

# Metodologia para Identificação de Glaucoma em Imagens de Retina

Maíla L. Claro<sup>1</sup>, Rodrigo M. S. Veras<sup>1</sup>, André M. Santana<sup>1</sup>

<sup>1</sup>Programa de Pós-graduação em Ciência da Computação – Universidade Federal do Piauí  
Teresina – PI – Brasil

claromaila@gmail.com, {rveras, andremacedo}@ufpi.edu.br

**Abstract.** *Glaucoma é uma doença que danifica o nervo óptico. Ela é considerada a segunda principal causa de cegueira no mundo. Vários sistemas de diagnóstico automático têm sido propostos. No entanto, esses sistemas não foram capazes de lidar com uma grande diversidade de imagens. Portanto, tais métodos não são viáveis para uso em programas de triagem. Realizamos um extenso estudo para definir o melhor conjunto de atributos para a representação da imagem. No total, avaliamos 16.469 características. Nossa abordagem de detecção de glaucoma usa descritores de textura e Redes Neurais Convolucionais (CNNs). Avaliamos nossa proposta em um total de 873 imagens de quatro bancos de dados públicos e concluímos que a junção da GLCM e CNNs pré-treinadas juntamente com o uso do classificador Random Forest são promissoras na detecção desta patologia, obtendo uma acurácia de 93,35% e um índice Kappa considerado Excelente.*

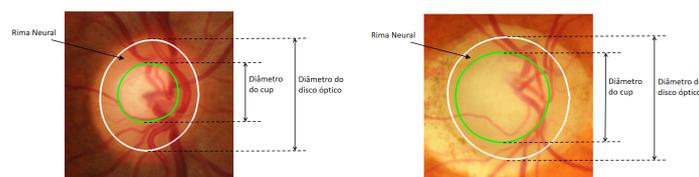
## 1. Introdução

No campo oftalmológico, foram desenvolvidos um grande número de sistemas de Diagnóstico Auxiliado por Computador (CAD - *Computer Aided Diagnosis*) para o auxílio na detecção de vários tipos de doenças oculares, entre elas o glaucoma [Kotyk et al. 2016, Orlando et al. 2017]. Estes sistemas têm potencial para fornecer uma solução alternativa aos programas de triagem em massa, que precisam examinar um grande número de imagens de fundo de olho de forma eficiente e robusta.

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 e, consequentemente, a perda da visão do paciente. Uma estimativa sugere que, em 2020, cerca de 80 milhões de pessoas terão essa doença [Quigley and Broman 2006].

O diagnóstico de glaucoma pode ser gerado a partir da análise de imagens digitais de retina. O mesmo só é possível devido a quantidade de perda de fibras do nervo óptico ter um efeito direto na configuração da rima neural. À medida que as fibras ópticas nervosas vão morrendo, a escavação (*cup*) torna-se mais larga em relação ao disco óptico (DO), o que acarreta em um valor da relação Escavação/Disco (*Cup to Disc Ratio* - CDR) aumentado [Mittapalli and Kande 2016], como pode ser visto na Figura 1.

A literatura apresenta vários estudos para processamento e análise de imagens da retina para o diagnóstico automático do glaucoma. No entanto, é comum que os pesquisadores avaliem suas propostas em um conjunto restrito de imagens (geralmente privadas).



**Figura 1. (a) Imagem de uma retina saudável e (b) Imagem de uma retina glaucomatosa. O DO está marcado em branco e a escavação (cup) em verde.**

Eles usam estratégias dependentes das características dos bancos de dados nos quais foram testados.

O objetivo deste trabalho é realizar a detecção automática do glaucoma em uma maior diversidade de imagens da retina, constituída pela união de diferentes bancos de dados. Para desenvolver o método proposto, realizamos vários experimentos utilizando descritores de textura, forma e Redes Neurais Convolucionais (CNNs) pré-treinadas como entrada para os classificadores *Random Forest* (RF), *Support Vector Machine* (SVM) e *MultiPayer Perceptron* (MLP). Com base nestes experimentos, realizamos uma análise comparativa para obtermos os descritores e classificadores que alcançaram métricas de desempenho mais altas, para em sequência realizar uma combinação entre eles. Assim, este trabalho apresenta um conjunto de recursos para a detecção automática do glaucoma.

Ao longo do manuscrito, por razões de espaço, omitimos algumas referências. No entanto, detalhes podem ser encontrados no texto da dissertação.

## 2. Trabalhos Relacionados

Realizou-se uma revisão da literatura para identificar as principais técnicas utilizadas no diagnóstico do glaucoma em imagens da retina.

Kotyk et al. 2016, os autores propuseram um sistema semi-automatizado para detectar tanto o DO quanto a escavação. Essas duas regiões são utilizadas no cálculo do CDR. O canal vermelho do modelo de cor RGB é utilizado nas imagens da retina da base RIM-ONE-1. As imagens são posteriormente processadas usando o descritor de textura de Haralick, e os valores obtidos são utilizados no classificador MLP para determinar a exatidão do sistema. A acurácia do sistema proposto foi da ordem de 86,43%.

Claro et al. 2016 aplicaram informação de textura com o uso da GLCM juntamente com a entropia nas imagens da retina para detectar o glaucoma. Os modelos de cores utilizados foram o *Red, Green, Blue* - (RGB), *Hue, Saturation, Intensity* - (HSI) e *Luminosity*, cromaticidade U e cromaticidade V - (LUV). A base de dados utilizada foi a RIM-ONE-1 e obtiveram uma acurácia da ordem de 93,03% aplicando o classificador MLP.

Orlando et al. 2017 apresentaram resultados de uma análise da utilização das CNNs pré-treinadas, a partir de imagens da retina para detecção automática do glaucoma. A Vgg-s e a Overfeat foram aplicadas nas imagens da base de dados DRISHTI para geração dos vetores de características. São utilizadas técnicas de pré-processamento da imagem com o objetivo de melhorar a imagem para a etapa de extração de características, sendo estas técnicas a equalização de histograma, pintura dos vasos e um corte na imagem

na região da cabeça do nervo óptico. A área sob a curva ROC (AUC) foi utilizada como métrica de avaliação tendo obtido resultado de 0,78 no classificador *Logistic Regression*.

### 3. Materiais e Métodos

O objetivo deste trabalho é realizar o diagnóstico do glaucoma em imagens da retina. Para isso aplicou-se na região do disco óptico experimentos com os descritores de textura, forma e (CNNs). Realizamos uma avaliação do potencial de descrição, em separado, de cada um destes descritores e concluímos que a união de diferentes características leva a melhores taxas de classificação. Assim, formamos uma nova representação de imagem usando o conjunto de recursos que produziu os melhores resultados.

#### 3.1. Método Proposto

O método proposto segue no fluxograma da Figura 2. Esta metodologia parte de uma imagem segmentada do DO seguida pela extração de características de textura, forma e CNNs. A partir da análise dos resultados obtidos, observamos que a concatenação das características da GLCM com a penúltima camada totalmente conectada das CNN-CaffeNet, CNN-Vgg-f, CNN-Vgg-m e CNN-Vgg-s obtiveram melhores resultados. Na etapa de seleção de atributos é utilizado o algoritmo de razão de ganho de informação. Por fim é aplicado o classificador RF para informar se a imagem é saudável ou glaucomatosa.

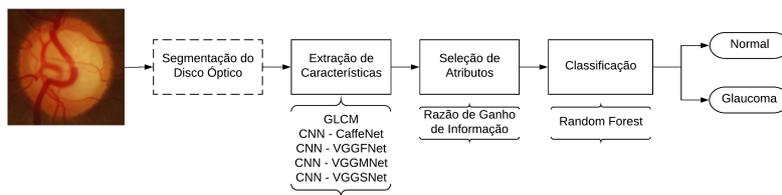


Figura 2. Fluxograma do método proposto neste estudo.

#### 3.2. Extração de Características, Seleção de Características e Classificação

De acordo com Salam et al. 2016 a identificação de imagens glaucomatosas pode ser eficientemente realizada usando os descritores de textura e forma. Assim, neste estudo, os resultados obtidos foram avaliados pelos descritores de textura e forma, bem como as CNNs. Para mais detalhes sobre os descritores aplicados, leia a seção 3.2 da dissertação.

Após a criação do vetor de características, realizou-se uma seleção de atributos. Esta seleção tem como objetivo eliminar atributos desnecessários e, conseqüentemente, simplificar o modelo de predição, melhorar a performance dos algoritmos, reduzir o custo computacional, além de fornecer um melhor entendimento sobre os resultados.

O algoritmo utilizado para seleção de atributos foi a razão de ganho, que é uma métrica de ganho de informação. Este algoritmo é um tipo de filtro, que tende a superestimar a qualidade dos atributos com muitos valores [Quinlan 1986]. Neste contexto, realizou-se 15 testes, variando o tamanho do vetor de características em 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% e 90%, para que possamos excluir os atributos menos significativos.

O software utilizado para classificação foi o *Waikato Environment for Knowledge Analysis* (WEKA) [Hall et al. 2009]. Os classificadores utilizados nos experimentos deste

trabalho foram o SVM, MLP e RF. Os parâmetros utilizados para cada classificador foram os padrões em WEKA e o método de validação utilizado foi o *k-fold cross-validation* (com  $k = 10$ ).

### 3.3. Métricas de Avaliação

Para avaliar o desempenho dos algoritmos de classificação foi utilizada a matriz de confusão. Fundamentado nestes valores da matriz, calculamos a Precisão (P), Área Sob a Curva ROC (AUC) e Acurácia (A) [Powers 2007].

Outra medida utilizada foi o Índice Kappa (K), que tem sido recomendado como medida apropriada da exatidão. Este índice toma todos os elementos da matriz em consideração, ao invés de apenas aqueles que retratam a quantidade de classificações verdadeiras, o que ocorre quando se calcula a exatidão global da classificação. De acordo com Landis and Koch 1977, k assume valores entre 0 e 100. A qualidade dos resultados podem ser avaliados pelo valor do índice Kappa, como segue:  $K \leq 20$ : Ruim;  $20 < K \leq 40$ : Moderado;  $40 < K \leq 60$ : Bom;  $60 < K \leq 80$ : Muito Bom;  $K > 80$ : Excelente.

## 4. Resultados e Discussão

Um dos desafios na detecção automática do glaucoma é a capacidade de diagnosticar a doença em bancos de dados com características distintas. No presente trabalho, usamos quatro bancos de dados: o banco de dados DRISHTI e três versões da RIM-ONE, totalizando 873 imagens, sendo 489 saudáveis e 384 glaucomatosas. A Tabela 1 mostra os resultados dos três classificadores usando os descritores individuais.

**Tabela 1. Resultados obtidos para cada descritor e classificador. Valores em negrito representam melhores resultados.**

Descritor	SVM				MLP				RF			
	P(%)	AUC(%)	A(%)	K(%)	P(%)	AUC(%)	A(%)	K(%)	P(%)	AUC(%)	A(%)	K(%)
LBP	31.40	50.00	56.01	0	74.40	80.90	74.45	48.12	74.40	82.50	74.45	47.91
GLCM	81.40	80.90	80.18	60.57	<b>85.10</b>	<b>92.90</b>	<b>84.99</b>	<b>69.69</b>	<b>83.10</b>	<b>90.50</b>	<b>83.16</b>	<b>65.70</b>
HOG	78.00	77.70	78.00	55.42	72.70	81.10	72.73	44.90	75.00	84.60	75.02	49.24
Tamura	69.50	68.50	69.64	37.54	70.40	78.90	70.33	39.98	70.80	79.10	70.90	40.19
GLRLM	63.40	62.70	63.68	25.67	74.30	80.30	74.34	47.70	75.60	83.90	75.71	50.44
Morphology	53.80	53.00	52.11	5.78	62.60	69.70	63.00	22.60	66.20	72.20	66.32	31.27
CNN-AlexNet	82.40	82.00	82.47	64.30	83.50	88.00	83.50	66.49	80.30	89.30	80.29	59.77
CNN-CaffeNet	80.80	80.20	80.87	60.91	81.15	87.90	81.55	62.40	81.60	90.20	81.55	62.18
CNN-Vgg-f	80.70	80.20	80.75	60.75	79.90	88.10	79.95	59.59	82.20	89.40	82.13	63.33
CNN-Vgg-m	82.10	81.80	82.13	63.68	83.30	89.20	83.27	66.04	81.80	89.50	81.78	62.97
CNN-Vgg-s	80.80	80.40	80.87	61.06	81.60	88.50	81.67	62.72	82.70	90.00	82.60	64.63
CNN-Vgg-16	78.20	77.90	78.23	55.78	80.10	87.80	80.06	59.51	80.90	88.70	80.87	60.77
CNN-Vgg-19	<b>82.70</b>	<b>82.40</b>	<b>82.70</b>	<b>64.89</b>	81.40	87.90	81.44	62.30	81.00	87.80	80.98	61.11

Comparando os resultados apresentados na Tabela 1 o RF obteve a maioria dos resultados com um Kappa acima de 60%. Podemos inferir que ele obteve melhor desempenho que os outros classificadores. Para configurar nosso vetor de características, decidimos utilizar os cinco descritores que alcançaram o melhor resultado no classificador RF. Desta forma, concatenamos os descritores: GLCM, CNN-CaffeNet, CNN-Vgg-f, CNN-Vgg-m e CNN-Vgg-s.

Realizamos testes para identificar o melhor subconjunto de características. Assim, usamos vetores de características com apenas 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% e 90%, dos atributos mais significativos, totalizando 90 testes

(15 vetores de características com diferentes comprimentos  $\times$  6 descritores). A Tabela 2 mostra o número de atributos que obtiveram melhores resultados.

Os descritores GLCM e CNNs obtiveram taxas de acerto de 82,59% para 86,48%. Estes descritores obtiveram resultados considerados “Muito Bom” de acordo com o índice Kappa. No entanto, ao concatená-los, o resultado melhorou significativamente, alcançando uma acurácia de 93,35% com apenas 4% dos atributos, o que corresponde a 659 atributos de 16.469. Como o índice Kappa foi 86,48%, este resultado é considerado “Excelente”.

**Tabela 2. Número de atributos e melhores resultados obtidos na etapa de seleção de atributos.**

Descritor	Data (%)	N. of Attributes	P(%)	AUC(%)	A(%)	K(%)
GLCM	100	85	86.50	95.00	86.48	72.52
CNN-CaffeNet	3	123	83.30	90.40	83.27	65.79
CNN-VGGFNet	70	2867	82.60	89.60	82.59	64.37
CNN-VGGMNet	2	82	83.50	89.90	83.50	66.38
CNN-VGGSNet	70	2867	82.60	89.80	82.59	65.67
GLCM+CNNs	4	659	<b>93.40</b>	<b>98.00</b>	<b>93.35</b>	<b>86.48</b>

Para compararmos este trabalho com Kotyk et al. 2016 e Claro et al. 2016, aplicamos os métodos destes nas 873 imagens disponíveis. Nossa proposta obteve melhor desempenho quando comparada a eles, com uma acurácia de 93,35% e Kappa de 86,48% que é um resultado considerado “Excelente”, enquanto os outros dois artigos atingiram um índice considerado “Muito Bom”.

Orlando et al. 2017 também utiliza redes neurais convolucionais pré-treinadas como descritor. No entanto, essas redes foram diferentes daquelas aplicadas neste estudo. Como não tivemos acesso ao código original do autor, executamos nosso método no banco de dados da DRISHTI (o mesmo que é utilizado em [Orlando et al. 2017]). A única métrica de avaliação apresentada pelos autores é a AUC com o valor de 76,20%. Nossa proposta obteve melhor desempenho neste banco de dados com uma AUC de 88,30%. Orlando et al. 2017 opta por utilizar apenas as CNNs como descritores. Em nosso descritor proposto, incluímos informações de textura (GLCM). Os resultados nos levam a concluir que as características da textura são fundamentais na caracterização da existência ou não de glaucoma em imagens da retina.

## 5. Conclusão

Muitos trabalhos identificados na literatura, demonstram que as pesquisas têm se direcionado para a área de processamento digital de imagens, na qual o diagnóstico automático do glaucoma é uma destas subáreas. Neste estudo, foi exposto uma estratégia para o diagnóstico automático do glaucoma, que utiliza descritores de textura e CNNs.

De acordo com a literatura, percebeu-se que os autores costumam utilizar suas próprias bases de dados ou apenas uma base de dados. No entanto, essa situação não representa um ambiente real do aplicativo do sistema. Assim, avaliamos nossa abordagem em quatro bancos de dados de imagens diferentes, cada uma com características específicas. A extração de características e a classificação são considerados as etapas mais importantes no diagnóstico automático de diferentes sistemas computacionais. Neste trabalho utilizados os descritores de textura, forma e CNNs, além de um algoritmo de seleção de atributos para auxílio no desempenho do método. O melhor resultado foi alcançado

através da concatenação de GLCM com CNNs, obtendo um índice Kappa “Excelente” e uma acurácia de 93,35%. Além disso, verificamos que o método proposto funciona bem mesmo quando há imperfeições na segmentação do DO.

Os resultados apresentados são promissores. Entretanto, podem ser melhorados. Como trabalhos futuros, propõe-se a extração de características na região da escavação (*cup*), além da região do disco óptico. Além disso, no presente trabalho, não realizamos um pré-processamento da imagem para remoção dos vasos sanguíneos. Assim, pretendemos investigar o efeito dessa operação no resultado final da classificação.

## Agradecimentos

Os autores gostariam de agradecer à Fundação de Amparo a Pesquisa do Estado do Piauí (FAPEPI) pelo patrocínio da nossa pesquisa.

## Referências

- Claro, M., Santos, L., Silva, W., Araújo, F., Moura, N., and Macedo, A. (2016). Automatic glaucoma detection based on optic disc segmentation and texture feature extraction. *CLEI Electronic Journal*, 19(2):5–5.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- Kotyk, T., Chakraborty, S., Dey, N., Gaber, T., Hassanien, A. E., and Snasel, V. (2016). Semi-automated system for cup to disc measurement for diagnosing glaucoma using classification paradigm. In *Proceedings of the Second International Afro-European Conference for Industrial Advancement*, pages 653–663. Springer.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- Mittapalli, P. S. and Kande, G. B. (2016). Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma. *Biomedical Signal Processing and Control*, 24:34–46.
- Orlando, J. I., Prokofyeva, E., del Fresnob, M., and Blaschko, M. (2017). Convolutional neural network transfer for automated glaucoma identification. In *12th International Symposium on Medical Information Processing and Analysis*, pages 101600U–101600U. International Society for Optics and Photonics.
- Powers, D. M. (2007). Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation, school of informatics and engineering, flinders university, adelaide, australia. Technical report, TR SIE-07-001, *Journal of Machine Learning Technologies* 2: 1 37-63.
- Quigley, H. A. and Broman, A. T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*, 90(3):262–267.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.
- Salam, A. A., Khalil, T., Akram, M. U., Jameel, A., and Basit, I. (2016). Automated detection of glaucoma using structural and non structural features. *SpringerPlus*, 5(1):1519.