

Impacto da Resolução na Detecção de Retinopatia Diabética com uso de Deep Learning*

Francis B. Moreira¹, Beatriz D. Schaan², Josiane Schneiders²,
Mateus A. dos Reis², Matheus S. Serpa¹, Philippe O. A. Navaux¹

¹ Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

{fbmoreira, msserpa, navaux}@inf.ufrgs.br

²Hospital de Clínicas de Porto Alegre
Porto Alegre – RS – Brasil

{bschaan, jschneiders, mreis}@hcpa.edu.br

Abstract. *In the context of healthcare systems, preventive screening is one of the most effective ways to prevent disease progression. Most diseases can be treated when detected in their initial stages. The demand for preventive screening is increasing, and this demand cannot be efficiently covered by the overloaded medical doctors. Therefore, there is a current need for a methodology to automate and increase the efficiency of preventive screening. In this paper, we discuss the development of deep convolutional neural networks to detect diabetic retinopathy, and the impact of image and network resolution on prediction accuracy. We have achieved 0.93 area under the receiver operating characteristic curve by increasing the analysis resolution of the inception v3 architecture.*

Resumo. *No contexto dos sistemas de saúde, a prevenção é uma das maneiras mais eficazes evitar a progressão de doenças. Muitas delas podem ser tratadas quando diagnosticadas em estágio inicial. A demanda por exames preventivos está aumentando e não consegue-se atender essa demanda com eficiência pelos médicos sobrecarregados. Portanto, existe a necessidade de uma metodologia para automatizar e aumentar a eficiência de exames de triagem. Neste artigo, discutimos o desenvolvimento de redes neurais convolucionais profundas para detectar retinopatia diabética, e o impacto da resolução de imagem e da rede na precisão de predição. Atingimos 0,93 de área sob a curva de característica operacional do receptor ao aumentar a resolução da arquitetura inception v3.*

1. Introdução

O diabetes mellitus é uma doença crônica associada à hiperglicemia devido a falha na regulação da produção de insulina, com prevalência crescente, e afetando em torno de 12% da população brasileira [Telo et al. 2016]. Os níveis aumentados de glicose nos vasos sanguíneos da retina podem causar alterações a longo prazo denominadas de retinopatia diabética (RD), as quais, caso não detectadas precocemente, podem evoluir para

*Este trabalho foi parcialmente financiado pelo projetos Petrobras 2016/00133-9, 2018/00263-5 e pelo projeto "GREEN-CLOUD: Computação em Cloud com Computação Sustentável" (#16/2551-0000 488-9), da FAPERGS e do CNPq, programa PRONEX 12/2014.

perda progressiva da visão, até cegueira. Como método de prevenção, realiza-se exame de fundo de olho por médicos oftalmologistas a fim de identificar alterações precoces na retina [Stratton et al. 2000]. No entanto, à medida que a população cresce e aumenta a prevalência do diabetes, torna-se difícil o sistema de saúde absover essa demanda. Além disso, o número de oftalmologistas disponíveis é abaixo do necessário, pois sua formação é longa e demanda muitos recursos [Resnikoff et al. 2012]. Realizar retinografias do fundo de olho por profissional técnico não médico é método de rastreamento aceito, porém, as retinografias obtidas precisam ser laudadas por oftalmologistas.

Nesse contexto, o aprendizado de máquina oferece uma excelente oportunidade para realizar a detecção de RD em larga escala dentro de um orçamento limitado. A retinografia seria analisada primeiro por uma rede neural convolucional, que serve como triagem, fazendo com que somente aqueles pacientes que apresentem alguma alteração no exame sejam encaminhados para oftalmologista para avaliação e tratamento, assim diminuindo a demanda do serviço especializado [Gulshan et al. 2016].

Este trabalho foi realizado em colaboração com o Hospital das Clínicas de Porto Alegre (HCPA), como parte do projeto *Diabetic retinopathy screening in patients with diabetes mellitus: validation of innovative method (machine learning)*. Neste artigo, mostramos os resultados obtidos com redes neurais profundas e convolucionais usando a arquitetura *inception v3*. Demonstramos que uma resolução mais alta aumenta a precisão média da rede neural. Atingimos mais de 0,93 de área abaixo da curva na característica de operação do receptor do nosso melhor modelo.

Na Seção 2, discutimos os trabalhos relacionados que abrangem as técnicas utilizadas e sua aplicação à RD. Na Seção 3, mostramos os resultados obtidos com nossos modelos, detalhando a metodologia e os efeitos observados com as alterações nas imagens. Na Seção 4, apresentamos as conclusões e discutimos nossas direções futuras.

2. Trabalhos Relacionados

O aprendizado de máquina passou por uma revolução na última década. O aumento do poder computacional permitiu que os pesquisadores de inteligência artificial criassem modelos muito mais complexos. Embora tenha sido formalmente comprovado que as redes neurais com uma única camada oculta podem ser aproximadores universais [Hornik et al. 1989], a especialização de múltiplas camadas mostrou que esses "aproximadores" podem se tornar precisos o suficiente para realizar tarefas humanas.

Um caso particularmente conhecido é a rede neural convolucional profunda [LeCun et al. 1995]. Ela simula o comportamento do olho humano, filtrando subáreas do olho através de convoluções. Ou seja, em vez de conectar todas as camadas de neurônios como em uma rede neural regular, ela constrói uma camada combinando setores de neurônios da camada anterior, como pode ser visto na Figura 1. Cada neurônio observa apenas uma parte da imagem, como no olho humano. As convoluções entre diferentes regiões criam a identificação de formas, definindo características que identificarão as classes definidas no problema. Assim, as CNNs (*Convolutional neural networks*) profundas têm camadas que executam automaticamente a engenharia e escolha de características, trabalho geralmente feito pelo usuário. Ao combinar camadas de convolução, *max-pooling*, unidades lineares retificadas e camadas totalmente conectadas, uma rede neural convolucional identifica com precisão objetos em uma imagem

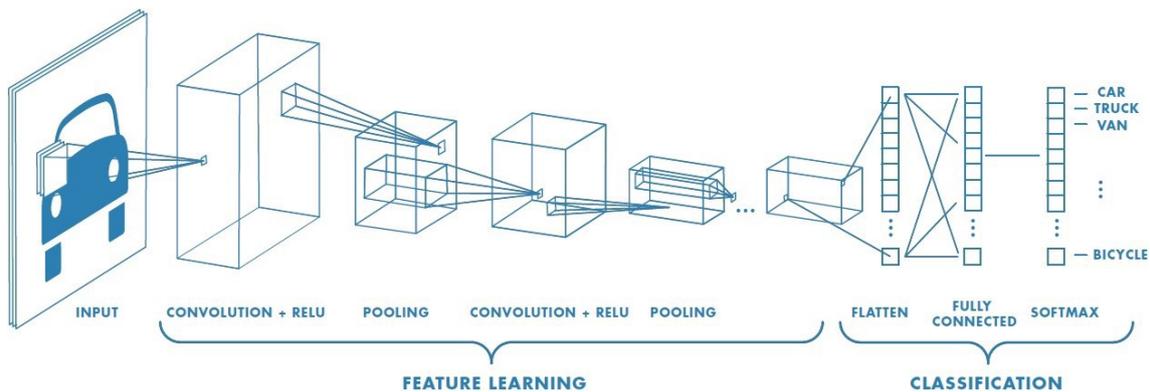


Figura 1. Abstração de rede neural convolucional. Fonte: <https://www.mathworks.com/>

Uma das arquiteturas mais populares e atualmente usadas da rede neural é a arquitetura Inception [Szegedy et al. 2015]. A ideia por trás dela é usar recursos de vários níveis de resolução de maneira eficiente, usando uma rede neural muito profunda e explorando as unidades de processamento gráfico de uso geral (GPGPU) para treinar eficientemente a rede neural.

Essas ideias foram recentemente usadas para detectar RD em retinografias. No trabalho de Gulshan et al. [Gulshan et al. 2016], é usada uma rede neural convolucional baseada na arquitetura Inception, versão 3. Os autores usaram dados de três hospitais indianos (Hospitais de Olhos Aravind, Sankara Nethralaya e Narayana Nethralay) e o conjunto de dados EyePACS [Cuadros and Bresnick 2009]. Para a validação, eles usaram imagens adicionais do EyePACS e o conjunto de dados Messidor-2 [Decencièrre et al. 2014].

Um passo importante em seu procedimento foi a reavaliação das imagens por médicos oftalmologistas experientes. O modelo obtido alcançou uma área sob a curva (AUC) da característica de operação do receptor (ROC) de 0,991 para EyePACS e 0,990 para Messidor-2. Assim, ao selecionar um ponto de alta sensibilidade, o modelo atinge 97,5% de sensibilidade e 93,4% de especificidade para o EyePACS; e atinge 96,1% de sensibilidade e 93,9% de especificidade para o Messidor-2. Esses resultados são muito superiores a outros trabalhos com o mesmo objetivo [Pratt et al. 2016, Tufail et al. 2017, Wong and Bressler 2016, Gargeya and Leng 2017]. Um estudo recente de Voets et al. [Voets et al. 2018] tentou reproduzir esses resultados com o conjunto de dados EyePACS disponível na competição Kaggle para retinopatia diabética ¹, mas não foi capaz de fazê-lo. Após uma mudança significativa na função de ativação da camada final, o algoritmo de reprodução conseguiu atingir 0,94 AUC para EyePACS e 0,80 AUC para Messidor. Os autores da reprodução apontam quatro possíveis razões para essa diferença. Primeiro, na reprodução as retinografias do banco não foram reavaliadas por oftalmologistas. Segundo, a lista publicada de hiperparâmetros do estudo original não detalha o método de normalização e o procedimento de validação usado, de modo que o algoritmo de reprodução não é capaz de atingir o mesmo nível de ajuste no modelo do que o original. Terceiro, é possível que haja erros no estudo ou metodologia originais. Quarto, o estudo da reprodução apresenta erros em sua metodologia. O estudo demonstra a dificuldade de reproduzir um modelo sem o mesmo conjunto de dados ou código disponível. Os mo-

¹ <https://www.kaggle.com/c/diabetic-retinopathy-detection/data>

delos em uso são muito complexos e tal falta de informação tornará a reprodução de um modelo incompleta.

O estudo original de Gulshan foi estendido por Krause et al. [Krause et al. 2018]. O foco deste estudo foi usar a adjudicação para quantificar erros na classificação da retinopatia diabética com base em graduadores individuais e decisão majoritária. Eles usam dados aprimorados para treinar uma rede neural melhor, usando a arquitetura Inception v4 e aumentando a resolução das imagens para treinamento. O trabalho constatou que, dentre as discrepâncias entre adjudicação por especialistas em retina e decisão majoritária de oftalmologistas, as falhas mais comuns são em geral, os microaneurismas ausentes (36%), artefatos (20%) e hemorragias classificadas incorretamente (16%). Eles então comparam o desempenho da decisão da maioria do oftalmologista (OMD) e o algoritmo de aprendizado profundo (DLA), usando o consenso adjudicado de especialistas em retina como padrão de referência. Para retinopatia diabética moderada ou pior, o OMD obtém 83,8% de sensibilidade e 98,1% de especificidade, enquanto o DLA obtém 97,1% de sensibilidade, 92,3% de especificidade e AUC de 0,986. Assim, eles descobriram que, usando um pequeno número de notas de consenso adjudicadas de especialistas em retina como um conjunto de dados de ajuste e imagens de alta resolução como entrada, o modelo melhorou na AUC de 0,934 para 0,986 para detecção moderada ou pior de retinopatia diabética.

Neste artigo, baseamos nosso código na reprodução de Voets et al. [Voets et al. 2018], mas adaptamos ideias do trabalho mais recente de Krause et al. [Krause et al. 2018], como resolução mais alta para imagens de entrada.

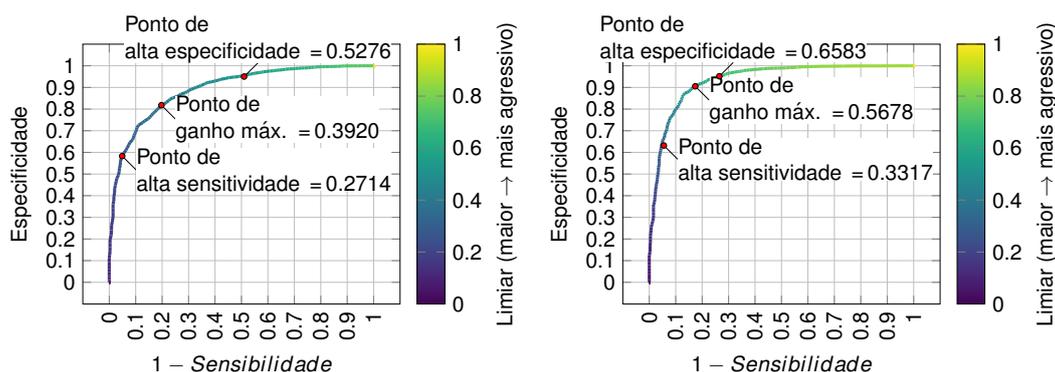
3. Experimentos

3.1. Desenvolvimento do Algoritmo

Para acesso inicial à viabilidade do uso do aprendizado de máquina como uma ferramenta de triagem, treinamos dez redes neurais com base na arquitetura Inception v3. Usamos os mesmos parâmetros disponíveis no trabalho de Gulshan et al. [Gulshan et al. 2016] e sua reprodução [Voets et al. 2018]. Nossa base de código inicial é a base de código da reprodução, disponível em <https://github.com/mikevoets/jama16-retina-replication>. Todas as redes foram inicializadas com pesos do conjunto de dados ImageNet para todos os camadas, exceto a camada totalmente conectada na parte superior, que recebeu treinamento. Como entrada, usamos o conjunto de dados EyePACS da Kaggle, redimensionando cada imagem para 299x299 pixels. Usamos as mesmas técnicas de aumento de dados para aumentar artificialmente o conjunto de dados: invertemos as imagens pelo eixo vertical (i.e. troca horizontal da esquerda para a direita), pelo eixo horizontal (troca de cima e baixo), saturação aleatória, matiz aleatória, brilho aleatório, e contraste aleatório.

3.2. Curvas das características operacionais dos receptores dos conjuntos

A curva característica operacional do receptor do modelo de conjunto composto pelas dez redes neurais é ilustrada na Figura 2(a). Os três pontos escolhidos indicam um alto ponto de especificidade, um ponto máximo de ganho e um ponto alto de sensibilidade. O ponto de alta especificidade atinge 95,02% de especificidade e 51,05% de sensibilidade. O ponto de ganho máximo atinge 81,78% de especificidade, 80,26% de sensibilidade. O ponto de alta sensibilidade atinge 58,30% de especificidade, 95,10% de sensibilidade.



(a) Resolução de imagem de entrada de 300×300 pixels

(b) Resolução de imagem de entrada de 500×500 pixels

Figura 2. Características operacionais do receptor do modelo de conjunto composto por dez redes neurais

Para aprimorar esse modelo, discutimos com os médicos do HCPA sobre a doença. Descobrimos que as lesões e detalhes da doença podem ser representados em segmentos muito pequenos da imagem, o que nos levou a aumentar a resolução da imagem de entrada, assim como o trabalho de Krause et al.

A curva característica operacional do receptor do modelo de conjunto composto pelas novas dez redes neurais é ilustrada na Figura 2(b). Podemos observar uma melhoria na troca de um ponto de alta sensibilidade, que agora atinge mais de 60% de precisão.

O ponto de especificidade atinge 95,17% de especificidade e 73,49% de sensibilidade. O ponto de ganho máximo atinge 90,53% de especificidade, 82,85% de sensibilidade. O ponto de sensibilidade atinge 64,46% de especificidade, 95,1% de sensibilidade.

4. Conclusões e Trabalhos Futuros

No contexto da saúde, a sensibilidade em diagnosticar determinada doença é a métrica mais importante de um mecanismo de triagem. A detecção de doenças inexistentes não é um problema quando comparado a uma doença não detectada que requer tratamento.

No entanto, se o objetivo de um modelo preditivo é ser eficiente na redução da carga de oftalmologistas em saúde pública, a precisão do modelo representa quanto ele pode economizar. Neste artigo, mostramos como uma alteração simples pode afetar o desempenho de uma CNN.

Dada a mesma sensibilidade, imagens de alta resolução proporcionaram uma melhoria de 6% na precisão das detecções. Para a curva de característica de operação do receptor, a área sob a curva mudou de 0,89 para 0,93.

Atualmente, estamos trabalhando para melhorar o modelo. Nossa ideia atual é usar o Inception v4 e obter um conjunto de dados especializado para ajustar o modelo, como foi feito por Krause et al. Pesquisas sobre hiperparâmetros também devem ser consideradas, dada a área muito maior sob a curva obtida em suas pesquisas. Finalmente, a maior fonte de diferença é o conjunto de dados disponível publicamente. Temos como objetivo juntar os dados de pacientes do HCPA aos dados já existentes para criar um conjunto maior de treinamento com configurações variadas de câmera, o que deve ajudar a rede a generalizar melhor os artefatos de diferentes câmeras.

Referências

- [Cuadros and Bresnick 2009] Cuadros, J. and Bresnick, G. (2009). Eyepacs: an adaptable telemedicine system for diabetic retinopathy screening. *Journal of diabetes science and technology*, 3(3):509–516.
- [Decencière et al. 2014] Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., et al. (2014). Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234.
- [Gargeya and Leng 2017] Gargeya, R. and Leng, T. (2017). Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7):962–969.
- [Gulshan et al. 2016] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410.
- [Hornik et al. 1989] Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.
- [Krause et al. 2018] Krause, J., Gulshan, V., Rahimy, E., Karth, P., Widner, K., Corrado, G. S., Peng, L., and Webster, D. R. (2018). Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*, 125(8):1264–1272.
- [LeCun et al. 1995] LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995.
- [Pratt et al. 2016] Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., and Zheng, Y. (2016). Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, 90:200–205.
- [Resnikoff et al. 2012] Resnikoff, S., Felch, W., Gauthier, T.-M., and Spivey, B. (2012). The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners. *British Journal of Ophthalmology*, 96(6):783–787.
- [Stratton et al. 2000] Stratton, I. M., Adler, A. I., Neil, H. A. W., Matthews, D. R., Manley, S. E., Cull, C. A., Hadden, D., Turner, R. C., and Holman, R. R. (2000). Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (ukpds 35): prospective observational study. *Bmj*, 321(7258):405–412.
- [Szegedy et al. 2015] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- [Telo et al. 2016] Telo, G. H., Cureau, F. V., de Souza, M. S., Andrade, T. S., Copês, F., and Schaan, B. D. (2016). Prevalence of diabetes in brazil over time: a systematic review with meta-analysis. *Diabetology & metabolic syndrome*, 8(1):65.
- [Tufail et al. 2017] Tufail, A., Rudisill, C., Egan, C., Kapetanakis, V. V., Salas-Vega, S., Owen, C. G., Lee, A., Louw, V., Anderson, J., Liew, G., et al. (2017). Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*, 124(3):343–351.
- [Voets et al. 2018] Voets, M., Møllersen, K., and Bongo, L. A. (2018). Replication study: Development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *arXiv preprint arXiv:1803.04337*.
- [Wong and Bressler 2016] Wong, T. Y. and Bressler, N. M. (2016). Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *Jama*, 316(22):2366–2367.