Interpretando Modelos para Diagnóstico de Retinopatia Diabética

Henrique C. Barbosa[1], **Anísio M. Lacerda**[1]

[1]Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)

henriquecb@dcc.ufmg.br, anisio@dcc.ufmg.br

Resumo. *O uso de Machine Learning no contexto da medicina trás diversas implicações a respeito da veracidade dos resultados dos modelos já que o impacto dos resultados é grande na vida das pessoas. Pensando nisso, este trabalho tem como objetivo aplicar métodos de interpretabilidade em modelos que façam a predição do grau de retinopatia diabética, uma doença ocular causada pela diabétes, em imagens de fundo de olho. Inicialmente, foi escolhido o algoritmo apresentado por [Voets et al. 2019] para ser o baseline do trabalho e o algoritmo foi rodado com o dataset público do Kaggle disponibilizado na competição de detecção de retinopatia do APTOS. Os resultados obtidos foram consideravelmente bons, pois foi obtido uma AUC de 0,96 enquanto o artigo apontou uma AUC de 0,95. Com estes resultados, ficou decido que este modelo será usado como baseline deste trabalho e os próximos passos do mesmo é aplicar métodos de explicabilidade para entender quais as features nas imagens que o modelo leva em conta na hora de realizar as predições.*

1. Introdução

Machine Learning (ML), um termo muito utilizado nos dias de hoje, pode ser definido como uma série de métodos (ou modelos) usados pelo computador para fazer boas predições de dados futuros baseadas em um conjunto prévio de dados do passado. Esses modelos vêm sendo cada vez mais usados em diversas áreas devido á sua rápida evolução e grande abrangência, caminhando pela medicina, linguística e até mesmo economia.

Uma das principais motivações de uso destes modelos é poder detectar possíveis doenças, quando aplicados neste contexto por exemplo, em sua fase inicial de forma que os médicos não conseguiriam. Focando neste ponto, a Retinopatia Diabética (RD), uma doença comum em pacientes diabéticos que causa perda gradual de visão, se torna um bom alvo para essa aplicação já que seu diagnóstico muitas vezes é feito muito tardiamente devido à sua falta de proliferação de sintomas.

Contudo, existe um grande problema presente nas aplicações dos modelos de ML, a falta de explicação dos computadores no momento de informarem os resultados. A maior dificuldade nessa questão é que apenas uma métrica, como a precisão da classificação por exemplo, não é suficiente para englobar todas as variáveis que podem ter levado à esse resultado. E essa lacuna de informações presente entre o *input* e o *output*, faz com que muitos ainda duvidem e fiquem receosos a respeito de utilizar modelos de ML para realizar predições, principalmente quando estão relacionados à temas mais sensíveis e críticos.

Pensando nessa questão, muitas pesquisas estão voltadas para criar modelos melhor explicativos para justificar as decisões tomadas. Entretanto, ainda há uma gama de

modelos que precisam ser avalidados e ter essa mudança de paradigma para que seja possível uma validação mais acertiva do que o levou a gerar tais resultados. Ademais, em muitos casos, é muito mais importante se saber o por quê daquele resultado ter sido predito do que o resultado por si só.

Sendo assim, o objetivo desse trabalho é a reprodução de um modelo de Machine Learning para detecção de retinopatia diabética com base em imagens de exame de fundo de olho, obtendo um modelo baseline para que futuramente possa ser usado como alvo para os métodos de interpretabilidade.

2. Referencial Teórico

Inicialmente, podemos definir interpretabilidade, no contexto de ML, como a capacidade de explicar ou apresentar em termos compreensíveis para um humano (Doshi-Velez & Kim)[Doshi-Velez and Kim 2017] o resultado daquele sistema. Como Christoph Molnar apresenta em seu livro [Molnar 2019], a interpretabilidade de modelos se torna um importante fator ja que para certos problemas ou tarefas, não é suficiente obter a previsão (o quê). O modelo também deve explicar como chegou à previsão (o porquê), porque uma previsão correta resolve apenas parcialmente o seu problema original.

Além deste fator, a RD é outro ponto importante para o trabalho, pois é a doença alvo que o modelo escolhido irá detectar. Schmidt-Erfurth et al [Schmidt-Erfurth et al. 2018] mostra alguns pontos importantes que incentivam o uso de ML para doenças na retina como o fato da RD ter sido considerada uma epidemia mundial além de que um terço de uma estimativa de 285 milhões de pessoas com diabetes tem sinais de RD e um terço deles tem RD com risco de visão.

Para execução deste trabalho, os trabalhos [Gulshan et al. 2016][Schmidt-Erfurth et al. 2018][Voets et al. 2019] foram usados como base de entendimento de modelos para detecção de RD em imagens de fundo de olho. O algorítmo escolhido foi o apresentado em Voets et al [Voets et al. 2019] por ter sido disponibilizado o código criado. Baseado em Gulshan et al [Gulshan et al. 2016], o algoritmo usado é uma rede neural,uma grande função matemática com milhões de parâmetros, que calcula a gravidade da retinopatia diabética a partir das intensidades dos pixels em uma imagem de fundo de olho. Durante o processo de treinamento, os parâmetros da rede neural são inicialmente configurados como valores aleatórios. Em seguida, para cada imagem, o grau de gravidade fornecido pela função é comparado com o grau conhecido, e a medida que a função roda e compara os resultados com o grau conhecido, seus parâmetros vão sendo ajustados para um melhor resultado. Um ponto importante levantado por Almazroa et al [Almazroa et al. 2017] é a necessidade da validação dos resultados por especialistas da área, dando mais segurança nos resultados obtidos.

Em Ribeiro et al [Ribeiro et al. 2016], é mostrado um ponto interessante a respeito de explicar as predições dos modelos de ML usando modelos-agnósticos devido à flexibilidade dada para a escolha do seu modelo. Lopez-Paz et al [Lopez-Paz et al. 2017] traz uma abordagem interessante criando um algoritmo (*Neural Causal Coefficient*) para avaliar causalidade em imagens, mostrando que a existem de sinais observáveis que revelam as disposições causais dos objetos, além de uma relação entre a direção da causalidade e a diferença entre os objetos e seus contextos. Zhang et al [Zhang et al. 2017] apresenta

uma ferramenta que não apenas diagnosticar uma doença, como também gerar relatórios sobre o diagnóstico através de uma combinação de um modelo que lida com imagens de cancêr de bexiga e um modelo que lida com a linguagem para avaliar laudos médicos.

3. METODOLOGIA

Inicialmente, foi necessário um embasamento teórico buscando na literatura a respeito do funcionamento dos modelos de predição e sobre os métodos usados para o interpretabilidade de modelos. Após o domínio dos conceitos, o projeto focou em encontrar um bom modelo de predição que será usado para aplicarmos os métodos de interpretabilidade. O modelo usado foi a reprodução de [Gulshan et al. 2016] por [Voets et al. 2019] foi o dataset público disponibilizado no Kaggle [Kag] pela APTOS (Asia Pacific Tele-Ophthalmology Society). Com isso, os próximos passos focarão na interpretabilidade deste modelo e em uma validação dos resultados por uma equipe de oftalmologistas. A Figura 1 mostra o desenho da metodologia.

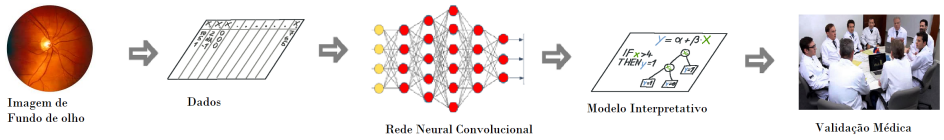

Figura 1. Fluxograma da metodologia proposta.

Dataset

O dataset escolhido foi o da competição *APTOS 2019 Blindness Detection* que ocorreu no Kaggle, uma plataforma para competições de Machine Learning. Esse conjunto conta com 5590 imagens de fundo de olho dividas em dois conjuntos, um de treino (com 3662 imagens) e um de teste (com 1928 imagens),e todas foram classificadas por um médico de acordo com a gravidade da RD em uma escala de 0 a 4, na qual:

Classificação	Legenda
0	Sem RD
1	Suave
2	Moderada
3	Grave
4	RD Proliferativa

Figura 2. Classificações das imagens do dataset

O Kaggle informa que as imagens podem conter ruído, sejam esses por estar fora de foco, subexposição ou superexposição. As imagens foram coletadas de várias clínicas usando uma variedade de câmeras por um longo período de tempo, o que introduzirá mais variações. Na Figura 3, podemos ver dois exemplos de imagens do dataset:

Detecção de Retinopatia Diabética

Voets et al [Voets et al. 2019] apresenta um algoritmo para detecção da Retinopatia Diabética baseado em Gulshan et al [Gulshan et al. 2016]. Nessa etapa, esse algoritmo

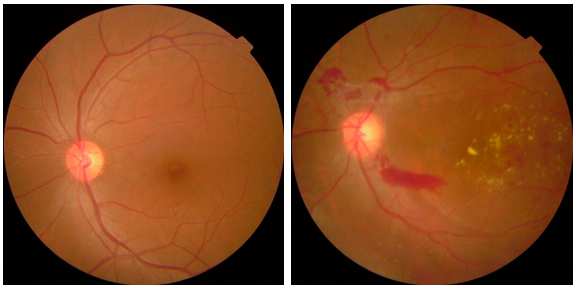

Figura 3. Imagens sem RD (Esquerda) e com RD Proliferativa (Direita)

será reproduzido para avaliar seus resultados e desempenho. Logo após, serão propostas algumas mudanças para possíveis melhorias para atingir uma melhor acurácia dos resultados.

O algoritmo foi criado por meio de deep learning, que é o processo de treinamento de uma rede neural (uma grande função matemática com milhões de parâmetros) para calcular a severidade da retinopatia diabética baseada na intensidade dos pixéis em uma imagem de fundo de olho. Esse modelo envolve um procedimento de treinamento de uma rede neural para executar a tarefa de classificar imagens.

A rede neural usada foi a mesma utilizada no estudo original e na sua primeira reprodução: o modelo InceptionV3 proposto por Szegedy et al [Szegedy et al. 2015] que inicialmente agrupam pixels de intensidade semelhtantes em features locais para depois agrupá-las novamente em features globais. A arquitetura da Inception pode ser vista na Figura 4.

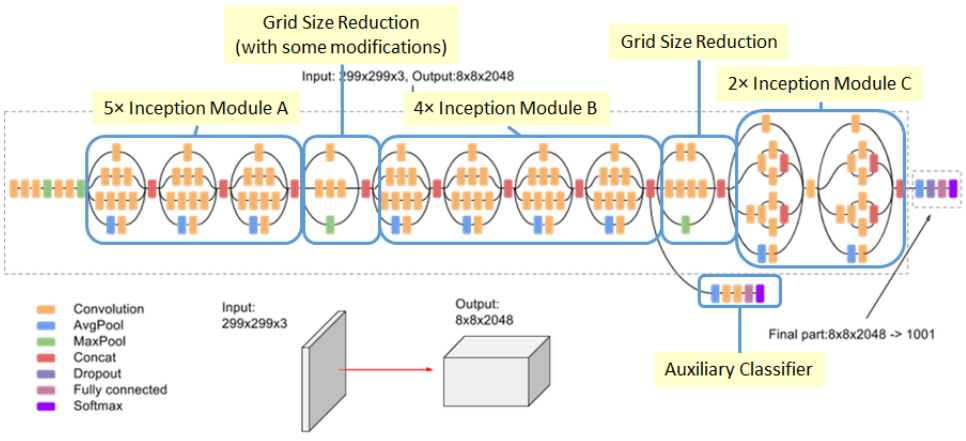

Figura 4. Fonte: Sik-Ho Tsang - Review: Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015 (2018)

A InceptionV3 é uma rede arquiteturada para reduzir a quantidade de parâmetros usadas o que, de acordo com os autores, reduz as chances de ocorrer overfitting e permitem a rede ir mais fundo para o seu aprendizado. A rede foi dividida em 3 módulos que trabalham exatamente para reduzir a quantidade de parâmetros por meio de fatoração das convoluções, pois é mostrado que duas convoluções 3x3 são tão eficientes quanto uma 5x5 e ainda reduz a quantidade de conexões/parâmetros usados.

Em [Voets et al. 2019] não é informado nenhum tipo de tratamento visual nas imagems apenas para redimensioná-las. As imagens, antes de serem dadas como a entrada para a rede, passarão por um pré-processamento com o objetivo de detectar a máscara circular da imagem e redimensionando o seu diâmetro para ter 299 pixels de largura.

4. Resultados

Foi comparado o resultado apresentados nos artigos de [Gulshan et al. 2016] e [Voets et al. 2019] com o modelo do segundo artigo sendo executado com um dataset diferente do que foi usado para sua construção. Os datasets usados possuem tamanhos diferentes e o que [Gulshan et al. 2016] usou foi um conjunto de imagens privado do EYEPacs enquando [Voets et al. 2019] usou o conjunto de imagens público do EYEPacs disponibilizado na competição *Diabetic Retinopathy Detection* do Kaggle.

A métrica usada foi a área abaixo da curva ROC, AUC, e todos os datasets usados possuem a mesma classificação de 0 a 4 das imagens. Os resultados comparativos dos modelos podem ser vistos na tabela a seguir:

Fonte do Algoritmo	Treino	Teste	AUC
Gulshan, 2016	~82032	~25635	0,99
Voets, 2019	57146	8790	0,95
Voets et al, 2019 (POC I)	3662	1928	0,96

Figura 5. Tabela comparativa com os resultados dos modelos

5. Conclusão

Inicialmente, a escolha do modelo foi uma tarefa complexa devido a dificuldade de achar bons preditores para essa aplicação. Alguns trabalhos foram considerados mas a reprodução visitada neste artigos foi um forte candidato como baseline para esse projeto.

Percebe-se pelos resultados apresentados que o modelo proposto reproduzido por [Voets et al. 2019] perfomou bem no dataset apresentado no [Kag] além de ter tido um bom resultado pelo que foi mostrado no artigo. Com isso, este modelo se tornou o baseline do projeto e este será usado como alvo para aplicação dos métodos de interpretabilidade.

Como trabalho futuro, o primeiro passo será rodar o modelo novamente em uma base de dados diferente que será disponibilizada pelo Instituto de Olhos e que será usada como o dataset do projeto. Como o dataset usado neste artigo foi muito pequeno comparado ao dataset originalmente usado, pode ter acontecido overfitting e sendo isso o motivo de um resultado maior que o artigo que reproduziu o modelo.

Com essa validação do modelo com um outro dataset, o passo seguinte será decidir e aplicar os métodos de explicabilidade para tentar verificar quais as features que estão sendo decisivas para as predições dos modelos e, por fim, ter uma validação médica para melhor entendimento a respeito do significado das features explicadas.

Referências

Aptos 2019 blindness detection. https://www.kaggle.com/c/aptos2019-blindness-detection/data.

Almazroa, A., Sun, W., Alodhayb, S., Raahemifar, K., and Lakshminarayanan, V. (2017). Optic disc segmentation for glaucoma screening system using fundus image.

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., and Webster, D. R. (2016). Development and validation of a deep learning algorithm for detectionof diabetic retinopathy in retinal fundus photographs. *JAMA*.

Lopez-Paz, D., Nishihara, R., Chintala, S., Scholkopf, B., and Bottou, L. (2017). Discovering causal signals in images. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Molnar, C. (2019). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Christoph Molnar.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Model-agnostic interpretability of machine learning.

Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B. S., Waldstein, S. M., and Bogunović, H. (2018). Artificial intelligence in retina.

Szegedy, C., Vanhoucke, V., Ioffe, S., and Shlens, J. (2015). Rethinking the inception architecture for computer vision.

Voets, M., Møllersen, K., and Bongo, L. A. (2019). Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs.

Zhang, Z., Xie, Y., Xing, F., McGough, M., and Yang, L. (2017). Mdnet: A semantically and visually interpretable medical image diagnosis network. *Proceedings of the IEEE conference on computer vision and pattern recognition*.