



Classificando Graus de Pterígio Utilizando Aprendizado de Máquina

Classifying Degrees of Pterygium Using Machine Learning

Clasificación de Grados de Pterigion Mediante Aprendizaje Automático

Erik Jhones Freitas do Nascimento¹, David Victor Cavalcante¹, Ana Carolina Silva Abreu¹, Diego Parente Paiva Mesquita², Amauri Holanda de Souza Júnior³

RESUMO

Descritores: Pterígio;
Aprendizado de Máquina;
Redes Neurais de
Computação

Objetivos: Pterígio é uma condição ocular resultante do crescimento de tecido subconjuntival ao longo da córnea, podendo comprometer a visão quando em grau avançado. Essa condição é particularmente comum entre trabalhadores frequentemente expostos à luz, muitos dos quais vivem em regiões com acesso precário à rede de saúde, o que naturalmente dificulta o diagnóstico. Este trabalho apresenta uma abordagem automatizada para a classificação de graus de avanço de Pterígio usando redes neurais a partir de imagens. **Métodos:** Para validar a abordagem proposta, é utilizado e disponibilizado um novo banco de dados de imagens em diversos graus de avanço da patologia. **Resultados:** Dentre os modelos avaliados, destacam-se as redes neurais convolucionais, obtendo acurácia média de 95%. **Conclusão:** Uma nova base de dados foi disponibilizada. Adicionalmente, planeja-se disponibilizar o método na forma de um aplicativo móvel.

ABSTRACT

Keywords: Pterygium;
Machine Learning; Neural
Networks

Objective: Pterygium is an eye condition resulting from the growth of subconjunctival tissue along the cornea, which can compromise vision when in an advanced degree. This condition is particularly common among workers frequently exposed to light, many of whom live in regions with poor access to the health network, which naturally makes diagnosis difficult. This work presents an automated approach for the classification of degrees of advance of Pterygium using neural networks from images. **Methods:** To validate the proposed approach, a new image database is used and made available in varying degrees of pathology progress. **Results:** Among the models evaluated, the convolutional neural networks stand out, obtaining an average accuracy of 95%. **Conclusion:** A new database was made available. Additionally, we intend to make the method available as a mobile application.

RESUMEN

Descriptores: Pterigion;
Aprendizaje automático;
Redes Neuronales de la
Computación

Metas: La pterigion es una afección ocular resultante del crecimiento de tejido subconjuntival a lo largo de la córnea, que puede comprometer la visión cuando está en un grado avanzado. Esta condición es particularmente común entre los trabajadores frecuentemente expuestos a la luz, muchos de los cuales viven en regiones con escaso acceso a la red de salud, lo que naturalmente dificulta el diagnóstico. **Métodos:** Este trabajo presenta un enfoque automatizado para la clasificación de grados de avance de la pterigion utilizando redes neuronales a partir de imágenes. **Resultados:** Para validar el enfoque propuesto, se utiliza una nueva base de datos de imágenes y se pone a disposición en diversos grados de progreso de la patología. Entre los modelos evaluados destacan las redes neuronales convolucionales, obteniendo una precisión media del 95%. **Conclusión:** Se puso a disposición una nueva base de datos. Además, el método está disponible como aplicación móvil.

¹ Bacharel (a) em Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Maracanaú, CE, Brasil.

² Doutorando em Ciência da Computação, Aalto University (A?), Espoo, Finlândia.

³ Professor Associado do PPGCC, Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Maracanaú, CE, Brasil.

INTRODUÇÃO

O Pterígio é uma patologia ocular caracterizada pelo crescimento fibrovascular, resultante de tecido subconjuntival, através do limbo até a córnea. A taxa de prevalência de Pterígio varia de 0,7% a 30% entre diferentes populações ao redor do mundo. Notavelmente, a maior concentração dos casos ocorre em regiões equatoriais⁽¹⁾. Dentre as possíveis causas para o surgimento dessa patologia, acredita-se que a principal é a exposição prolongada aos raios ultravioleta (UV)⁽²⁾. O Pterígio possui 4 graus que são classificados com base no tamanho do avanço do tecido fibrovascular ao longo da córnea, sendo esses graus definidos como: Grau I (menor que 2 mm); Grau II (entre 2 a 4 mm); Grau III (entre 4 a 6 mm); Grau IV (maior que 6 mm)⁽³⁾. A Figura 1 mostra os graus de avanço em olhos afetados.

Pterígio é prevalente entre trabalhadores que são frequentemente expostos à luz solar, tais como pescadores, fazendeiros e construtores civis. A maior parte dessas pessoas raramente estão cientes da existência do Pterígio. Portanto, a patologia só é diagnosticada quando um profissional da saúde a visita para realizar exames básicos de saúde, que incluem o rastreamento de doenças oculares. Devido à quantidade de testes que precisam ser realizados para cada ciclo de verificação, uma ferramenta moderna de fácil pulverização seria instrumental para tornar o processo de triagem mais rápido⁽⁴⁾. Idealmente, tal ferramenta deveria ser capaz de prover um diagnóstico primário mais completo, incluindo uma estimativa do grau de severidade da patologia.

Sistemas de Auxílio ao Diagnóstico (*Computer-Aided Diagnosis - CAD*) visam auxiliar médicos no diagnóstico de diversas patologias. Usualmente, esses sistemas utilizam técnicas de processamento de sinais/imagens e aprendizado de máquina para fornecer estimativas relacionadas ao diagnóstico de interesse, provendo assim um diagnóstico suplementar. CADs também podem auxiliar no atendimento de mutirões e na prevenção de fadiga por parte do médico-especialista, prevenindo assim a ocorrência de diagnósticos errôneos⁽⁵⁻⁶⁾.

Neste trabalho, é proposto o desenvolvimento de um CAD para a classificação de graus de avanço de Pterígio a partir de imagens de olhos possivelmente afetados por essa patologia. Para isso, é realizado um estudo comparativo de técnicas de aprendizado de máquina (*machine learning, ML*) baseadas em redes neurais. Em particular, são considerados os modelos de regressão logística⁽⁷⁾, *perceptrons* multicamadas

(*MultiLayer Perceptrons, MLPs*)⁽⁸⁾ e redes neurais convolucionais (*Convolutional Neural Networks, CNNs*)⁽⁸⁾. Além dos códigos usados neste estudo, uma nova base de dados de imagens rotuladas, incluindo olhos saudáveis e diferentes graus de Pterígio, será disponibilizada publicamente, juntamente com um protótipo de aplicativo para classificação de imagens em dispositivos móveis, tais como *smartphones* e *tablets*⁸.

Dentre os modelos avaliados, as redes neurais convolucionais obtiveram maior êxito na predição do grau de avanço de Pterígio, com *F1-score* e acurácia média de aproximadamente 95%. Observou-se também que um modelo linear mais simples não é capaz de capturar as nuances que caracterizam os diferentes graus de Pterígio. Percebeu-se também que, a utilização de escala de cinza ao invés de escala RGB degrada a performance dos classificadores.

Além do estudo de métodos clássicos de ML para classificação de graus de Pterígio, a fim de fomentar o desenvolvimento de novas metodologias, o presente artigo também contribui com a disponibilização de uma nova base de dados contendo imagens pré-processadas. Vale ressaltar também o potencial impacto socioeconômico do uso de estratégias automáticas para auxílio ao diagnóstico de Pterígio, beneficiando populações com acesso limitado à rede de saúde.

Trabalhos Relacionados

Apesar do uso difundido de CNN como uma das principais ferramentas para diagnósticos automáticos de patologias oftalmológicas, verificou-se que sua utilização para a classificação de Pterígio em específico não é amplamente estudada. Foram encontrados na literatura somente quatro trabalhos que tratam da classificação desta patologia. No entanto, esses trabalhos⁽⁹⁻¹¹⁾ empregam somente uma abordagem de classificação binária, Pterígio ou não Pterígio; ou apresentam métodos para segmentar e identificar a doença⁽⁴⁾. A literatura não possui qualquer trabalho que apresente um método que realize predições a nível do grau do desenvolvimento da doença. Essa escassez de abordagens pode ser explicada pela ausência de um banco de dados público, que possibilite a avaliação de diversos modelos.

MÉTODOS

Nesta seção, são apresentados os materiais e o método

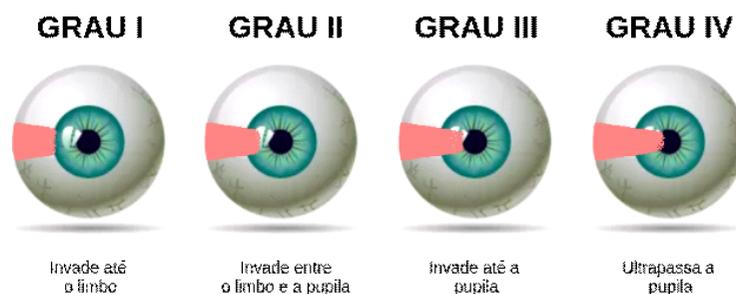


Figura 1 – Os quatro graus de avanço de Pterígio. A cor rosa indica tecido que avança sobre o olho. No acometimento mais leve, somente o limbo é afetado. No caso mais extremo, o tecido fibrovascular anômalo invade a pupila.

* A base de imagens e os produtos de softwares serão disponibilizados após o processo de revisão.

proposto neste trabalho. O método engloba etapas que vão desde a obtenção da base de imagens e pré-processamento das imagens ao treinamento e avaliação da qualidade preditiva dos modelos de ML. A Figura 2 ilustra, de maneira geral, as etapas da metodologia adotada. Nas subseções subsequentes, cada uma dessas etapas será discutida mais detalhadamente.

Base de Dados

Para projetar classificadores automáticos capazes de discriminar vários graus de pterígio, é necessário definir um conjunto de dados com imagens rotuladas de olhos saudáveis (classe 0) e olhos acometidos por Pterígio (classes 1 a 4). Esse conjunto pode então ser usado para treinar e validar diversos classificadores.

O conjunto de dados obtido para este trabalho é composto por 404 imagens no formato JPEG (*Joint Photographic Experts Group*). Cada imagem possui tamanho 3000 x 3000 *pixels* em escala RGB (*Red Green Blue*). As imagens foram disponibilizadas pelo Instituto de Oftalmologia de Maranguape, situado no município homônimo, no estado do Ceará. As imagens foram capturadas através de um auto refrator de *Greens TCI*. A base de dados possui 29 amostras para olhos sem a patologia (Grau 0); 20 amostras para Pterígio grau I; 47 para grau II; 207 para grau III; 101 para grau IV. Além disso, cada imagem foi rotulada por especialistas na área.

Pré-processamento

O pré-processamento das imagens é realizado para facilitar a tarefa de aprendizagem e divide-se em dois passos. O primeiro consiste no redimensionamento das imagens para o tamanho 128x128 *pixels*, com o objetivo de reduzir o número de parâmetros, bem como o custo computacional, dos algoritmos de classificação. Ressalta-

se que foram testados outros tamanhos de imagens, porém, o tamanho 128x128 foi o que obteve a melhor relação de custo e desempenho.

O segundo passo consiste em aumentar o conjunto de dados de treinamento com ligeiras modificações das imagens já existentes, num processo também conhecido como *data augmentation*⁽¹²⁾. Para cada imagem foram gerados novos exemplares rotacionados, transladados, expandidos ou contraídos. Além de facilitar o aprendizado, o objetivo desse passo é tornar os modelos treinados robustos a pequenas alterações na forma como a imagem foi adquirida. Essa é uma forma de prevenir o bem conhecido problema de *overfitting* em aprendizado de máquina, particularmente importante em base de dados com poucas amostras e modelos muito expressivos. Mais especificamente, foram geradas 24 amostras adicionais a partir de cada imagem da base original. Portanto, a base de dados de treino final contém 10100 imagens. A Figura 3 mostra o resultado do pré-processamento de imagens da base de dados.

Treinamento dos Modelos Neurais

Neste trabalho, o problema de estimar o grau de Pterígio é tratado como um problema de aprendizado supervisionado⁽⁸⁾, no qual as imagens correspondem às entradas e os graus da patologia definem as classes do problema.

Foram considerados os seguintes modelos de classificação: regressão logística multiclasse (*Logistic Regression*, LR), *perceptrons* multicamadas (MLPs) e redes convolucionais (CNNs). Vale a pena ressaltar que um modelo de regressão logística pode ser interpretado como a uma MLP sem camadas ocultas e equipado com a função *softmax*. Essa escolha visa cobrir diferentes níveis de complexidade, variando desde um modelo linear (i.e., LR),

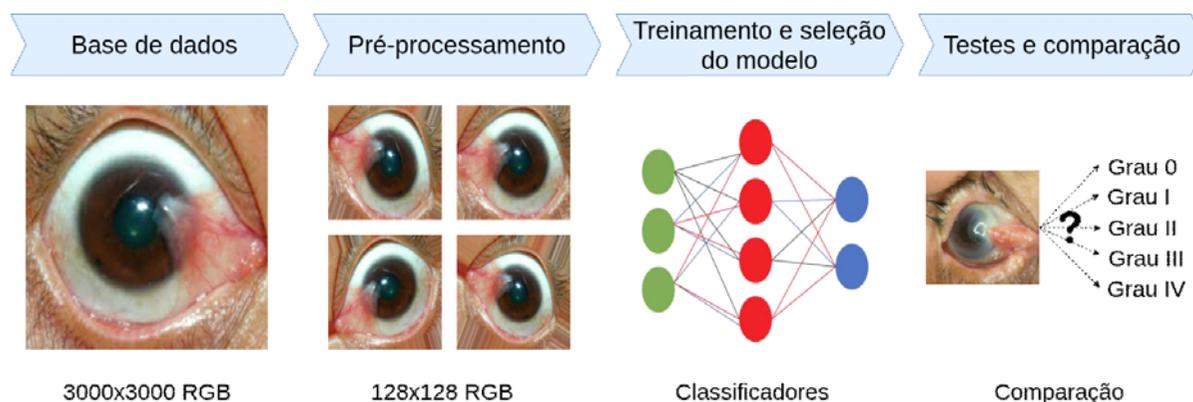


Figura 2 – A metodologia adotada compreende as etapas de: *i*) obtenção de uma base de dados contendo tanto imagens de olhos saudáveis como olhos afetados com pterígio em seus diversos graus; *ii*) pré-processamento das imagens a fim de facilitar o aprendizado; *iii*) treinamento e seleção de modelo; *iv*) comparação entre modelos.

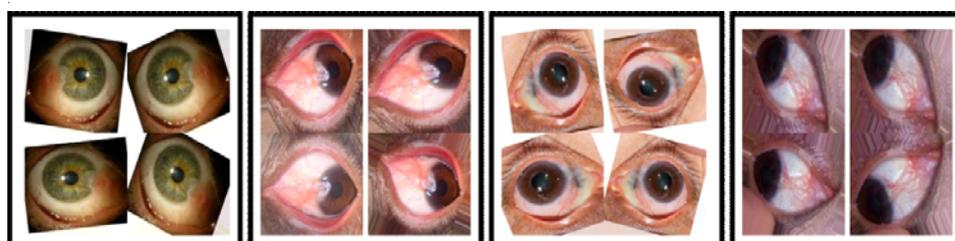


Figura 3 – Exemplo de imagens redimensionadas e pré-processadas através da rotação das imagens originais.

passando por uma alternativa mais expressiva não-linear (i.e., MLP), até redes neurais que representam o estado da arte para classificação de imagens (CNNs), capazes de explorar a regularidade presente em imagens e aprender, por exemplo, representações invariantes à translação (*shift-invariance*) ao longo de suas camadas^(8,13).

Como pré-processamento adicional para os modelos LR e MLP, para cada imagem, os canais RGB foram vetorizados e concatenados. A porcentagem de imagens selecionadas para treino e teste foram, respectivamente, 80% e 20%, dos quais 20% das amostras separadas para treino foram usadas para validação de modelo. As partições das imagens foram decididas de maneira aleatória, usando amostragem uniforme sem reposição. Todos os modelos foram treinados usando *Adam*⁽¹⁴⁾ sobre a função de custo de entropia cruzada (*cross-entropy*), com gradientes estimados de maneira “estocástica” com *mini-batches* de tamanho 64. Como os *mini-batches* são escolhidos sem reposição, as estimativas de gradiente são viesadas, e este procedimento é melhor referenciado como *random shuffling*⁽¹⁵⁾.

O modelo MLP possui 5 camadas intermediárias, cada uma com 100 neurônios. O modelo CNN possui 5 camadas convolucionais — com 16, 32, 32, 64 e 64 filtros respectivamente — e cada um dos filtros possui tamanho 3. Empregou-se ainda taxa de *dropout* igual 0.25 para todas as camadas convolucionais e *max-pooling* de tamanho 2 após cada camada. Além disso, a componente densa da CNN possui 1 camada oculta com 512 neurônios. A função de ativação ReLU foi utilizada em todas as camadas, exceto na componente densa, que emprega a função sigmoide. Cada arquitetura foi treinada por até 300 épocas, com parada prematura (*early stopping*) e paciência de 30 épocas. Os modelos com menor erro de validação foram selecionados para a etapa de teste. Todos os métodos foram implementados na linguagem de programação *Python*, utilizando a biblioteca *TensorFlow* sem utilização de *hardware* especializado.

Testes e Comparação

O processo descrito na subseção anterior (treinamento

dos modelos neurais) foi repetido 10 vezes, com partições de treino e testes distintas. Para comparar os classificadores desenvolvidos, utilizou-se as métricas comumente utilizadas e aceitas pela comunidade científica para análise de desempenho de sistemas CAD⁽¹⁶⁾: acurácia (ou taxa de acerto), precisão (*precision*), sensibilidade (*recall*) e *F1-score*. Vale ressaltar que as três últimas são definidas no contexto de classificadores binários. As equações (1), (2) e (3) descrevem a sensibilidade, a precisão e o *F1-score*, respectivamente:

$$recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

em que TP e TN denotam os números de verdadeiros positivos e negativos, e FN e FP os números de falsos positivos e negativos, respectivamente. Em outras palavras: a sensibilidade é a taxa de acerto entre as amostras que deveriam ser classificadas como positivas; a precisão é a fração de verdadeiros positivos dentre todas as amostras classificadas como positivas; e o *F1-score* é a média harmônica entre a precisão e a sensibilidade. Como este trabalho engloba um problema multiclasse e as estatísticas citadas são apropriadas para problemas binários, os valores de *recall*, *precision* e *F1-score* são ponderados pelo número de amostras para cada classe.

RESULTADOS E DISCUSSÃO

A Tabela 1 reporta a média e o desvio padrão para cada uma das métricas de avaliação utilizadas neste trabalho, extraídas no conjunto de teste ao longo de 10 repetições. Avaliou-se os modelos MLP e LR tanto para em escala de cinza quanto para escala RGB. A escala de cinza foi obtida através da média dos pixels ao longo dos canais RGB.

Tabela 1 – Comparação entre os resultados dos classificadores: média e desvio padrão. Os valores de *recall*, *precision* e *F1-score* são médias ponderadas, com pesos dados pelo número de amostras para cada classe. Note que CNNs atingem melhor performance quando comparadas aos demais métodos, em todas as métricas.

Modelo	<i>recall</i>	<i>precision</i>	<i>F1-score</i>	acurácia
CNN	0.95±0.01	0.95±0.01	0.95±0.01	0.95±0.01
MLP (cinza)	0.83±0.05	0.83±0.05	0.83±0.05	0.83±0.05
MLP (RGB)	0.93±0.02	0.93±0.02	0.93±0.02	0.93±0.02
LR (cinza)	0.39±0.06	0.43±0.03	0.36±0.04	0.39±0.06
LR (RGB)	0.49±0.07	0.52±0.03	0.47±0.06	0.49±0.07

Tabela 2 – *F1 scores* por classe (grau de Pterígio). CNNs apresentam a melhor performance para todas as classes. Dentre os diferentes graus de Pterígio, o Grau 1 é o de mais difícil classificação.

Modelo	classe 0	classe 1	classe 2	classe 3	classe 4
CNN	0.97±0.01	0.92±0.02	0.94±0.01	0.96±0.01	0.95±0.01
MLP (cinza)	0.82±0.08	0.74±0.10	0.77±0.06	0.87±0.04	0.81±0.05
MLP (RGB)	0.94±0.03	0.90±0.06	0.92±0.03	0.95±0.02	0.92±0.03
LR (cinza)	0.21±0.05	0.07±0.06	0.19±0.09	0.53±0.11	0.27±0.11
LR (RGB)	0.41±0.12	0.30±0.07	0.27±0.08	0.61±0.10	0.38±0.08

Os resultados mais promissores foram os obtidos com a CNN, com acurácia, *recall*, *precision* e *F1-score* médios de 95%. Em contraste, o modelo de regressão logística multiclasse alcançou um baixo resultado, com acurácia de 49% (RGB). Note que MLP e LR se beneficiaram significativamente do uso de múltiplos canais de cores, apesar do aumento do número de parâmetros dos modelos. Claramente, observa-se a importância das nuances de cor na caracterização dos graus de Pterígio.

Para efeito de completude, analisou-se também a utilização de redes convolucionais pré-treinadas. Em particular, foi testada uma rede convolucional *VGG16* treinada na base de dados *ImageNet* para obtenção de representações iniciais para nossa CNN. No entanto, não observou-se nenhuma vantagem preditiva para o problema de interesse. Da forma similar, testes preliminares com métodos de segmentação de olhos, não forneceram forte indício de benefício para a performance dos modelos avaliados.

A Tabela 2 exibe os *F1-scores* estratificados por classe, para cada um dos métodos. Como esperado, a CNN apresenta performance superior à dos demais métodos. Adicionalmente, é possível concluir que classe 1 (Grau 1 de Pterígio) é a de mais difícil classificação, em geral.

A Figura 4 apresentada algumas classificações obtidas com as redes convolucionais. Observe que nos casos

reportados, as predições incorretas não diferem em mais de um grau do valor alvo.

Testes e Comparação

Com o intuito de tornar os artefatos de *software* mais acessíveis, este trabalho prevê a disponibilização de um aplicativo móvel para classificação dos graus de Pterígio: *Beye Pterygium*. A Figura 5 ilustra as telas do protótipo em desenvolvimento. Além de telas informativas sobre o que é Pterígio, os seus graus, e postos de saúde mais próximos, o aplicativo conta também com a possibilidade de carregar imagens de olhos para classificação automática usando redes convolucionais. Com isso, usuários podem fotografar seus olhos e obter um diagnóstico preliminar. A ideia é que o aplicativo possa alertar usuários sobre o avanço do grau de Pterígio e recomendá-los a buscar ajuda médica para tratamento e diagnóstico preciso/completo.

CONCLUSÃO

Este artigo apresentou uma nova base de dados juntamente com uma comparação de métodos para classificação dos diversos graus de Pterígio. Dentre os métodos avaliados, redes neurais convolucionais obtiveram os melhores resultados, atingindo em torno

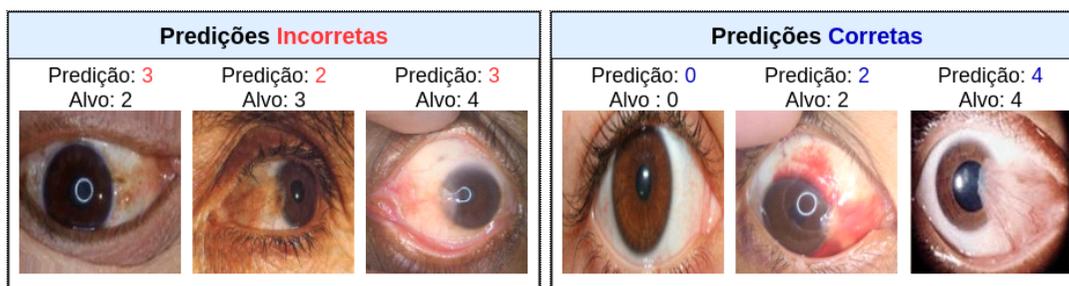


Figura 4 – Exemplos de classificações obtidas usando CNNs.

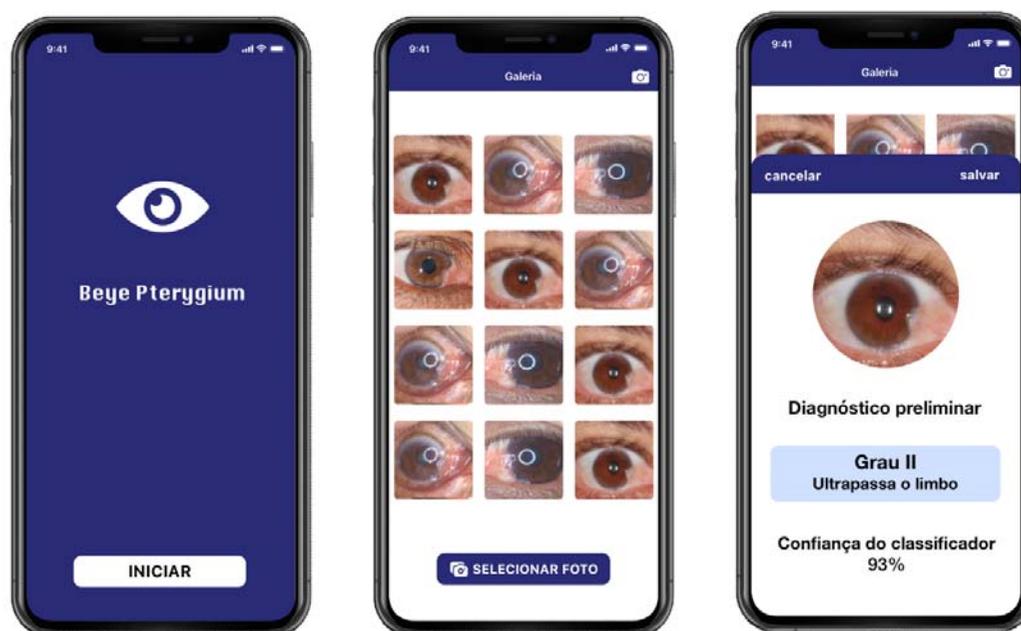


Figura 5 – *Beye Pterygium*: Telas do protótipo de aplicativo móvel para classificação de graus de Pterígio. Dentre outras funcionalidades, o aplicativo permite submeter imagens para classificação automática usando CNNs. Junto com a estimativa de grau de Pterígio, o aplicativo também fornece uma medida de confiança associada à classificação.

de 95% de acurácia.

Adicionalmente, será disponibilizado um aplicativo móvel para o diagnóstico automático do Pterígio, de maneira a fornecer auxílio a especialistas no processo de triagem e diagnóstico desta patologia. Espera-se também que a disponibilização do banco de dados vá fomentar a pesquisa em novas metodologias computacionais para a classificação de Pterígio.

Dentre potenciais trabalhos futuros, destaca-se o uso

de funções custo que levem em conta a natureza ordinal dos graus de Pterígio, ao invés de variáveis categóricas (classes) que ignoram o relacionamento entre os graus. Além disso, planeja-se uma avaliação mais ampla da performance de modelos de ML e estratégias de pré-processamento. Por fim, tendo em foco a utilização desses modelos em aparelhos móveis, é de extrema relevância analisar a robustez dos modelos à utilização de imagens obtidas sob as mais diversas situações de luz e qualidade de captura.

REFERÊNCIAS

1. Noh NRM, Yong G-Y, Lee, S-T, Low KF, and Pan S-W. Patient satisfaction of pterygium surgery using fibrin glue versus suture autograft. *Asian Journal of Ophthalmology*. 2019;16(3):186–192.
2. Baxter SL, Nguyen BJ, Kinori M, Kikkawa DO, Robbins SL, and Granet DB. Identification and correction of restrictive strabismus after pterygium excision surgery. *American journal of ophthalmology*. 2019;202:6–14.
3. Coutts SJ and Coombes A. Pterygium: Prevalence and severity in an amazonian ophthalmic setting. *Revista Brasileira de Oftalmologia*. 2012;71(6):372–376.
4. Zulkifley MA, Abdani SR, and Zulkifley NH. Pterygium-net: a deep learning approach to pterygium detection and localization. *Multimedia Tools and Applications*. 2019; 78(24):34563–34584.
5. Ramos ARC, dos Santos Neto AC, and Carvalho Filho AO. Shape Analysis 3D to false positive reduction. *Journal of Health Informatics*. 2017; 9(4)Szolovits, Peter, ed. *Artificial intelligence in medicine*. 1a. ed. New York: Routledge; 2019. 52 p.
6. Szolovits, Peter, ed. *Artificial intelligence in medicine*. 1st ed. New York: Routledge; 2019. 52 p.
7. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, and Rubin DB. *Bayesian Data Analysis*. 3rd Edition. Chapman and Hall/CRC; 2013.
8. Goodfellow I, Bengio Y, and Courville A. *Deep Learning*. 1st edition. MIT Press; 2016.
9. Abdani SR, Zulkifley MA, and Hussain A. Compact convolutional neural networks for pterygium classification using transfer learning. In: *IEEE International Conference on Signal and Image Processing Applications*. 2019.
10. Pagnin AF. Identificação e caracterização de pterígio utilizando floresta de caminhos ótimos e técnicas de otimização [dissertação]. Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2011.
11. Zaki WMDW, Daud MM, Abdani SR, Hussain A, and Mutalib HA. Automated pterygium detection method of anterior segment photographed images. *Computer methods and programs in biomedicine*. 2018; 154:71–78.
12. Wong SC, Gatt A, Stamatescu V, and McDonnell MD. Understanding Data Augmentation for Classification: When to Warp?. In *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Gold Coast, QLD, 2016, p. 1-6.
13. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*. 2nd edition. MIT Press; 2003.
14. Kingma DP and Ba J. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference for Learning Representations*, San Diego, 2015.
15. HaoChen J, and Sra S. Random Shuffling Beats SGD after Finite Epochs. In *Proceedings of the International Conference on Machine Learning*, 2624–2633. PMLR, 2019.
16. Bland M. *An introduction to medical statistics*. 3th ed. New York: Oxford University Press; 2000.