

Métodos de Classificação de Glaucoma em Imagens de Fundo de Olho

Débora Ferreira de Assis
Departamento de Engenharia
de Teleinformática
Universidade Federal do Ceará
Email: debora.ferreira@lesc.ufc.br

Pedro Motta
Departamento de Engenharia
de Teleinformática
Universidade Federal do Ceará
Email: pedro.motta@lesc.ufc.br

Paulo César Cortez
Departamento de Engenharia
de Teleinformática
Universidade Federal do Ceará
Email: cortez@lesc.ufc.br

Abstract—Glaucoma is an asymptomatic eye disease that, if not treated, can lead to blindness. The World Health Organization (WHO) estimated that by 2020 glaucoma should affect 80 million people and by 2040 will be 111.5 million. In this context, the present work aims to compare automatic classification methods to assist the specialist physician in the diagnosis of glaucoma. For this purpose, a model based on the extraction of non-geometric characteristics of optical disk images from the RIM-ONE r2 dataset had been developed. These characteristics were submitted to Principal Component Analysis (PCA) for dimensionality reduction, the resulting components served as input to the classifiers: Logistic Regression (RL), Decision Tree Gradient Boosting (DTGB), Support Vector Machine (SVM), k nearest neighbors (k-NN) and Multilayer Perceptron (MLP). To evaluate the results we used the accuracy, sensitivity, specificity, positive predictive value, negative predictive value and area under curve (AUC). The results demonstrate good performance with all classifiers, especially MLP, the test results reached an accuracy of 97.83 %, sensitivity 100 %, specificity 96.15 %, vpp 95.24 %, vpn 100 % and auc 97.62 %.

Keywords—Glaucoma; Classification; MLP; PCA.

1

I. INTRODUÇÃO

O glaucoma é uma doença ocular crônica que, se não for tratada, pode levar a perda de visão. Dentre as doenças oculares, o glaucoma é a segunda doença que mais causa cegueira no mundo (21), ficando atrás somente da catarata que é uma doença reversível. Dentre vários distúrbios oculares, o glaucoma é o que mais se associa ao valor de pressão intraocular alterada (22), a qual provoca lesões no nervo óptico, comprometendo assim o campo visual.

Essa doença tem tratamento, porém não tem cura. Nos estágios iniciais o glaucoma tende a ser assintomático e seus sintomas só aparecem apenas nos estágios mais avançados da doença. A descoberta precoce desta patologia é imprescindível, pois existem tratamentos que evitam a sua progressão, consequentemente, a perda da visão do paciente (21).

Segundo as projeções da Organização Mundial de Saúde (OMS), estima-se que o glaucoma deve afetar 80 milhões de pessoas em 2020, e esse número tende a crescer em 2040, alcançando 111,5 milhões de pessoas (14). Por consequência,

pesquisas que auxiliem no diagnóstico de glaucoma possuem grande importância para a saúde pública (15).

Um exame utilizado para obter informações acerca do fundo de olho e detectar glaucoma é a Fundoscopia. A partir desse exame, é possível visualizar as partes central e periférica da retina, o disco óptico e a mácula. A Fundoscopia é preferível aos demais exames visto que é mais econômica e mais fácil de ser realizada do que outros métodos de análise de retina disponíveis (23). No entanto, sua análise é limitada pela visão humana que incorpora fatores subjetivos, tornando-a dependente do especialista. Devido a isso é crescente os estudos para identificação automática de glaucoma, auxiliando assim em um diagnóstico médico mais preciso e eficiente.

Recentemente diversos trabalhos foram desenvolvidos para a detecção de glaucoma a partir de imagens de fundo do olho. Algumas abordagens utilizaram a segmentação do disco óptico para extrair as características geométricas, em outras como o trabalho de (11), além de usar características geométricas também utilizaram de características não geométrica, como informações de textura e intensidades, para descrever as imagens e identificar o glaucoma. Já em (13) foi utilizado somente as informações estatísticas das imagens para classificar o glaucoma.

Neste contexto, este trabalho apresenta uma metodologia de fácil implementação e de menor custo computacional para a avaliação de métodos de classificação na identificação do glaucoma, a partir de imagens do fundo de olho. Esta metodologia utiliza de aprendizagem de máquina supervisionada, na qual são comparados os desempenhos de diferentes classificadores com base nas características não geométricas selecionadas das imagens do fundo do olho.

II. LEVANTAMENTO BIBLIOGRÁFICO

A maioria dos trabalhos usam como principais atributos: características não geométricas das imagens e outras informações tais como texturas, bordas e entropias. Em (2) foram extraídos dos coeficientes da transformada de Gabor características como média, variância, assimetria, curtose, energia, entropias de Shannon e Rényi. Essas variáveis foram submetidas à análise de componentes principais (PCA), as 23 componentes resultantes foram aplicadas aos classificadores SVM e Naive Bayes (NB), sendo que o SVM obtém o melhor resultado

¹Agradecimentos à CAPES - Código de Financiamento 001. Trabalho suportado pelo projeto Proc. 311436/2015-3 CNPq.

com uma acurácia de 93,1%. Outro trabalho que também utilizou informações estatísticas foi (13), no qual extraiu das imagens de fundo de olho características estatísticas como: média, suavidade e terceiro momento. Estas características foram aplicadas nos classificadores NB, MLP, SVM e K-NN. Neste trabalho, o melhor resultado foi produzido pelo algoritmo K-NN com acurácia, sensibilidade e especificidade em torno de 95%.

Em (24) foram obtidas informações de correntropia a partir de componentes EWT decompostos. Essas características foram usadas para a classificação de imagens normais e de glaucoma usando o classificador SVM de mínimos quadrados (LS-SVM). Foram utilizadas uma base pública RIM-ONE e uma base privada, o melhor resultado se deu com a base pública alcançando uma acurácia de 98,33%.

Em (12) foram extraídas características de Haralick, que é uma matriz de co-ocorrência baseada em quatorze características estatísticas definidas sobre uma imagem, a classificação foi realizada com uma rede neural de propagação reversa (BPNN), cujos resultados alcançaram uma acurácia de 96,16%. Já no trabalho de (5) foram extraídas, das imagens pré-processadas, características dos índices de diversidade de Shannon e McIntosh como descritores de textura, a classificação foi realizada com o SVM que atingiu uma acurácia de 88,35% com alta taxa de especificidade e sensibilidade baixa.

No artigo (6) foi proposto um método de análise de textura através do LBP (*Local Binary Pattern*) para representar a região do disco óptico e funções geoestatísticas para descrever quantitativamente a textura da representação LBP. O SVM foi utilizado para a classificação final de imagens, com os melhores resultados atingem uma acurácia de 91,2%.

As principais informações para classificação de glaucoma são através das características geométricas (como NRR, BVR e ISNT), sendo que alguns trabalhos utilizam somente essas informações para classificação de glaucoma. Em outros trabalhos, os autores optaram por utilizar informações geométricas juntamente com as características não geométricas.

Em (8) foi proposto uma abordagem 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 foram realizados através do SVM com diferenças nos kernel, neste caso, o classificador SVM Twin obtém o melhor resultado com uma acurácia de 94,4%. Em (11) também foi proposto uma combinação das propriedades geométricas (CDR, NRR e ISNT) e não geométricas (textura e intensidade) para melhorar a precisão do diagnóstico automatizado de glaucoma. As imagens utilizadas foram selecionadas das bases MESSIDOR, HEI-MED, DRIVE, STARE e DiaRetDB0. As características extraídas foram submetidas ao PCA, e a classificação foi feita com o SVM, obtendo-se uma taxa de acerto foi de 91%.

O modelo de redes neurais convolucionais (CNN) extraem características das imagens e classificam em um único processo. No entanto, alguns autores optam por utilizar a CNN somente para extração das características e classificam com

métodos de aprendizado de máquinas.

Em (4), um sistema automatizado baseado em uma CNN foi desenvolvido para extrair características que distinguem padrões normais e glaucomatosos que foram inseridas no SVM para classificar as imagens em saudáveis ou glaucomatosas. Assim, o modelo proposto atingiu 88,2% de taxa de acerto. Em (10) foi utilizado uma CNN com 18 camadas efetivamente treinadas para extrair características robustas das imagens do fundo de olho. Tais características foram classificadas com o LDA, cujo resultado atinge 98% de taxa de acerto.

As técnicas mais comuns para classificação de imagens utilizando CNN são através de um único processo de extração de características e classificação, os trabalhos descritos a seguir fazem uso dessas técnicas. Além disso, alguns dos autores usam métodos para selecionar as informações mais importantes e otimizar as camadas do modelo proposto.

No trabalho de (1), a arquitetura não supervisionada de uma CNN foi empregada para extrair características da imagem do fundo do olho das bases DRIONS-DB, HRF-dataset, sjchoi86-HRF e PVR. Posteriormente, o modelo de *deep-belief network* (DBN) foi usado para selecionar as características mais discriminativas. Por fim, a decisão final foi realizada pelo classificador linear *softmax* para diferenciar entre glaucoma e não glaucoma, cuja acurácia foi de 84,5%.

O artigo de (9) fez uso de um Algoritmo Genético (GA) para otimizar arquiteturas de CNN. Inicialmente as imagens foram redimensionadas para que todas possuíssem as mesmas dimensões, logo após foi executado o GA para encontrar melhor arquitetura da CNN. Por fim, a CNN classificou as imagens em saudáveis ou não, com uma acurácia de 91%.

Um outro método utilizando CNN é através de aprendizado por transferência (*learning transfer*), no qual permite que uma rede neural CNN, que é usada em determinada tarefa, seja aplicada em outra tarefa. Este é o caso de (?) que comparou três modelos: classificação logística simples, uma CNN usando o Tensorflow, e o modelo GoogleNet Inception v3 pré-treinado. O modelo de classificação logística simples mostrou uma acurácia de teste de 77,2%. Já a CNN e GoogleNet Inception v3 alcançaram uma precisão de 87,9% e 84,5% em dados de teste, respectivamente.

Em (7) foram exploradas 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 é uma junção da ESPERANZA, RIM-ONE (v1, r2 e v3) e DRISHTI-GS. O melhor desempenho alcançado foi utilizando aprendizagem por transferência com o modelo VGG19, o qual produziu um acurácia de 88,05%.

III. MATERIAIS E MÉTODOS

As imagens utilizadas são da base pública An Open Retinal Image Database for Optic Nerve Evaluation (RIM-ONE), que é um grupo de bancos de dados oftalmológicos projetados para servir de referência para o desenvolvimento de algoritmos de segmentação do nervo óptico e de diagnóstico do glaucoma

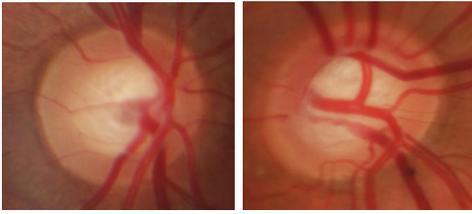


Fig. 1. Fundo do olho de paciente saudável e paciente com glaucoma, respectivamente.

auxiliado por computador (7). A base utilizada foi a versão 2, um exemplo está mostrado na Figura 1, que é composta por 200 imagens referente a 'glaucoma/suspeita de glaucoma' e 255 normal, totalizando 455 imagens.

As imagens do dataset RIM-ONE-r2 foram adquiridas com diferentes dimensões em *pixels*, portanto, todas as imagens foram redimensionadas para 256×256 . As imagens estão no formato JPEG, no sistema RGB (red, green, blue), todas foram convertidas em escala de cinza com 8 bits.

Algumas dificuldades foram encontradas nesse conjunto de dados como a qualidade de algumas imagens, no entanto, sua escolha foi por ser uma base pública e pela quantidade de imagens relativamente grande, em comparação a outras bases públicas. Além disso, o conjunto de dados possui distinção entre saudáveis e com 'glaucoma/suspeita de glaucoma'. Na aplicação da metodologia, a classe normal foi rotulada por 0 (zero) e a classe 'glaucoma/suspeito de glaucoma' por 1 (um).

No método proposto, inicialmente todas imagens foram pré-processadas, transformadas para a escala de cinzas e com dimensões iguais (256×256). Em seguida, foram extraídas as características selecionadas das imagens que passaram pela análise de componentes principais para redução de dimensionalidade para a entrada dos classificadores. As componentes foram classificadas por métodos de aprendizagem de máquinas, já com os hiper parâmetros ajustados, com as duas saídas possíveis da classificação foram: saudáveis ou glaucoma/ suspeita de glaucoma. A Figura 2 mostra visualmente os passos do método proposto.

O computador utilizado nas análises possui um processador Intel(R) Core(TM) i3-4330 CPU 3.50GHz, com memória RAM de 12,0GB e o sistema operacional de 64 Bits.

O tempo para o computador extrair as características das imagens foi de 2 min 16 s 61 ms, e o PCA para redução de dimensionalidade dos dados decorreu em 30 ms.

TABLE I
CUSTO COMPUTACIONAL REFERENTE A CADA CLASSIFICADOR.

Classificadores	Hiper parâmetros	Classificação
RL:	290 ms	4,29 ms
k-NN:	370 ms	4,71 ms
GBDT:	4 min 24s	330 ms
SVM:	15,2 s	251 ms
MLP:	1 min 3 s	358 ms

O custo computacional descrito na primeira coluna de valores da tabela III, refere-se a otimização dos hiper parâmetros

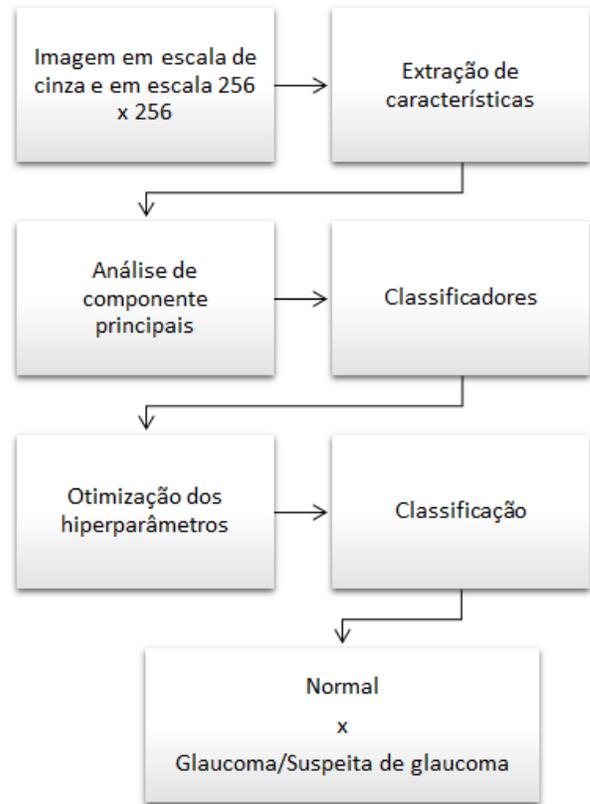


Fig. 2. Diagrama de blocos do método proposto.

implementados uma única vez para cada classificador. Após os hiper parâmetros ajustados, foram executados 100 vezes a classificação e extraída a média para mensurar o tempo de processamento de cada classificador. Na segunda coluna de valores da tabela III é mostrado o tempo médio de execução de cada classificador.

A regressão logística foi o classificador mais rápido, e o mais demorado foi o MLP. Apesar disso, observa-se que o tempo de execução de todos os classificadores é bastante rápido, menos que 1 segundo.

A. Características não geométricas

Com base nos trabalhos relacionados, testaram-se várias configurações de descritores de características. Cada descritor foi avaliado individualmente e combinados para classificar as imagens.

Inicialmente foram testados separadamente cada descritor, o melhor desempenho foi obtido com o HOG (*Histogram of Oriented Gradients*). Os resultados obtidos foram acima da média, a acurácia nos classificadores permaneceu entre 82% e 87%, apesar disso a sensibilidade ficou abaixo da média, ou seja, o algoritmo não classificou adequadamente os indivíduos com glaucoma. Já os indivíduos saudáveis foram classificados corretamente, o que é mostrado na especificidade do teste em todos os classificadores permaneceu acima de 90%.

Segundo o trabalho de Wang et al. (20), quando o LBP é combinado com o HOG, melhora consideravelmente o desempenho de detecção em alguns conjuntos de dados.

Com base nessa informação, adicionando-se o LBP, os resultados melhoram para alguns classificadores como a regressão logística e MLP que atingiram 89% de taxa de acerto. O restante permaneceu com acurácia próximas ou iguais, no entanto, foi visto que em todos os casos a sensibilidade aumenta levemente. Desse modo, como aumentou a sensibilidade, foi decidido permanecer com o LBP no modelo.

Outro descritor utilizado foi os momentos de Zernike, por ser mais adequado ao método. Com a adição desse descritor, os resultados melhoraram, exceto com o classificador k-NN que reduziu em 2% a taxa de acerto. Prosseguindo os experimentos, permutam-se os momentos de Zernike pela entropia de Shannon das imagens para verificar se os resultados apresentaram melhoras. Comparando com os resultados usando somente os descritores HOG e LBP, houve melhora apenas com os classificadores k-NN e MLP.

Quando adicionado os momentos de Zernike e a entropia de Shannon, os resultados apresentam levemente uma melhora, o MLP chega a atingir 92% de taxa de acerto. Como a sensibilidade aumentou, ou seja, passou a classificar corretamente mais indivíduos com glaucoma, então esses descritores permanecem no modelo.

Adicionando-se as informações do filtro de Gabor para descrever a textura das imagens, a acurácia de alguns testes melhoraram, a taxa de especificidade aumentou, indicando que o modelo classifica corretamente as pessoas doentes. No entanto, produzem-se muitos falsos negativos, ou seja, o teste indica que a pessoa é saudável quando na realidade é doente. Como os resultados da sensibilidade não foram adequados, os filtros de Gabor foram excluídos do método e as informações de dissimilaridade do GLCM foram adicionadas. Observam-se que, para todos os classificadores, apenas a taxa de especificidade permanece acima da média, produzindo também melhora na acurácia. Porém, a sensibilidade continua abaixo da média, provando que esse descritor não contribui para uma classificação adequada para as imagens da base Rim-ONE v2.

Em alguns trabalhos como (19) foi mostrado que, separadamente as informações dos filtros de Gabor e GLCM podem não ser bons descritores, mas quando avaliados em conjunto tendem a apresentar resultados melhores. Com base nisso, foram adicionados ambos os descritores de textura e os resultados mostraram que, em alguns classificadores a acurácia aumentou, assim como a sensibilidade também. Já para outros classificadores não houve uma melhora significativa mas, ao contrário disso, houve uma redução como é o caso da regressão logística. Entretanto, o maior valor alcançado dentre todos os outros foi de 94,51% para o MLP, atingindo um valor para a AUC de 95%. Por isso foram escolhidos esses descritores para compor o método proposto.

Os descritores que apresentaram melhores desempenho para o conjunto de imagens foram: entropia de Shannon da imagem, LBP (*Local Binary Pattern*), HOG (*Histogram of Oriented Gradients*), medida de dissimilaridade do GLCM (*gray-level*

co-occurrence matrix), momentos de Zernike e informações estatísticas do filtro de Gabor.

O LBP é utilizado para descrição local de textura que é baseada na suposição de que a informação de uma textura pode ser decomposta em conjuntos de pequenas unidades, chamadas de unidades texturais (16). Para o conjunto de imagens utilizadas nesse trabalho, o cálculo do LBP é tratado usando o raio igual a 8 e o número de pontos igual a 24. O método utilizado foi o uniforme, indicando que cálculo computado é uma forma invariante de rotação e em tons de cinza de LBPs.

O HOG é um descritor de características globais que descreve informações da aparência e forma do objeto em uma imagem através da distribuição de gradientes de intensidade ou direções de borda. Os parâmetros utilizados foram: 4 posições na orientação, 128×128 pixels de cada célula, e 2 células em cada bloco.

A entropia da informação é uma forma de medir a taxa média de incerteza acerca das informações produzidas por uma fonte estocástica. Quanto maior o valor da entropia mais variações de tonalidades terá a imagem.

Os momentos de Zernike são uma medida global de como a massa está distribuída.

Os filtros de Gabor são usados em análise de textura para detecção de bordas. Quando um filtro Gabor é aplicado a uma imagem, fornece a maior resposta nas bordas e nos pontos onde a textura é alterada. Foram utilizadas as informações estatísticas do filtro: média, variância, assimetria, curtose, energia e entropia.

A matriz de co-ocorrência de nível de cinza, do inglês *gray-level co-occurrence matrix* (GLCM), também é um método de extração de características baseado em textura. A GLCM armazena a probabilidade de que dois valores de intensidade de cinza estejam envolvidos por uma determinada relação espacial. A partir desta matriz de probabilidades, diferentes medidas estatísticas são extraídas para caracterizar a textura presente na imagem (17). Nesse trabalho, foi usado apenas a informação de dissimilaridade do GLCM dos ângulos θ e $\pi/2$, com dois deslocamentos da distância do par de pixels.

B. Classificação

As características obtidas foram processadas através da análise de componentes principais PCA (*Principal Component Analysis*) para redução de dimensionalidade. É um método estatístico que converte o conjunto de dados em combinações lineares não correlacionados, chamadas de componentes principais, a fim de reduzir a dimensionalidade das variáveis, as componentes resultantes sempre são menores que a quantidade de variáveis. No total, foram 96 informações extraídas (variáveis), com a análise de componentes principais essa quantidade reduziu-se para 41 componentes.

As componentes resultantes são usadas como entrada para os classificadores: Regressão Logística (RL), k-Vizinhos Mais Próximos (k-NN), Árvore de Decisão com Aumento de Gradiente (GBDT), Máquinas de Vetores de Suporte (SVM) e Perceptron Multicamadas (MLP).

Para otimizar os hiperparâmetros dos classificadores foi utilizado o *Grid Search*, ou pesquisa de grade, com validação cruzada *5-fold*. O método *textitGrid Search* faz uma busca exaustiva no intervalo de hiperparâmetros, ou seja, para cada hiperparâmetro é especificado um conjunto de valores candidatos que serão postos no modelo e avaliados para obter qual a melhor configuração de hiperparâmetros para os dados em questão. Além de usar o *Grid Search*, também foram ajustados alguns hiperparâmetros manualmente, os quais estão descritos a seguir.

- RL: a inversão da força de regularização foi o valor padrão $C=1$, valores menores especificam uma regularização mais forte. E para a regularização foi usada a penalidade L1.
- k-NN: o número de vizinhos (k) foi igual a 3, a função peso usada para previsão foi a medida de distância euclidiana, ou seja, os vizinhos mais próximos de um ponto de consulta terão uma influência maior do que os vizinhos mais distantes.
- XGB: a proporção de subamostras de colunas ao construir cada árvore é de 0,5; a profundidade máxima da árvore é de 2; o número de árvores é 1000, o peso L1 da regularização é igual a 0,01 e a proporção da subamostra da instância de treinamento é igual a 0,5.
- SVM: o parâmetro de penalidade C do termo de erro é igual a 10, com kernel RBF e coeficiente gama igual 0,001.
- MLP: a função de ativação utilizada foi a RELU com valor de alfa igual a 0,0001 (parâmetro de regularização L2), a taxa de aprendizagem constante para atualização dos pesos, o solucionador para otimização de peso é um otimizador baseado em gradiente estocástico: ADAM, e `hidden_layer_sizes = (100,)`, isto é, uma camada de entrada, uma de saída, e uma camada oculta com 100 unidades ocultas (neurônios).

C. Métricas de Avaliação

As métricas para comparação do desempenho dos classificadores são: acurácia, sensibilidade, especificidade, valor preditivo positivo, valor preditivo negativo e AUC. Antes da definição é importante entender alguns conceitos básicos.

TABLE II
MATRIZ DE CONFUSÃO - VALIDADE DE UM TESTE DE DIAGNÓSTICO

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

A acurácia (A) indica a proporção de acerto, ou seja, se o teste classificou corretamente os indivíduos doentes e saudáveis. A sensibilidade (S) é a capacidade do teste detectar a doença quando o indivíduo realmente está doente. A especificidade (E) é a capacidade do teste diagnosticar ausência da doença, quando os indivíduos são saudáveis.

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

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

$$VPP = \frac{VP}{VP + FP}, \quad VPN = \frac{VN}{VN + FN}.$$

O valor preditivo positivo (VPP) é a proporção de verdadeiros positivos entre todos os indivíduos com teste positivo, este valor indica a probabilidade de uma pessoa com determinadas características que indicam glaucoma ter realmente a doença. O Valor preditivo negativo (VPN) é a proporção de verdadeiros negativos entre todos os indivíduos com teste negativo, este valor indica a probabilidade de uma pessoa com determinadas características que indicam olho saudável realmente não ter a doença.

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. A Curva ROC é uma métrica de verificação de qualidade do modelo, que determina a chance de classificar corretamente um indivíduo com glaucoma (sensibilidade), contra a chance do teste ser positivo quando na realidade o indivíduo é saudável (1-especificidade). A qualidade do modelo é obtida através da área sob a curva (area under curve - AUC), que fornece uma medida de taxa de acerto. Portanto, quanto mais distante da linha diagonal melhor será o modelo (18).

IV. RESULTADOS E DISCUSSÕES

Para verificar qual configuração de separação treino e teste obteve melhor desempenho, segundo os classificadores, foram testadas diferentes porcentagens de separações para treinamento e teste: 70/30, 80/20 e 90/10.

Em relação a separação com proporção 70/30, a quantidade de imagens para treino foram 318, e de teste foram 137. A separação foi aleatória, no teste a quantidade de imagens com glaucoma foram 57 e de indivíduos saudáveis foram 80. O melhor resultado foi para o MLP, exceto para o valor de especificidade e valor preditivo positivo, que foi menor dentre todos. Apesar disso, como o intuito do estudo é detectar indivíduos com glaucoma, deseja que os valores de sensibilidade sejam altos, por esse motivo o MLP destacou-se dentre os outros. Os demais classificadores apresentaram acurácia maior que 85%, o que já indica que o modelo proposto é adequado.

TABLE III
RESULTADOS COM A SEPARAÇÃO 70/30

Cla	A (%)	S(%)	E (%)	VPP (%)	VPN (%)	AUC (%)
RL:	85,40	75,44	92,50	87,76	84,09	85,92
KNN:	86,13	75,44	93,75	89,58	84,27	86,93
GBDT:	89,78	80,70	96,25	93,88	87,50	90,69
SVM:	88,32	82,46	92,50	88,68	88,10	88,39
MLP:	89,05	85,96	91,25	87,50	90,12	88,81

Em relação a separação com proporção 80/20, a quantidade de imagens para treino foram 364 e de teste foram 91. A separação foi aleatória, e no teste a quantidade de imagens com glaucoma foram 38 e de indivíduos saudáveis foram 53. O melhor resultado continua sendo com a rede MLP, alcançando maiores valores de acurácia AUC, especificidade e, conseqüentemente, valor preditivo positivo. O classificador SVM apresentou maiores valores de sensibilidade e valor preditivo negativo.

TABLE IV
RESULTADOS COM A SEPARAÇÃO 80/20

Cla	A (%)	S(%)	E (%)	VPP (%)	VPN (%)	AUC (%)
RL:	89,01	81,58	94,34	91,18	87,72	89,45
KNN:	87,91	78,95	94,34	90,91	86,21	88,56
GBDT:	91,21	81,58	98,11	96,88	88,14	92,51
SVM:	93,41	92,11	94,34	92,11	94,34	93,22
MLP:	94,51	89,47	98,11	97,14	92,86	95,00

Em relação a separação com proporção 90/10, a quantidade de imagens para treino foram 409 e de teste foram 46. A separação foi aleatória, e no teste a quantidade de imagens com glaucoma foram 20 e a quantidade de imagens sem glaucoma (indivíduos saudáveis) foram 26. O melhor resultado continua sendo para a rede MLP, alcançando maiores valores em todas as métricas avaliadas. Já o k-NN e o GBDT apresentaram desempenho semelhante. Vale ressaltar que os valores de sensibilidade e especificidade foram acima da média, indicando que o modelo está classificando corretamente indivíduos doentes e saudáveis.

Nesse caso específico, a rede MLP apresentou o valor preditivo positivo de 100%, ou seja, em todos os casos que o teste deu positivo, o paciente realmente tinha a doença; e para o valor preditivo negativo, o classificador errou apenas em uma imagem, classificando um indivíduo como saudável quando na realidade ele possui glaucoma. O valor preditivo positivo indicou que, nos 21 casos que o teste foi positivo para o glaucoma, em 20 o paciente estava realmente doente; e o valor preditivo negativo indicou que nos 25 casos que o teste foi negativo para a doença, em todos os indivíduos eram realmente saudáveis. A AUC resultou em 97,62%, ou seja, em 97,62% das vezes em que for utilizado, o teste discriminará os verdadeiro-positivos e os verdadeiro-negativo; isto também significa que em 2,38% das vezes fornecerá um falso resultado.

TABLE V
RESULTADOS COM A SEPARAÇÃO 90/10.

Cla	A (%)	S(%)	E (%)	VPP (%)	VPN (%)	AUC (%)
RL:	89,13	95,00	84,62	82,61	95,65	89,13
KNN:	91,30	85,00	96,15	94,44	89,29	91,87
GBDT:	91,30	85,00	96,15	94,44	89,29	91,87
SVM:	95,65	100,00	92,31	90,91	100,00	95,45
MLP:	97,83	100,00	96,15	95,24	100,00	97,62

A curva ROC 3 para os classificadores com a separação 90/10 mostra que todos os classificadores tiveram desempenho próximo do ideal, isto porque a curva se encontra distante da diagonal, aproximando-se do valor 1.

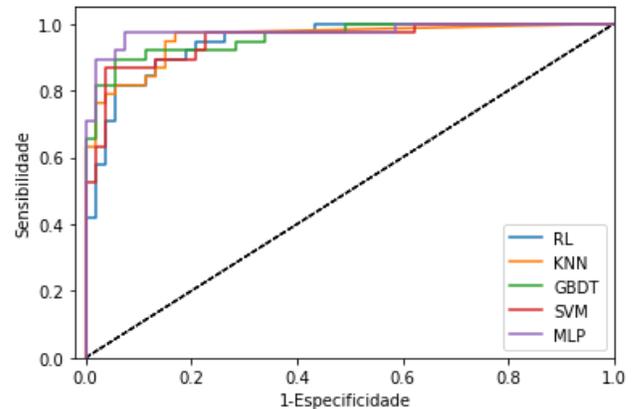


Fig. 3. Curva ROC para separação 90/10.

A qualidade das imagens prejudica o classificador na separação de classes, ocorrendo assim o erro no momento de identificar a qual classe pertence. Algumas características são responsáveis pela classificação errada das classes, 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 da imagem mais eficiente será a extração de características relevantes, como bordas e contornos, para caracterizar as diferentes imagens. A Figura 4 (a) mostra um disco óptico com glaucoma classificado erroneamente - a imagem está bastante borrada e os contornos e bordas de difícil distinção - um falso negativo, o teste indica que o paciente está saudável, quando na verdade está com glaucoma. E a Figura 4(b) um falso positivo, quando o teste indica que o indivíduo está doente quando na realidade ele é saudável.

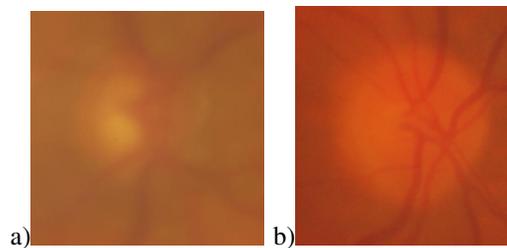


Fig. 4. Classificação errônea típica de uma imagem, falso negativo (a) e falso positivo (b).

TABLE VI
COMPARAÇÃO DE DESEMPENHO DO MÉTODO PROPOSTO COM OUTROS MÉTODOS.

Autores	Dados	Qtd	Método	A (%)	S (%)	E (%)
Acharya et al. (2)	Kasturba Medical College	510	SVM	93,10	89,75	96,20
Septiarini et al. (13)	Dr. YAP Eye and Dr. Sardjito Hospital	84	K-NN	95,24	95,12	95,35
Maheshwari et al (24)	Kasturba Medical College	60	LS-SVM	98,33	100,00	96,67
Samanta et al. (12)	RIM-ONE r2	455	BPNN	96,16	99,51	90,44
Araújo et al. (5)	RIM-ONE r2	455	SVM	88,35	84,50	91,37
De Sousa et al. (6)	RIM-ONE r2	455	SVM	91,20	95,00	88,20
Haleem et al. (8)	RIM-ONE	124	SVM Twin	94,40	92,20	95,30
Salam et al. (11)	-	100	SVM	91,00	100,00	87,00
Al-Bander et al. (4)	RIM-ONE r2	455	SVM	88,20	85,00	90,80
Raghavendra et al. (10)	Kasturba Medical College	1426	CNN	98,13	98,00	98,30
Abbas (1)	DRIONS-DB, HRF e PVR	1200	CNN	84,50	98,01	99,00
Lima et al. (9)	RIM-ONE r2	455	CNN	91,00	-	-
Ahn et al. (3)	Kim's Eye Hospital	1542	CNN	87,90	-	-
Gómez-Valverde et al. (7)		2313	VGG19	88,05	87,01	89,01
Método proposto	RIM-ONE r2	455	MLP	97,62	100,00	96,15

Comparando-se os resultados obtidos com essa metodologia proposta com os trabalhos relacionados, observam-se que alguns resultados mostraram uma acurácia mais alta como é o caso do trabalho de (24) em que utilizou informações de correntropia e classificou com SVM de mínimos quadrados; e o de (10) que usou CNN com 18 camadas para extrair características e classificá-las. Apesar desses resultados serem melhores, a metodologia proposta alcança resultados promissores para o conjunto RIM-ONE r2, e com taxas de acertos maiores do que em outros trabalhos que utilizaram a mesma base de dados. Além disso, o custo computacional da metodologia proposta é menor. No entanto, ainda deve ser estudado um aprimoramento nos descritores de características e nos classificadores para reduzir os falsos negativos.

V. CONCLUSÕES

Esse trabalho propõe um método usando uma rede MLP simples com três camadas para a detecção de glaucoma em imagens de fundo de olho, empregando para tanto características não geométricas. Usa a base de dados RIM-ONE r2 que é empregada em vários trabalhos relacionados. O método proposto no processo de classificação é comparado com outros cinco classificadores, obtendo resultados promissores. Dentre as separações estudadas, para treinamento e teste, todas apresentaram resultados animadores, principalmente para a separação 90/10, que indica que as técnicas de extração de características juntamente com o PCA são recursos fundamentais para a classificação de pacientes com 'glaucoma/suspeita de glaucoma' ou saudáveis. Além disso, o classificador MLP mostrou-se ser com um desempenho comparável aos melhores classificadores, alcançando uma precisão de 97,62%. Por fim, por ser de simples possui baixo custo computacional para a aplicação descrita neste trabalho. Destaca-se seu desempenho superior a todos os trabalhos avaliados para aqueles que usam a mesma base de dados. Dessa forma, o método proposto

pode ser usado para o auxilia ao diagnóstico de glaucoma, oferecendo recursos automáticos e precisos para esta doença.

VI. TRABALHOS FUTUROS

Atualmente muitos trabalhos estão envolvendo redes neurais convolucionais, principalmente pela facilidade e praticidade devido este tipo de mecanismo extrair as características e realizar a classificação em um único procedimento. Com base nesses estudos, para trabalhos futuros é considerado utilizar alguma configuração de CNN para classificar de forma automática o glaucoma. Ademais, empregar aprendizagem por transferência para extração de características das imagens e aplicar nos classificadores estudados anteriormente. Além disso, segmentar o disco óptico e escavação para obter informações geométricas do disco óptico, para que assim seja possível combinar características geométricas e não geométricas para descrever e classificar as imagens do disco óptico em glaucoma ou saudáveis. Desse modo, busca uma metodologia que reduza os falsos negativos e possa distinguir melhor a classe de pessoas com glaucoma.

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.

REFERENCES

- [1] Q. Abbas, "Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning". Int J Adv Comput Sci Appl, v. 8, n. 6, p. 41-5, 2017.
- [2] U. R. Acharya, E. L. Eugene, K. P. Noronha, L. C. Min, K. P. Nayak, and S. V. Bhandary, "Decision support system for the glaucoma using Gabor transformation", Biomedical

- Signal Processing and Control, Elsevier, v. 15, p.18-26, 2015.
- [3] J. M. Ahn, S. Sangsoo, k. S. Ahn, S. H. Cho, K. B. Lee and U. S. Kim, "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.
- [4] B. Al-Bander, W. Al-Nuaimy, M. A. Al-Tae and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach," 2017 14th International Multi-Conference on Systems, Signals & Devices (SSD), Marrakech, p. 207-210, 2017.
- [5] J. D. L. Araújo, A. C. Paiva, J. D. S. Almeida, O. P. S. Neto, J. A. de Sousa, A. C. Silva and G. B. Júnior, "Diagnóstico de glaucoma em imagens de fundo de olho utilizando Índices de diversidade de Shannon e McIntosh". Anais do XVII Workshop de Informática Médica (WIM 2017). SBC, 2017.
- [6] J. A. de Sousa, A. C. Paiva, J. D. S. de Almeida, A. C. Aristofanes, G. B. Junior and M. Gattass, "Texture based on geostatistic for glaucoma diagnosis from fundus eye image". Multimedia Tools and Applications, Springer, v. 76, n. 18, p. 19173-19190, 2017.
- [7] J. J. Gómez-Valverde, A. Antón, G. Fatti, B. Liefers, A. Herranz, A. Santos, C. Sánchez and M. J. Ledesma-Carbayo. "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.
- [8] M. S. Haleem, L. Han, J. Van Hemert, A. Fleming, L. R. Pasquale, P. S. Silva, B. J. Song and L. P. Aiello, "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.
- [9] A. Lima, L. B. Maia, P. T. C. Santos, G. B. Junior, D. D. S. Almeida and A. C. Paiva, "Evolving Convolutional Neural Networks for Glaucoma Diagnosis". In: 18th Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2018). SBC, 2018.
- [10] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H. Tan, U. R. Acharya, "Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images". Information Sciences, Elsevier, v. 441, p. 41-49, 2018.
- [11] A. Salam, T. Khalil, M. U. Akram, A. Jameel and I. Basit, "Automated detection of glaucoma using structural and non structural features". Springerplus, Springer, v. 5, n. 1, p. 1519, 2016.
- [12] S. Samanta, S. K. S. Ahmed, M. A. M. Mohammed, S. S. Nath, N. Dey, S. S. Chowdhury, "Haralick features based automated glaucoma classification using back propagation neural network", Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, Springer, p. 351-358 2015.
- [13] A. Septiarini, D. M. Khairina, A. H. Kridalaksana and H. Hamdani, "Automatic glaucoma detection method applying a statistical approach to fundus images", Healthcare informatics research, v. 24, n. 1, p. 53-60, 2018.
- [14] R. Turbiani, "Principal causa de cegueira irreversível, glaucoma avança com envelhecimento da população", BBC, 2019. Disponível em: <https://www.bbc.com/portuguese/geral-47979880>.
- [15] Jonas, J. B., T. Aung, R. R. Bourne, A. M. Bron, R. Ritch and S. Panda-Jonas, "Glaucoma," The Lancet. Volume 390, Issue 10108, Pages 2183-2193, ISSN 0140-6736, 2017.
- [16] M. Pietikäinen, "Local Binary Patterns", Scholarpedia, v. 5, n. 3, p.9775, 2010.
- [17] J. G. Martins, Y. M. G. Costa, D. Bertolini, L. S. Oliveira, "Uso de descritores de textura extraídos de glcm para o reconhecimento de padrões em diferentes domínios de aplicação", XXXVII Conferencia Latinoamericana de Informática, p. 637-652, 2011
- [18] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez and M. Müller, "pROC: display and analyze ROC curves," R Package Version, v. 1, 2015.
- [19] F. Mirzapour and H. Ghassemian, Hassan, "Using GLCM and Gabor filters for classification of PAN images" . 21st Iranian Conference on Electrical Engineering (ICEE), IEEE, p. 1-6, 2013.
- [20] X. Wang, T. Han and S. Yan, "An HOG-LBP human detector with partial occlusion handling". In: 2009 IEEE 12th international conference on computer vision. IEEE, p. 32-39, 2009.
- [21] M. Claro, R. M. S. Veras and A. M. Santana, "Metodologia para Identificação de Glaucoma em Imagens de Retina". In: Anais Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde. SBC, p. 103-108, 2019.
- [22] H. J. C. Barbosa, D. M. Ferreira, R. R. Azavedo Júnior and A. R. M. Szpilman, "Perfil clínico e epidemiológico dos pacientes atendidos em um ambulatório de oftalmologia em Vila Velha/ES". Revista Brasileira de Pesquisa em Saúde/Brazilian Journal of Health Research, v. 20, n. 3, p. 77-84, 2019.
- [23] M. D. Abràmoff, M. K. Garvin and M. Sonka, "Retinal imaging and image analysis". IEEE reviews in biomedical engineering, v. 3, p. 169-208, 2010.
- [24] S. Maheshwari, R. B. Pachori and U. R. Acharya, "Automated diagnosis of glaucoma using empirical wavelet transform and correntropyfeatures extracted from fundus images". IEEE journal of biomedical and healthinformatics, v. 21, n. 3, p. 803-813, 2017