

# Diagnóstico de glaucoma baseado em Deep Features

Lisle Faray de Paiva<sup>1</sup> ,  
José Mateus Boaro Carvalho<sup>1</sup> , Arthur Guilherme Santos Fernandes<sup>1</sup> ,  
Caio Manfredini da Silva Martins<sup>1</sup> , Geraldo Braz Junior<sup>1</sup>

<sup>1</sup>Laboratório de Visão e Processamento de Imagens - VIPLab  
Núcleo de Computação Aplicada - NCA  
Universidade Federal do Maranhão (UFMA)  
Av. dos Portugueses, 1966 - Bacanga, São Luís - MA, 65080-805, Brazil

{lisle, boaro, arthurgsf, caiomanfredini, geraldo}@nca.ufma.br

**Abstract.** *Glaucoma is a silent disease which permanently damages the optic nerve. Asymptomatic in its early stages, it's the second leading cause of blindness in the world. Several automatic diagnostic systems have been proposed, however, these systems are not capable of handling a wide range of images. Therefore, such methods are not feasible for use in screening programs. This paper proposes a methodology for efficiently detecting glaucoma that is capable of handling diverse images through feature extraction using Convolutional Neural Networks (CNNs). In this proposal, a total of 1090 images from four public data-sets were evaluated and it was concluded that the combination of pre trained CNNs and Specific Optimized Networks together with the use of the Logistic Regression classifier is promising in the detection of this pathology, obtaining accuracy's of 86.8% e 86.3%.*

**Resumo.** *O glaucoma é uma doença silenciosa que danifica permanentemente o nervo óptico. Assintomática em seu estágio inicial, é a segunda maior causa de cegueira no mundo. Vários sistemas de diagnóstico de glaucoma automático têm sido propostos, no entanto, existe ainda grande dificuldade de lidar com uma grande diversidade de imagens. Em razão disso, tais métodos não são viáveis para o uso em programas de triagem. Este trabalho propõe uma metodologia com a finalidade de detectar de uma maneira eficiente o glaucoma, que seja capaz de lidar com imagens diversas, através da extração de características usando Redes Neurais Convolucionais (CNNs). Nesta proposta, foi avaliado um total de 1090 imagens de quatro bases de dados públicas e foi concluído que a junção de CNNs pré-treinadas e Redes Específicas juntamente com o uso do classificador Regressão Logística são promissores na detecção desta patologia, obtendo acurácias de 86.8% e 86.3%.*

## 1. Introdução

O glaucoma é uma das principais doenças causadoras de cegueira em todo o mundo e deverá afetar cerca de 80 milhões de pessoas até 2020 [Quigley and Broman 2006]. Sua principal consequência é a perda irreparável de visão. Em seus estágios iniciais o glaucoma tende a ser assintomático e seus sintomas só aparecem apenas nos estágios mais avançados da doença. A detecção precoce desta patologia é imprescindível, pois existem tratamentos que evitam sua progressão e, conseqüentemente, a perda da visão do paciente.

O diagnóstico de glaucoma pode ser dado a partir da análise de imagens de fundo de olho, uma vez que a quantidade de perda de fibras do nervo óptico tem 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, o que acarreta em um valor da relação Escavação/Disco(**Cup to Disc ratio** - CDR) aumentado [Mittapalli and Kande 2016].

Devido ao impacto do glaucoma e a incapacidade de uma triagem eficiente da população tem-se métodos de diagnóstico através do uso de técnicas de aprendizado de máquina a fim de diagnosticar automaticamente e indicar acompanhamento especializado para tratamento. Abordagens diferentes têm sido desenvolvidas para o diagnóstico de glaucoma com CNNs, como em [Raghavendra et al. 2018] que é utilizado uma CNN de 18 camadas e alcança a acurácia de 98.13% e em [Abbas 2017] é utilizado uma CNN e uma *Deep-belief Network* (DBN) atingindo a precisão de 84%.

Neste trabalho é desenvolvido um método para a extração das características utilizando Redes Neurais Convolucionais afim de detectar o glaucoma, visando contribuir para a área de saúde auxiliando os profissionais para uma triagem eficiente e precisa, uma vez que esta triagem é feita manualmente.

## 2. Metodologia

Este trabalho desenvolve um método para o diagnóstico de glaucoma baseado na extração de características usando Redes Neurais Convolucionais. Primeiramente as imagens da base de dados são alimentadas para duas arquiteturas: a VGG16<sup>1</sup> pré-treinada com os pesos da Imagenet e a VGG16 ajustada com rede específica com parâmetros otimizados. Após o treinamento, são extraídos os vetores características de cada rede e concatenados. Esses vetores por sua vez serão passados para um classificador para realizar o diagnóstico. A Figura 1 ilustra o método proposto com o fluxo das etapas. Cada uma das etapas são descritas em detalhes nas subseções seguintes.

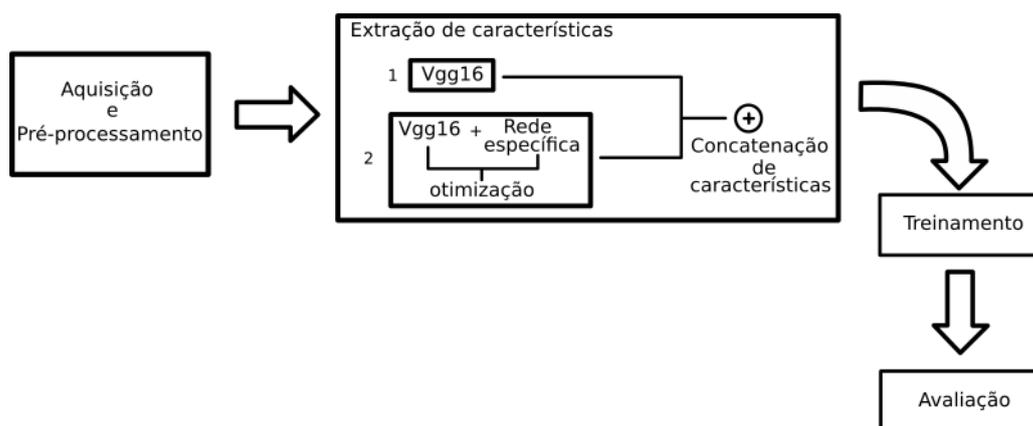


Figura 1. Fluxo de etapas da metodologia proposta

### 2.1. Aquisição e Pré-processamento

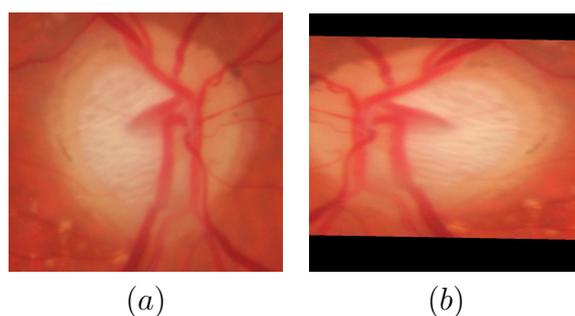
A primeira etapa da metodologia consiste na aquisição da base de dados. A base de dados é composta pela junção de quatro bases já existentes: Rim-One r1 [Fumero et al. 2011],

<sup>1</sup>Very Deep Convolutional Networks for Large-Scale Image Recognition

Rim-One r2 [Fumero et al. 2011], iChallenge Gon [of Technology 2018], Kaggle 1000 fundus image [linchundan 2018]. Esta possui no total 1090 imagens, sendo 303 imagens de olhos glaucomatosos e 787 imagens de olhos normais.

Por se tratar de uma base de dados desbalanceada, inicialmente é realizado geração sintética das imagens com a finalidade de distribuir os quantitativos de imagens pela classes. Os parâmetros escolhidos para gerar variações das imagens originais foram: intervalo de cisalhamento, variação de largura e altura, variação do ângulo de rotação, variação de zoom, preenchimento lateral com 0 e rotação 180°, os quais foram aplicadas aleatoriamente.

Na Figura 2, pode-se observar a diferença na imagem original e a imagem sintética. Para cada imagem de glaucoma, são geradas 10 imagens sintéticas, enquanto que para cada imagem de olho normal são geradas 4 imagens. A diferença na quantidade de imagens geradas com glaucoma para as normais é a maneira escolhida para balancear a base. Após a geração de imagens sintéticas é realizado o redimensionamento das imagens para altura 612 e largura 900, com a finalidade de padronizar a entrada da CNN que, por sua vez, só recebe imagens de mesmo tamanho.



**Figura 2. (a) Imagem normal; (b) Imagem sintética**

## 2.2. Extração de características

Após o pré-processamento, realiza-se a extração de características. Este processo consiste na retirada da última camada da rede (seu classificador) e obter a saída da camada de *pooling* logo após a última camada de convolução. Neste trabalho são extraídos 2 vetores característicos de duas arquiteturas: a rede VGG16 [Han et al. 2015] original e a VGG16 com *fine tuning* estimado pelo otimizador. A rede neural VGG16 foi escolhida para este trabalho por ser uma rede de arquitetura simples, assim dando maior destaque para o método. A extração de cada um será explicado mais a fundo nas subseções a seguir.

Após a extração, ocorre a concatenação dos vetores característicos da rede VGG16 propriamente dita e o vetor da VGG16 com a rede específica estimada pelo otimizador. Com a concatenação dos dois vetores, o classificador tem uma maior quantidade de características mais proeminentes com o intuito de obter melhores resultados. Assim existem 2 vetores de características, o primeiro gerado pela rede genérica e o segundo gerado pela rede específica.

Uma Rede Neural Convolutiva (CNN, do inglês *Convolutional Neural Network*) é uma classe de rede neural artificial do tipo *feed-forward*, que vem sendo aplicada com

sucesso no processamento e análise de imagens digitais. As redes convolucionais são inspiradas nos processos biológicos [Matsugu et al. 2003]. Nelas o padrão de conectividade entre os neurônios é inspirado na organização do córtex visual dos animais. Neurônios corticais individuais respondem a estímulos apenas em regiões restritas do campo de visão conhecidas como campos receptivos. Os campos receptivos de diferentes neurônios se sobrepõem parcialmente de forma a cobrir todo o campo de visão.

Uma CNN demanda mínimo pre-processamento quando comparada a outros algoritmos de classificação de imagens. Isso significa que a rede "aprende" os filtros que em um algoritmo tradicional precisariam ser implementados manualmente. Essa independência de um conhecimento a priori e do esforço humano no desenvolvimento de suas funcionalidades básicas pode ser considerada a maior vantagem de sua aplicação.

Para este trabalho, foi escolhida a CNN VGG16, nomeada pelo grupo *Visual Geometry Group* que a desenvolveu. A rede proposta por [Han et al. 2015] tem sua arquitetura demonstrada na Figura 3, composta por convoluções seguidas por pooling.

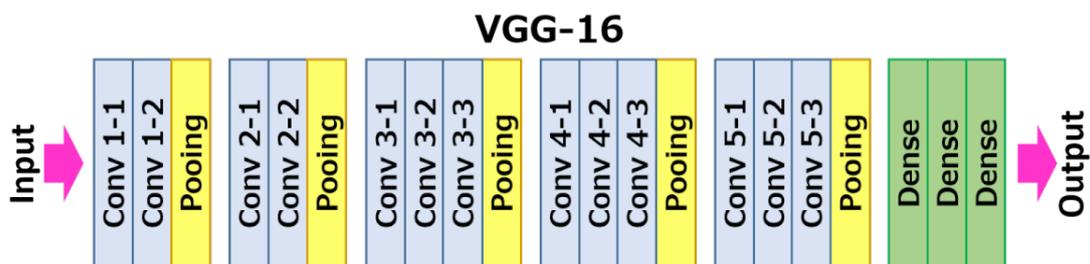


Figura 3. Arquitetura da VGG16

### 2.2.1. Extração de Características com VGG16 pré treinada

A extração das características usando a VGG16 pré treinada consiste inicialmente em carregar a arquitetura juntamente dos pesos da *Imagenet*, com a finalidade de aproveitar o treinamento básico de suas características. Depois é removido seu classificador e inserido uma camada de *Global Average Pooling* para a saída da rede ser um vetor de características.

### 2.2.2. Extração de Características com VGG16 otimizada

Este método também implementa extração de características da rede VGG16 otimizada para o problema específico.

O processo de otimização consiste em analisar resultados de modelos anteriores para sugerir novos parâmetros para um próximo modelo. Assim, a cada novo modelo, o resultado tende a ser mais positivo. Os parâmetros que o algoritmo de otimização pôde escolher são determinados em um espaço de busca referenciado na Tabela 1.

O otimizador foi implementado sobre a biblioteca *HyperOpt* que disponibiliza o algoritmo *Tree-Structured Parzen Estimator*(TPE) [Bergstra et al. 2013]. Segundo [Fernandes et al. 2019], o algoritmo TPE se utiliza de decisões bayesianas para escolha

**Tabela 1. Espaço de Busca**

Parâmetros	Opções de Busca		
Quantidade de convoluções	0	1	2
Camada de saída	block5 convolução 3	block5 convolução 2	
Pooling	Average Pooling	Max Pooling	
Quantidade de neurônios	256	128	
Quantidade de denses	1	2	
Classificador	Sigmoid	Softmax	
Dropout	0.5	0.4	0.3
Class weight	TRUE	FALSE	

de parâmetros, tendo como principal característica, a convergência mais rápida para a solução ótima se comparado com a decisão aleatória ou em *grid*.

Após a estimação do melhor modelo, utiliza-se o melhor modelo estimado para extração das características. A extração das características da VGG16 com *fine tuning* é um processo similar ao da rede pura, explicado na seção anterior. A suposição é que as características geradas de maneira específica podem incluir informações complementares para determinar os padrões nas imagens.

### 2.3. Treinamento

Após ter as características extraídas e armazenadas, seguiu-se para a etapa de classificação. Nesta fase, é utilizado o classificador Regressão Logística [Yu et al. 2011]. Este foi escolhido por trabalhar bem com características lineares, as quais são esperadas como saída das redes utilizadas.

Na regressão logística, a probabilidade de ocorrência de um evento pode ser estimada diretamente. No caso da variável dependente  $Y$  assumir somente dois possíveis estados e haver um conjunto de  $p$  variáveis independentes  $X_1, X_2, \dots, X_p$  o modelo de regressão logística pode ser escrito da seguinte forma:

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}} \quad (1)$$

onde,  $g(x) = B_0 + B_1X_1 + \dots + B_pX_p$ .

Os coeficientes  $B_0, B_1, \dots, B_p$  são estimados a partir do conjunto de dados, pelo método da máxima verossimilhança, em que encontra uma combinação de coeficientes que maximiza a probabilidade da amostra ter sido observada. Considerando uma certa combinação de coeficientes  $B_0, B_1, \dots, B_p$  e variando os valores de  $X$ , observa-se que a curva logística tem um comportamento probabilístico no formato da letra **S**.

Por causa de seu comportamento, a Regressão Logística é muito utilizada para classificação de classes discretas, como no caso do glaucoma. Para esta base é definido glaucoma como 1 e normais como 0. Assim classificando as características de acordo, com  $> 0,5$  como 1 e  $< 0,5$  como 0.

Utilizando o regressão logística, metodologia proposta realiza 3 experimentos de classificação. O primeiro usando apenas as características da VGG16 original, o segundo

usando as características da VGG16 otimizada e o terceiro concatenando os dois conjuntos de características.

Os três experimentos foram conduzidos com a base dividida em treino, teste e validação, sendo que 70% foi para o treino e 30% para validação e teste. Após a divisão o classificador é treinado e testado, tendo as métricas de avaliação armazenadas.

## 2.4. Avaliação

Com a extração e classificação concluídas, começa a etapa de avaliação das métricas. As métricas avaliadas são: sensibilidade, precisão, especificidade, *fscore* e acurácia.

Primeiramente é calculada a matriz de confusão. A partir dos índices da matriz, são calculadas as métricas demonstradas nas equações abaixo:

$$Sensibilidade = \frac{tp}{tp + fn} \quad (2)$$

$$Precisão = \frac{tp}{tp + fp} \quad (3)$$

$$Especificidade = \frac{tn}{tn + fp} \quad (4)$$

$$Acurácia = \frac{tp + tn}{tn + tp + fn + fp} \quad (5)$$

$$F1 - score = 2 \cdot \left( \frac{Precisão \cdot Sensibilidade}{Precisão + Sensibilidade} \right) \quad (6)$$

onde tp = verdadeiro positivo, tn = verdadeiro negativo, fp = falso positivo, fn = falso negativo.

## 3. Resultados

Neste trabalho é desenvolvido um método para diagnóstico de glaucoma baseado em extração de características utilizando Redes Neurais Convolucionais. São extraídas as características de duas redes: a VGG16 original e a VGG16 otimizada para o problema específico do diagnóstico. Após a extração são realizados três experimentos, sendo eles a classificação do vetor das características da VGG16 pré-treinada, do vetor da VGG16 com *fine tuning* e da concatenação dos dois vetores.

Para o aperfeiçoamento da VGG16, foi executado um total de 35 testes ao longo de 5 dias. O melhor modelo gerado atingiu a acurácia de 88,22% e os parâmetros escolhidos pelo otimizador estão referenciados na Tabela 2.

Utilizando o modelo gerado, foram realizados os três experimentos que combinam características geradas por cada rede. Os resultados obtidos são apresentados na Tabela 3.

Ao analisar os resultados, observamos que os melhores resultados acontecem com a concatenação de características, mesmo que a diferença seja pequena em relação a rede

**Tabela 2. Melhores parâmetros escolhidos**

Parâmetros	Opções escolhidas
Quantidade de convoluções	1
Camada de saída	block5 convolução 3
Pooling	Average Pooling
Quantidade de neurônios	256
Quantidade de denses	2
Classificador	Sigmoid
Dropout	0.3
Class weight	TRUE

**Tabela 3. Resultados da classificação das características**

Características	Sensibilidade	Especificidade	Precisão	F-score	Acurácia
VGG16 original	0,510	0,814	0,688	0,586	0,679
VGG16 fine tuning	0,863	0,868	0,835	0,849	0,863
Concatenação	0,868	0,868	0,848	0,857	0,868

estimada. As melhoras acontecem em todos os índices verificados. Esta melhora se dá ao fato de a base de dados do classificador ter sido bem maior em relação aos outros dois experimentos, já que ,enquanto que no primeiro e segundo experimento a base utilizada pelo classificador foi um vetor de características, no terceiro sua base foi a concatenação dos dois vetores. Embora melhores, os resultados da concatenação são similares aos da rede específica, isso se dá pelo fato do classificador ter aprendido o mesmo tanto na concatenação quanto na rede específica.

Verifica-se que a otimização da VGG16 para o problema específico do glaucoma atingiu métricas melhores do que a VGG16 original. Assim evidenciado pela discrepância em precisão com a original tendo uma precisão de 68.88% e a VGG16 com *fine tuning* obtendo uma precisão de 83.56%. A VGG16 original teve uma certa dificuldade em classificar os olhos glaucomatoso e normais, porque seus pesos originais são da base de dados *Imagenet* que não inclui imagens de olhos glaucomatosos ou normais. Já a rede especificada pelo otimizador teve a chance de treinar utilizando a base de dados de glaucoma assim tendo seus pesos ajustados para o problema específico e obtendo um melhor desempenho.

#### 4. Conclusão

Neste trabalho foi elaborado um método através da extração de características utilizando Redes Neurais Convolucionais com a finalidade de diagnosticar imagens de fundo de olhos glaucomatosos e normais. O método consiste na extração de características de 2 redes diferentes sendo elas a VGG16 original e a VGG16 com parametrização otimizada para problema específico.

O método obteve um resultado de 0,868 e 0,848 para sensibilidade e precisão especificamente, mesmo que empregado para uma base de dados complexa constituída pela junção de vários *datasets*.

Pretende-se explorar como trabalhos futuros outras arquiteturas de redes convo-

lucionais mais complexas como a *DenseNet* pro exemplo e uma estimação mais ampla da rede especializada como forma de melhoria dos resultados. Além disso, pretende-se aumentar a base de dados através do adicional de mais bases públicas.

## Referências

- Abbas, Q. (2017). Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning. *Int J Adv Comput Sci Appl*, 8(6):41–5.
- Bergstra, J., Yamins, D., and Cox, D. D. (2013). Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In *Proceedings of the 12th Python in science conference*, pages 13–20. Citeseer.
- Fernandes, A. G. S., da Silva Martins, C. M., de Moura Lima, A. C., Junior, G. B., de Almeida, J. D. S., and de Paiva, A. C. (2019). Meta aprendizagem de extração de características aplicada ao diagnóstico de glaucoma. In *Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 342–347. SBC.
- Fumero, F., Alayón, S., Sanchez, J., Sigut, J., and Gonzalez-Hernandez, M. (2011). Rimone: An open retinal image database for optic nerve evaluation. pages 1 – 6.
- Han, S., Mao, H., and Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149*.
- linchundan (2018). 1000 fundus images with 39 categories.
- Matsugu, M., Mori, K., Mitari, Y., and Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5-6):555–559.
- 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.
- of Technology, B. I. (2018). ichallenge-gon.
- 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.
- Raghavendra, U., Fujita, H., Bhandary, S. V., Gudigar, A., Tan, J. H., and Acharya, U. R. (2018). Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Information Sciences*, 441:41–49.
- Yu, H.-F., Huang, F.-L., and Lin, C.-J. (2011). Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, 85(1):41–75.