, M.; MROCHEN, M.; SCHOR, P.; CHAMON, W.; SEILER, T. Frentes de ondas (wavefronts) e limites da visão humana parte 1: fundamentos. **Arquivos brasileiros de oftalmologia**, SciELO Brasil, v. 65, n. 6, p. 679–684, 2002.

KAUSU, T.; GOPI, V. P.; WAHID, K. A.; DOMA, W.; NIWAS, S. I. Combination of clinical and multiresolution features for glaucoma detection and its classification using fundus images. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 38, n. 2, p. 329–341, 2018.

KINGMAN, S. Glaucoma is second leading cause of blindness globally. **Bulletin of the World Health Organization**, SciELO Public Health, v. 82, p. 887–888, 2004.

KOH, J. E.; ACHARYA, U. R.; HAGIWARA, Y.; RAGHAVENDRA, U.; TAN, J. H.; SREE, S. V.; BHANDARY, S. V.; RAO, A. K.; SIVAPRASAD, S.; CHUA, K. C. *et al.* Diagnosis of retinal health in digital fundus images using continuous wavelet transform (cwt) and entropies. **Computers in biology and medicine**, Elsevier, v. 84, p. 89–97, 2017.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.

LIMA, A.; MAIA, L. B.; SANTOS, P. T. C. dos; JÚNIOR, G. B.; ALMEIDA, J. D. S. de; PAIVA, A. C. de. Evolving convolutional neural networks for glaucoma diagnosis. In: ANAIS PRINCIPAIS DO XVIII SIMPOSIO BRASILEIRO DE COMPUTAÇÃO APLICADA A SAUDE. **Anais [...]**. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação (SBC), 2018. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/3687>>.

MÄENPÄÄ, T.; PIETIKÄINEN, M. Texture analysis with local binary patterns. In: **Handbook of pattern recognition and computer vision**. [S.l.]: World Scientific, 2005. p. 197–216.

MAGALHAES, L. **Olhos**. Toda Matéria, 2019. Disponível em: <<https://www.todamateria.com.br/olhos/>>. Acesso em: 02 out. 2019.

MAHESHWARI, S.; PACHORI, R. B.; ACHARYA, U. R. Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images. **Journal of biomedical and health informatics**, IEEE, v. 21, n. 3, p. 803–813, 2016.

MALLICK, S. **Histogram of Oriented Gradients**. Learn OpenCV, 2016. Disponível em: <<https://www.learnopencv.com/histogram-of-oriented-gradients/>>. Acesso em: 03 out. 2019.

MARTINS, T. G. dos S.; COSTA, A. L. F. de A.; MARTINS, R. V.; MARTINS, E. N.; ALVES, M. R.; HELENE, O.; SCHOR, P. Modelo para o ensino da oftalmoscopia direta. **Revista Brasileira de Ensino de Física**, v. 36, n. 2, p. 2303, 2014.

MAYRINK, V. T. M. **Avaliação do algoritmo gradient boosting em aplicações de previsão de carga elétrica a curto prazo**. Dissertação (Mestrado em Modelagem Computacional) — Programa de Pós-graduação em Modelagem Computacional. Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, 2016.

MEDEIROS, A. R. C.; ARAÚJO, Y. B. d.; VIANNA, R. P. d. T.; MORAES, R. M. d. Modelo de suporte à decisão aplicado à identificação de indivíduos não aderentes ao tratamento anti-hipertensivo. **Saúde em debate**, SciELO Public Health, v. 38, p. 104–118, 2014.

MONTAGNER, I. dos S. **Resumo de Local Binary Patterns**. Blog STOA, 2014. Disponível em: <<https://social.stoa.usp.br/igordsm/blog?npage=3>>. Acesso em: 03 out. 2019.

MURRTHY, K. **Gabor Filter: A Practical Overview**. Computer Vision Tutorials, 2014. Disponível em: <<https://cvtuts.wordpress.com/2014/04/27/gabor-filters-a-practical-overview/>>. Acesso em: 03 out. 2019.

NAGPAL, A. **L1 and L2 Regularization Methods**. Towards Data Science, 2017. Disponível em: <<https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c>>. Acesso em: 27 dez. 2019.

NETTO, R.; DORA, J. M. S.; SCOCCO, C. A.; KIWTKO, S. Exame de fundo de olho. In: EDITORA, A. (Ed.). **Exame Clínico: Consulta rápida**. 2. ed. Porto Alegre, 2004. v. 1, p. 119–134.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. **Pattern recognition**, Elsevier, v. 29, n. 1, p. 51–59, 1996.

OJALA, T.; PIETIKÄINEN, M.; MENPÄÄ, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **IEEE Transactions on Pattern Analysis & Machine Intelligence**, IEEE, n. 7, p. 971–987, 2002.

OLIVEIRA V. K., M. **O que é glaucoma?** AbcMed, 2012. Disponível em: <<https://www.abc.med.br/p/308210/o+que+e+glaucoma.htm>>. Acesso em: 04 dez. 2019.

PAULA G. A. de. **Modelos de regressão: com apoio computacional**. [S.l.]: IME-USP São Paulo, 2004.

PECCIA, F. **Batch normalization in Neural Networks**. Towards Data Science, 2017. Disponível em: <<https://towardsdatascience.com/batch-normalization-in-neural-networks-1ac91516821c>>. Acesso em: 29 dez. 2019.

PEIXOTO, E. B. A.; ARAKI, H.; CENTENO, J. **Utilização do algoritmo SVM (support vector machine) e árvores de decisão para classificação de imagens de alta resolução do sensor ads-40**. Recife: VI Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, 2016.

PETERSON, L. E. K-nearest neighbor. **Scholarpedia**, v. 4, n. 2, p. 1883, 2009. Revision #137311.

PIETIKÄINEN, M. Local Binary Patterns. **Scholarpedia**, v. 5, n. 3, p. 9775, 2010. Revision #188481.

PINHEIRO, P. **Exame de fundo de olho - o que é e como é feito**. Md. Saúde, 2019. Disponível em: <<https://www.mdsaude.com/oftalmologia/exame-de-fundo-de-olho/>>. Acesso em: 02 out. 2019.

POMARI, T.; REZENDE, E.; RUPPERT, G.; BALIEIRO, F.; CARVALHO, T. Associando redes convolucionais e características de iluminação para detectar falsificações em imagens. **Conference on graphics, pattern and images, 31 (SIBGRAPI)**, 2018.

PORTELA, C.; AMORIM, F.; MONTEIRO, G.; MARTINS, J.; MONTENEGRO, M. **Tutorial de SVM**. LAMFO, 2017. Disponível em: <<https://lamfo-unb.github.io/2017/07/13/svm/>>. Acesso em: 03 out. 2019.

RAO, P. B.; PRASAD, D. V.; KUMAR, C. P. Feature extraction using zernike moments. **International Journal of Latest Trends in Engineering and Technology**, v. 2, n. 2, p. 228–234, 2013.

RAVARANI. **Misleading modelling: overfitting, cross-validation, and the bias-variance trade-off**. Cambridge Coding, 2016. Disponível em: <<https://cambridgecoding.wordpress.com/2016/03/24/misleading-modelling-overfitting-cross-validation-and-the-bias-variance-trade-off/>>. Acesso em: 30 jan. 2020.

REBOUÇAS FILHO P. P.; CORTEZ P. C. et al. Modelo de contorno ativo crisp adaptativo 2d aplicado na segmentação dos pulmões em imagens de tc do tórax de voluntários sadios e pacientes com enfisema pulmonar. **Revista Brasileira de Engenharia Biomédica**, Rio de Janeiro, RJ, v. 29, n. 4, p. 363–376, 2013.

REFOSCO, L. M.; FENSTENSEIFER, G.; VARGAS, J. A. Manejo e diagnóstico do glaucoma agudo de ângulo fechado na emergência. **Acta méd.(Porto Alegre)**, v. 35, p. 6–6, 2014.

RIZWAN, M. **VGG16 – Implementation Using Keras**. EngMRK, 2018. Disponível em: <<https://engmrk.com/vgg16-implementation-using-keras/>>. Acesso em: 26 dez. 2019.

ROBIN, X.; TURCK, N.; HAINARD, A.; TIBERTI, N.; LISACEK, F.; SANCHEZ, J.-C.; MÜLLER, M. proc: display and analyze roc curves. **R Package Version**, v. 1, 2014.

SAHOO, S. **Residual blocks — Building blocks of ResNet**. Towards Data Science, 2018. Disponível em: <<https://towardsdatascience.com/residual-blocks-building-blocks-of-resnet-fd90ca15d6ec>>. Acesso em: 25 jan. 2020.

SALAM, A. A.; KHALIL, T.; AKRAM, M. U.; JAMEEL, A.; BASIT, I. Automated detection of glaucoma using structural and non structural features. **Springerplus**, Springer, v. 5, n. 1, p. 1519, 2016.

SAMANTA, S.; AHMED, S. S.; SALEM, M. A.-M. M.; NATH, S. S.; DEY, N.; CHOWDHURY, S. S. Haralick features based automated glaucoma classification using back propagation neural network. In: PROCEEDINGS OF THE 3RD INTERNATIONAL CONFERENCE ON FRONTIERS OF INTELLIGENT COMPUTING: THEORY AND APPLICATIONS (FICTA) 2014. **Proceedings [...]**. Advances in Intelligent Systems and Computing: Springer, 2015. p. 351–358.

SEGRE, L. **Eye anatomy: A closer look at the parts of the eye**. All About Vision, 2019. Disponível em: <<https://www.allaboutvision.com/resources/anatomy.htm>>. Acesso em: 05 fev. 2020.

SEPTIARINI, A.; KHAIRINA, D. M.; KRIDALAKSANA, A. H.; HAMDANI, H. Automatic glaucoma detection method applying a statistical approach to fundus images. **Healthcare informatics research**, v. 24, n. 1, p. 53–60, 2018.

SHUTLER, J. **Complex Zernike moments**. 2015. Disponível em: <https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/SHUTLER3/node11.html>. Acesso em: 03 out. 2019.

- SILVA M. A. da. **Retina**. Info Escola, 2018. Disponível em: <<https://www.infoescola.com/visao/retina/>>. Acesso em: 02 out. 2019.
- SINGH, A.; DUTTA, M. K.; PARTHASARATHI, M.; UHER, V.; BURGET, R. Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. **Computer methods and programs in biomedicine**, Elsevier, v. 124, p. 108–120, 2016.
- SOUSA, J. A. de; PAIVA, A. C. de; ALMEIDA, J. D. S. de; SILVA, A. C.; JUNIOR, G. B.; GATTASS, M. Texture based on geostatistic for glaucoma diagnosis from fundus eye image. **Multimedia Tools and Applications**, Springer, v. 76, n. 18, p. 19173–19190, 2017.
- SOUZA, A. **Algoritmo SVM (Máquina de Vetores de Suporte) a partir de exemplos e código (Python e R)**. 2019. Disponível em: <<https://pessoalex.wordpress.com/2019/04/10/algoritmo-svm-maquina-de-vetores-de-suporte-a-partir-de-exemplos-e-codigo-python-e-r/>>. Acesso em: 03 out. 2019.
- SRIVASTAVA, T. **Introduction to k-Nearest Neighbors: A powerful Machine Learning Algorithm (with implementation in Python & R)**. Analytics Vidhya, 2018. Disponível em: <<https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>>. Acesso em: 03 out. 2019.
- SRIVASTAVA, T. **Boosting (1): Árvores de Decisão e Gradient Boosting**. DataLab, 2019. Disponível em: <<https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>>. Acesso em: 03 out. 2019.
- TAHMASBI, A.; SAKI, F.; SHOKOUHI, S. B. An effective breast mass diagnosis system using zernike moments. In: 17TH IRANIAN CONFERENCE OF BIOMEDICAL ENGINEERING (ICBME). **Proceedings [...]**. Isfahan, Iran: IEEE, 2010. p. 1–4.
- TEWARI, S. **CNN Architecture Series — VGG-16 with implementation(Part I)**. Medium, 2019. Disponível em: <<https://medium.com/datadriveninvestor/cnn-architecture-series-vgg-16-with-implementation-part-i-bca79e7db415>>. Acesso em: 25 jan. 2020.
- THAKUR, N.; JUNEJA, M. Clustering based approach for segmentation of optic cup and optic disc for detection of glaucoma. **Current Medical Imaging Reviews**, Bentham Science Publishers, v. 13, n. 1, p. 99–105, 2017.
- THAM, Y.-C.; LI, X.; WONG, T. Y.; QUIGLEY, H. A.; AUNG, T.; CHENG, C.-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. **Ophthalmology**, Elsevier, v. 121, n. 11, p. 2081–2090, 2014.
- TOROK, L.; CONCI, A. **Método de Otsu**. Instituto de Computação, Universidade Federal Fluminense (UFF). Niterói, RJ, 2016. Disponível em: <<http://www2.ic.uff.br/~aconci/OtsuTexto.pdf>>. Acesso em: 3 out. 2019.
- VARELLA, C. A. A. Análise de componentes principais. **Seropédica: Universidade Federal Rural do Rio de Janeiro**, 2008.
- VASCONCELOS, S. **Análise de Componentes Principais (PCA)**. 2008. Disponível em: <<http://www2.ic.uff.br/~aconci/PCA-ACP.pdf>>. Acesso em: 03 out. 2019.

- WANG, C.-F. **A Basic Introduction to Separable Convolutions**. Towards Data Science, 2018. Disponível em: <<https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728>>. Acesso em: 29 dez. 2019.
- WANG, X.; HAN, T. X.; YAN, S. An hog-lbp human detector with partial occlusion handling. In: 12TH INTERNATIONAL CONFERENCE ON COMPUTER VISION. **Proceedings [...]**. Kyoto, Japan: IEEE, 2009. p. 32–39.
- WANG, Y.; SEE, J.; PHAN, R. C.-W.; OH, Y.-H. Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. **PloS one**, Public Library of Science, v. 10, n. 5, p. e0124674, 2015.
- WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 2, n. 1-3, p. 37–52, 1987.
- WU, H.; XIN, M.; FANG, W.; HU, H.-M.; HU, Z. Multi-level feature network with multi-loss for person re-identification. **IEEE Access**, IEEE, v. 7, p. 91052–91062, 2019.
- YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. **Insights into imaging**, Springer, v. 9, n. 4, p. 611–629, 2018.
- YU, S.; XIAO, D.; FROST, S.; KANAGASINGAM, Y. Robust optic disc and cup segmentation with deep learning for glaucoma detection. **Computerized Medical Imaging and Graphics**, Elsevier, v. 74, p. 61–71, 2019.
- ZAHOOR, M. N.; FRAZ, M. M. A correction to the article “fast optic disc segmentation in retina using polar transform”. **IEEE Access**, IEEE, v. 6, p. 4845–4849, 2018.
- ZHEN, Y.; WANG, L.; LIU, H.; ZHANG, J.; PU, J. Performance assessment of the deep learning technologies in grading glaucoma severity. **CoRR**, abs/1810.13376, 2018. Disponível em: <<http://arxiv.org/abs/1810.13376>>. Acesso em: 19 de dezembro de 2019.

ANEXO A – ARQUITETURAS DOS MODELOS CNN PRÉ-TREINADOS

Tabela 12 – Arquitetura da rede completa MobileNet.

Passos	Tamanho do filtro	Tamanho de entrada
Conv/ s2	3 x 3 x 3 x32	224 x 224 x 3
Conv dw/ s1	3 x 3 x32 dw	112 x 112 x 32
Conv/ s1	1 x 1 x 32 x 64	112 x 112 x 32
Conv dw/ s2	3 x 3 x 64 dw	112 x 112 x 64
Conv/ s1	1 x 1 x 64 x 128	56 x 56 x 64
Conv dw/ s1	3 x 3 x 128 dw	56 x 56 x 128
Conv/ s1	1x 1 x 128 x128	56 x 56 x 128
Conv dw/ s2	3 x 3 x 128 dw	56 x 56 x 128
Conv/s1	1 x 1 x 128 x 256	28 x 28 x 128
Conv dw/ s1	3 x 3 x 256 dw	28 x 28 x 256
Conv/ s1	1 x 1 x 256 x 256	28 x 28 x 256
Conv dw/ s2	3 x 3 x 256 dw	28 x 28 x 256
Conv/s1	1 x 1 x 256 x 512	14 x 14 x 256
5 x Conv dw/ s1	3 x 3 x 512 dw	14 x 14 x 512
5 x Conv/ s1	1 x 1 x 512 x 512	14 x 14 x 512
Conv dw/ s2	3 x 3 x 512 dw	14 x 14 x 512
Conv/ s1	1 x 1 x 512 x 1024	7 x 7 x 1024
Conv dw/ s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv/s1	1 x 1 x 1024	7 x 7 x 1024
Avg Pool/ s1	Pool 7 x 7	7 x 7 x 1024
FC/ s1	1024 x 1000	1 x 1 x 1024
Softmax/ s1	Classificador	1 x 1 x 1000

Fonte: (HOWARD *et al.*, 2017).

Tabela 13 – Arquitetura da rede completa VGG16.

Passos	Tamanho do filtro	Tamanho de entrada
2 x Convolução	3 x 3 x 64	224 x 224 x 64
Max Pooling	3 x 3 x 64	224 x 224 x 64
2 x Convolução	3 x 3 x 128	112 x 112 x 64
Max Pooling	3 x 3 x 128	112 x 112 x 128
2 x Convolução	3 x 3 x 256	56 x 56 x 128
Max Pooling	3 x 3 x 256	56 x 56 x 256
3 x Convolução	3 x 3 x 512	28 x 28 x 256
Max Pooling	3 x 3 x 512	28 x 28 x 512
3 x Convolução	3 x 3 x 512	14 x 14 x 512
Max Pooling	3 x 3 x 512	7 x 7 x 512
Flatten		1 x 1 x 25088
Densa		1x 1 x 4096
Densa		1 x 1 x 4096
Softmax	Classificador	1 x 1 x 1000

Fonte: (RIZWAN, 2018).

Tabela 14 – Arquitetura da rede completa VGG19.

Passos	Tamanho do filtro	Tamanho de entrada
2 x Convolução	3 x 3 x 64	224 x 224 x 64
Max Pooling	3 x 3 x 64	224 x 224 x 64
2 x Convolução	3 x 3 x 128	112 x 112 x 64
Max Pooling	3 x 3 x 128	112 x 112 x 128
4 x Convolução	3 x 3 x 256	56 x 56 x 128
Max Pooling	3 x 3 x 256	56 x 56 x 256
4 x Convolução	3 x 3 x 512	28 x 28 x 256
Max Pooling	3 x 3 x 512	28 x 28 x 512
4 x Convolução	3 x 3 x 512	14 x 14 x 512
Max Pooling	3 x 3 x 512	7 x 7 x 512
Flatten		1 x 1 x 25088
Densa		1x 1 x 4096
Densa		1 x 1 x 4096
Softmax	Classificador	1 x 1 x 1000

Fonte: (RIZWAN, 2018).

Figura 39 – Tamanhos das saídas e núcleos convolucionais da arquitetura ResNet.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fonte: (WU *et al.*, 2019).