

Classificação de Patologias Oculares com o Uso em Cascata do Modelo Swin Transformer

Filipe Correia Belfort¹, Antônio Pedro Vieira Lima¹, Italo Francyles Santos da Silva¹
Aristófanés Corrêa Silva¹

¹Núcleo de Computação Aplicada – Universidade Federal do Maranhão (UFMA)
– São Luís - MA - Brazil

{cpbelfort, antonio.lima, francyles, ari, }@nca.ufma.br

Abstract. *Each year, between 1 and 2 million people lose their vision around the world, with women being the most affected. The fundus examination is a non-invasive and low-cost method compared to other techniques, which can be used to diagnose different ocular abnormalities. This observation manual is time consuming and subject to misinterpretation among experts. In this context, this work proposes a cascade classification method to carry out the correct diagnosis of pathologies using the Swin Transformer network. In the test carried out with the Odir5K base, results of up to 71% F1-Score for the normal class were obtained, 0.67% for Sick class.*

Resumo. *A cada ano, entre 1 e 2 milhões de pessoas perdem a visão no mundo, sendo as mulheres as mais afetadas. O exame de fundo de olho é um meio não invasivo e de baixo custo comparado com outras técnicas, que pode ser usado para diagnosticar diferentes anormalidades oculares. Essa observação manual consome tempo e está sujeita a erros de interpretações entre os especialistas. Nesse contexto este trabalho propõe um método de classificação em cascata para realizar o correto diagnóstico das doenças utilizando a rede Swin Transformer, No teste feito com a base Odir5K obteve resultados de até 71% de F1-Score para classe normal, 0.67% para classe Doente.*

1. Introdução

Conforme a Organização Mundial da Saúde, pelo menos um bilhão de pessoas sofre de uma deficiência visual que poderia ter sido prevenida ou ainda não foi abordada [Emir and Colak 2023]. Anualmente, entre 1 e 2 milhões de pessoas perdem a visão no mundo, sendo as mulheres as mais afetadas. No caso das crianças com menos de 15 anos, cerca de 1,4 milhões delas são cegas [Rocha et al. 2014].

A miopia patológica é uma das principais causas de cegueira em países desenvolvidos, afetando aproximadamente 2% da população geral [Carvalho et al. 2014]. A catarata é a principal causa reversível de cegueira no mundo. Essa condição muitas vezes é adquirida ao longo do processo de envelhecimento [Domingues et al. 2016]. As retinopatias diabéticas e hipertensiva acometem grande parte da população mundial. Como afetam a retina, levam a alterações vasculares e conseqüente comprometimento da visão [Jacomini and Hannouche 2001, Hirakawa et al. 2019]. A degeneração macular relacionada à idade (DMRI) representa uma das principais causas de deficiência visual irreversível em muitos indivíduos com mais de 65 anos no mundo ocidental [Rickman et al. 2013]. O

glaucoma destaca-se como a principal causa de cegueira e deficiência visual irreversível no Brasil e no mundo. Estudos projetam que globalmente, em 2030, 95,4 milhões de indivíduos sejam portadores de glaucoma [Guedes 2021].

O exame de fundo de olho é uma técnica não invasiva e de baixo custo para diagnosticar diversas anormalidades oculares. No entanto, a observação manual é demorada e sujeita a erros de interpretação entre especialistas, o que pode levar a diferentes diagnósticos para a mesma imagem. [Araújo et al. 2018]. Atualmente, muitos trabalhos na área da tecnologia buscam produzir ferramentas que auxiliem no diagnóstico de doenças oculares, como o desenvolvido por [Berni 2021], que propuseram a utilização de redes neurais para auxiliar no diagnóstico de catarata, a partir de imagens de exames oftalmológicos realizados no biomicroscópio ocular e no retinógrafo. [Wang et al. 2023], apresentam a rede *MBSaNet* em uma abordagem multiclasse para a detecção de diversas doenças apresentadas pela base de dados *Ocular Disease Intelligent Recognition* (ODIR).

Neste cenário, este estudo propõe um método em cascata para a classificação das doenças oculares Retinopatia Diabética, Glaucoma, Catarata, DMRI, Retinopatia Hipertensiva e Miopia, utilizando a rede *Swin Transformer* proposta por [Liu et al. 2021] em conjunto com uma etapa de pré-processamento nas imagens. O trabalho também usa a base ODIR. Além dos resultados do experimento, será realizada uma análise comparativa entre o método proposto e os resultados obtidos por meio de experimentos com a rede DenseNet-121 proposta por [Huang et al. 2017].

O trabalho está organizado assim: seção 2 apresenta a base de dados e as etapas do método proposto; seção 3 mostra os resultados obtidos; e a seção 4 expõe a conclusão e trabalhos futuros.

2. Materias e métodos

Esta seção descreve de forma mais detalhada o método proposto para a classificação das imagens de fundo de olho patológico, além de maiores informações sobre a base de imagens do estudo.

2.1. Base de imagens

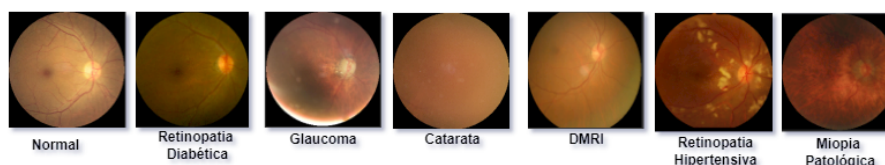
O conjunto de dados *Ocular Disease Intelligent Recognition* (ODIR)¹ consiste em informações oftalmológicas de 5.000 pacientes, incluindo idade, imagens coloridas do fundo de olho esquerdo e direito, e diagnósticos por palavras-chave. As imagens foram capturadas por diversas câmeras disponíveis no mercado, resultando em diferentes resoluções. Os pacientes foram classificados em oito categorias: Normal, Diabetes, Glaucoma, Catarata, Degeneração Macular Relacionada à Idade, Hipertensão, Miopia Patológica e Outras doenças/anormalidades, por leitores humanos treinados com controle de qualidade rigoroso. A Figura 1 fornece exemplos das imagens que compõem o dataset. Neste experimento a classe Outras Doenças/anormalidades foi descartada, devido a complexidade apresentada pelos próprios dados, onde dentro desta classe possuem outras subclasses, requerindo para este problema uma estratégia mais sofisticada.

2.2. Método proposto

Como retratado na Figura 2, após a etapa de pré-processamento, na fase de classificação, todo o método em cascata é baseado na rede proposta por [Liu et al. 2021]. A *Swin Trans-*

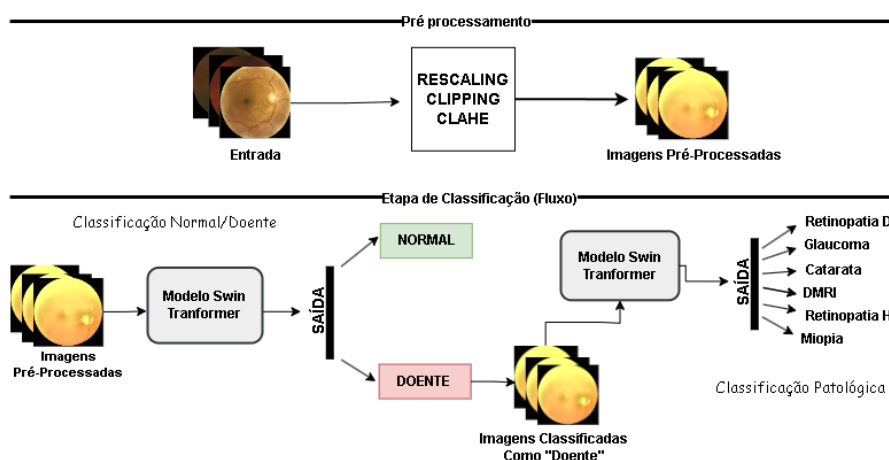
¹Disponível em: <https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k>

Figura 1. Exemplo de imagens de fundo de olho, com as patologias utilizadas no estudo



former (SwinT) propõe um novo paradigma para o processamento de imagens através da aplicação de um Transformer hierárquico com janelas deslocadas, que limitam o cálculo da aplicação do mecanismo de atenção nas regiões locais na imagem, reduzindo significativamente a complexidade computacional do modelo. O SwinT é um modelo de backbone de uso geral que pode ser aplicado a diversas tarefas de visão computacional, como classificação de imagem, detecção de objeto e segmentação semântica. Seus autores mostram que ela alcança resultados equiparados ou superiores aos de outras arquiteturas baseadas em transformer. Para a elaboração do método proposto, foi utilizada a mesma arquitetura SwinT concebida por [Liu et al. 2021].

Figura 2. Fluxograma do método proposto, a constituição de cada fase será maior aprofundada ao longo da sessão



2.2.1. Pré-processamento

Nesta etapa, é realizado um pré-processamento em todas as imagens. Inicialmente, é realizada uma operação de clipping para padronizar e centralizar as imagens de fundo de olho em retângulos, de modo que a imagem não fique distorcida na operação realizada a seguir. Em seguida, é aplicado um redimensionamento para 224x224. Por fim, é aplicado o algoritmo *CLAHE* (*Contrast Limited Adaptive Histogram Equalization*) utilizando a biblioteca OpenCV. Neste trabalho o *CLAHE* é aplicado no canal de cor Y, que corresponde à luminosidade da imagem, após a transformação da imagem do espaço RGB para YUV, e, em seguida, retorna-se para o espaço RGB. Essa aplicação visa realçar a visibilidade da imagem através da distribuição da intensidade dos pixels ao longo dela e ajuda a preservar os detalhes locais na imagem.

2.2.2. Classificação Normal/Doente e Classificação Patológica

Esta etapa visa categorizar as imagens entre as classes normal e doente. Neste caso, uma imagem de olho saudável já terá o resultado de sua classificação. Por outro lado, caso possua doença, ela será submetida a um modelo posterior para diagnosticar qual sua doença específica entre as 6 classes descritas anteriormente. Esta fase pretende realizar uma melhor predição da classe normal, tendo em vista que, após alguns testes, constatou-se que boa parte das imagens das classes de doença eram preditas como normais. Os resultados dos testes serão melhor descritos na subseção Resultados Preliminares.

Quanto à etapa de classificação patológica, será determinada a doença específica das imagens entre as 6 possíveis: DR, Catarata, Glaucoma, DMRI, RH, MP. Tal separação também visa tratar do problema em que muitas classes como retinopatia diabética, DMRI e RH eram erroneamente classificadas como normais, devido principalmente à falta de padronização na captura das imagens, gerando imagens da mesma classe com diferentes aspectos de cor e iluminação, o que afeta ainda mais a classe normal, devido a sua maior quantidade de amostras. Neste contexto, o método visa construir um modelo que pode oferecer uma classificação patológica mais precisa devido à rede poder dar um enfoque nas características existentes entre as próprias classes de imagens doentes.

3. Resultados e Discussão

Para os experimentos da etapa de classificação Normal/Doente, as imagens foram submetidas a várias transformações para aumento de dados, incluindo rotação aleatória de até 90 graus, recorte central, transposição e adição de brilho e contraste aleatórios. Isso foi crucial, especialmente devido à disparidade de amostras entre as classes. A base foi dividida em 1689 imagens de treino, 2016 de teste e 404 de validação respectivamente.

Para a etapa de classificação patológica, todas as outras classes sofreram as mesmas operações de aumento de dados supracitadas. Houve uma ampliação de dez vezes do total de imagens, exceto para a classe Retinopatia Diabética, que já possuía uma quantidade relevante de exemplos. Como resultado, obtiveram-se 9463 imagens para treino, 2369 para validação e 300 imagens de teste na classificação patológica. Ressalta-se que o conjunto de teste foi criado de modo balanceado com 50 imagens de cada classe, selecionadas aleatoriamente, para permitir uma melhor comparação com o trabalho de [Wang et al. 2023] que utilizou da mesma configuração.

As redes *Swin Transformer* (SwinT) foram treinadas por 50 épocas com batch size de 32, utilizando AdamW como otimizador e CrossEntropyLoss como função de perda. A rede Densenet-121 utilizada para comparação com o método proposto foi treinada pelo mesmo número de épocas utilizando o otimizador Adam, e a SparseCategoricalCrossentropy como função de perda. SwinT e *DenseNet-121* foram desenvolvidas usando, respectivamente, as bibliotecas PyTorch² e TensorFlow³. O F1-Score, métrica escolhida, é a média harmônica entre a precisão e a sensibilidade, ideal para avaliação do modelo.

Na etapa de classificação normal/doente, o modelo obteve uma média de *F1-Score* de até 0,71 para a classe Normal e 0,67 para a classe Doente, indicando a ocorrência considerável de falsos positivos para a classe normal e falsos negativos para a classe doente.

²Pytorch versão 2.2

³Tensorflow versão 2.15.1

Contudo, o método em cascata se mostrou promissor para a classificação da classe Normal em comparação com o uso de uma só SwinT para categorizar todas as classes da base. Além disso, o valor obtido para a classe Normal é superior às demais abordagens testadas, como mostrado na Tabela 1.

Tabela 1. Comparação dos resultado dos modelos na base ODIR5k

Métodos	F1-Score						
	N	RD	G	C	DMRI	RH	MP
Método proposto	0.71	0.55	0.83	0.95	0.71	0.45	0.94
SwinT (Classificação Direta)	0.37	0.47	0.58	0.93	0.60	0.30	0.90
DenseNet-121 (Classificação Direta)	0.31	0.26	0.28	0.83	0.29	0.40	0.86
Wang et. al 2023	0.56	0.70	0.84	0.95	0.89	0.90	0.99

No entanto, no fluxo completo do método, o modelo enviará imagens normais erroneamente para terem sua doença classificada, acarretando uma alta margem de erro. Após o fluxo completo, o modelo da etapa de classificação patológica obteve apenas 0,34 de acurácia média entre as 6 classes. Dado esse cenário, foi realizado um experimento usando o conjunto de teste balanceado, simulando uma classificação totalmente correta para verificar o desempenho da rede. Os resultados de *F1-Score* também são vistos na Tabela 1. Os principais erros observados se devem principalmente à semelhança entre aspectos das imagens entre as classes, como manchas, hemorragias e vasos sanguíneos alterados, características que podem confundir até mesmo os próprios especialistas, resultando em um problema de classificação mais complexo.

Comparando com trabalho de [Wang et al. 2023], que utilizou a *MBSanet* e técnicas mais sofisticadas de classificação, como fusão de características de múltiplas escalas, vê-se que o método proposto se equipara ou até mesmo se aproxima em relação as classes Glaucoma (G), Catarata (C) e Miopia (MP), superando seus resultados no caso da classe normal.

4. Conclusão e Trabalhos Futuros

Os resultados preliminares da etapa de classificação Normal/Doente são promissores, com margem para melhorias. Levando em consideração que, para um resultado confiável da etapa de classificação das doenças, esta etapa deve oferecer uma maior precisão para a divisão Normal/Doente. Os testes realizados na classificação patológica obtiveram resultados próximos ou equivalentes ao estado da arte para classes como a classe Normal, Glaucoma e Miopia, enquanto para as outras classes o F1-score se manteve abaixo. No entanto, levando em consideração a complexidade da classificação delas, devido às semelhanças entre as doenças e à falta de padronização da captura das imagens do dataset, os resultados são positivos. Sugere-se que futuras melhorias incluam o pré-processamento para normalização de cores, ajustes nos hiperparâmetros da rede e a aplicação de pesos nas classes para lidar com o desbalanceamento, estimando melhorias no método proposto.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 , Fundação de

Amparo a Pesquisa do Maranhão (FAPEMA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Empresa Brasileira de Serviços Hospitalares (Ebsers) Brazil (Proc. 409593/2021-4)

Referências

- Araujo, J. D. L. et al. (2018). Diagnóstico de glaucoma a partir de imagens de fundo de olho utilizando índices de diversidade.
- Berni, A. V. (2021). Uso de inteligência artificial para apoio na identificação de catarata.
- Carvalho, B., Freitas-Costa, P., Pinheiro-Costa, J., Falcão, M., Carneiro, Â., and Falcão-Reis, F. (2014). Avaliação dos resultados do tratamento antiangiogênico na neovascularização coroideia associada à miopia patológica. *Acta Médica Portuguesa*, 27(1):49–58.
- Domingues, V. O., Lawall, A. R. N., Battestin, B. B., de Lima, F. J. R., Lima, P. M., Ferreira, S. H., and Moraes, C. F. (2016). Catarata senil: uma revisão de literatura. *Revista de medicina e saúde de Brasília*, 5(1).
- Emir, B. and Colak, E. (2023). Análise de desempenho de modelos de rede neural convolucional pré-treinados para classificação de doenças oftalmológicas. *Arquivos Brasileiros de Oftalmologia*, 87:e2022–0124.
- Guedes, R. A. P. (2021). Glaucoma, saúde coletiva e impacto social.
- Hirakawa, T. H., Costa, W. d. C., Nakahima, F., Ferreira, A. I. C., Ribeiro, L. B., Ticianeli, J. G., and Sequeira, B. J. (2019). Conhecimento dos pacientes diabéticos usuários do sistema único de saúde acerca da retinopatia diabética. *Revista Brasileira de Oftalmologia*, 78:107–111.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Jacomini, C. Z. and Hannouche, R. Z. (2001). Retinopatia hipertensiva. *Rev bras hipertens*, 8:321–7.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022.
- Rickman, C. B., Farsiu, S., Toth, C. A., and Klingeborn, M. (2013). Dry age-related macular degeneration: mechanisms, therapeutic targets, and imaging. *Investigative ophthalmology & visual science*, 54(14):ORSF68–ORSF80.
- Rocha, M. N. A. M., Ávila, M. P. d., Isaac, D. L. C., Mendonça, L. S. d. M., Nakanishi, L., and Auad, L. J. (2014). Prevalência de doenças oculares e causas de comprometimento visual em crianças atendidas em um centro de referência em oftalmologia do centro-oeste do Brasil. *Revista Brasileira de Oftalmologia*, 73(4):225–229.
- Wang, K., Xu, C., Li, G., Zhang, Y., Zheng, Y., and Sun, C. (2023). Combining convolutional neural networks and self-attention for fundus diseases identification. *Scientific Reports*, 13(1):76.