

---

# Detecção e Interpretação de Retinopatia Diabética usando Aprendizado de Máquina

---

Fernanda Tempesta<sup>1</sup>

*Abstract - A retinopatia diabética é uma das maiores causas de cegueira entre adultos. Visando adquirir maior eficiência no processamento e interpretação das imagens de exame de fundo de olho, aplicou-se técnicas de aprendizado profundo para diagnosticar pacientes que sofrem de retinopatia diabética. Além disso, os mapas de atenção foram utilizados para melhor compreensão a respeito dos pontos que o modelo leva em consideração no momento de identificação de características e classificação das imagens. Por fim, pretendia-se avaliar se existia algum impacto nos resultados ao utilizar datasets de distintas proveniências.*

## 1. Introdução

Em um passado, o uso do termo Inteligência Artificial era empregado principalmente no contexto de jogos digitais, contudo a aplicação de algoritmos para reconhecimento de padrões têm sido amplamente utilizada para as mais diversas áreas, especialmente na medicina. Métodos baseados em *machine learning (ML)* e, particularmente, *deep learning (DL)* estão se tornando especialistas em identificar, localizar e diagnosticar doenças da mácula e da retina através da análise de imagens de exames de fundo de olho (Schmidt-Erfurth et al., 2018).

Segundo trabalho publicado no *National Center for Biotechnology Information (NCBI)*, a retinopatia diabética (RD) é umas das principais complicações relacionadas ao *diabetes mellitus* e a principal causa de perda de visão e de cegueira entre adultos em idade ativa (Stewart JM, 2017). Estima-se que até 2045 existam 693 milhões de pessoas com diabetes em todo o mundo (Cho et al., 2018). A recomendação de muitas diretrizes é que seja feita uma triagem anual para pacientes não diagnosticados com retinopatia ou com quadros leves da doença e um exame periódico de 6 meses para aqueles com retinopatia diabética moderada (Gulshan et al., 2016). Muitos desses exames são feitos por meio da interpretação manual de uma fotografia da retina, o que, além de levar um tempo considerável para a análises, está sujeito

a falhas nos diagnósticos devido à não atenção a detalhes particulares que são característicos da doença, sem contar o fato de que realizar tais exames periodicamente apresenta um custo elevado para o paciente.

Com o objetivo de auxiliar um diagnóstico mais preciso, técnicas baseadas em aprendizado de máquina estão sendo utilizadas para detectar doenças da retina e edema macular, uma vez que a detecção correta do caso fará toda a diferença na recomendação do tratamento apropriado (Carson Lam et al., 2018).

A ideia é que o uso de modelos baseados em *machine learning* e *deep learning* possa servir como uma ferramenta de apoio aos médicos durante a triagem, diagnóstico e orientação da terapia mais apropriada para cada caso e indivíduo de acordo com o desenvolvimento da enfermidade. A palavra final sempre caberá a um especialista, entretanto, devido à eficiência no processamento de dados, a computação tem potencial para se tornar um excelente suporte na análise de exames que requerem bastante precisão.

Já se pode notar um grande avanço no que diz respeito ao uso de tecnologias para apoio à tomada de decisão em áreas médicas, contudo, os chamados “*modelo caixa preta*” ainda são considerados uma barreira para o completo uso de tais técnicas no dia-a-dia das clínicas e hospitais. Isto é, apesar de apresentarem bons resultados, os modelos recebem uma imagem como entrada e oferecem um diagnóstico como saída, mas não dizem muito sobre o processamento das informações, melhor dizendo, sobre os pontos levados em consideração para chegar a tais conclusões. Sendo assim, explicabilidade é um dos itens que ainda precisa ser melhorado, especialmente no que se refere a diagnósticos médicos, para garantir uma maior confiabilidade nos resultados e, conseqüentemente, maior uso dessas ferramentas.

Desta forma o objetivo deste trabalho consiste em (i) utilizar um modelo de aprendizado capaz de realizar o diagnóstico da retinopatia diabética através de imagens de fundo de olho, (ii) identificar se há impacto nos resultados ao utilizar dataset de distintas proveniências em um mesmo modelo e (iii) explicar os pontos levados em consideração que justificam o diagnóstico.

---

<sup>1</sup>Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil. Correspondence to: Fernanda Tempesta <fernandatempesta@dcc.ufmg.br>.

## 2. Referencial Teórico

A seguir serão apresentados conceitos e assuntos pertinentes ao entendimento do problema a ser tratado no presente trabalho. Tais conceitos fazem referência às áreas de Inteligência Artificial, mais especificamente Aprendizado de Máquina e Aprendizado Profundo, Redes Neurais e Medicina.

### 2.1. Inteligência Artificial

Inteligência Artificial (IA) é um ramo da ciência da computação que se dedica a criar algoritmos capazes de analisar dados, reconhecer padrões e tomar decisões com respeito a esses dados simulando alguma tarefa feita pela inteligência humana, porém com maior eficiência e precisão (Lobo, 2018).

### 2.2. Aprendizado de Máquina e Aprendizado Profundo

A ideia básica por trás de aprendizado de máquina e aprendizado profundo é que ambos são capazes extrair características relevantes a partir de um conjunto de dados fornecidos como treinamento, com o intuito de classificar ou prever alguma informação para o problema em questão (Erickson et al., 2017).

Em um primeiro momento o modelo precisa receber como entrada um conjunto de imagens ou características, juntamente com um *label* informativo contendo a resposta esperada; a essa etapa damos o nome de *treinamento*. Em seguida, se o modelo tiver aprendido corretamente, ele será capaz de, por si mesmo, reconhecer e prever novos casos semelhantes aos utilizados durante o treinamento, mas não os mesmo.

Já o termo aprendizado profundo, ou *deep learning* como seria o nome original, é um caso particular do aprendizado de máquina, porém ao invés de ter somente uma unidade de processamento, comumente conhecida como neurônio, conta com várias delas. Cada um desses neurônios fornece informação sobre uma parte do problema, e, ao interconectar todos eles, obtém-se as chamadas *Redes Neurais Convolucionais*, cujo poder de processamento supera as estimativas do aprendizado de máquina tradicional.

### 2.3. Interpretabilidade

Em linhas gerais, *interpretabilidade* em aprendizado de máquina significa explicar ou apresentar razões compreensíveis e entendíveis por um ser humano (Doshi-Velez & Kim, 2017). Atualmente um dos maiores desafios enfrentados por algoritmos de aprendizado de máquina e aprendizado profundo é o fato de eles serem considerados “*modelo caixa preta*”, pois apesar de apresentarem boa performance e ótimos resultados, falham na qualidade de serem transparentes,

isto é, de justificarem os critérios que levaram em consideração para chegarem a determinada predição.

Sendo assim, a condição de interpretável se faz fundamental em modelos de diagnósticos médicos, uma vez que um resultado mal compreendido pode colocar em risco a vida de milhares de pessoas. Contudo, ainda não foi encontrada a melhor solução para conciliar boa performance com transparência, pois se por um lado os modelos de aprendizado de máquina optam por oferecer maior suporte à interpretabilidade, os de aprendizado profundo preferem priorizar o desempenho (Schmidt-Erfurth et al., 2018).

### 2.4. Retinopatia Diabética

A retinopatia diabética é uma doença que envolve alterações nos vasos sanguíneos da retina, devido à alta concentração crônica de açúcar no sangue de pessoas diabéticas, podendo causar sangramento ou vazamento de líquidos nos pequenos vasos sanguíneos da retina, o que causa distorções na visão. A falta de um diagnóstico precoce e de um acompanhamento apropriado pode levar o paciente diabético à uma perda da visão (Stewart JM, 2017).

Por esse motivo, estão sendo propostas técnicas baseadas em inteligência artificial capazes de fornecer um diagnóstico preditivo com maior rapidez e precisão de modo a evitar que pacientes cheguem a quadros críticos da doença.

### 2.5. Trabalhos Relacionados

O uso de modelos de *deep learning* para diagnóstico e classificação de imagens médicas vem crescendo cada vez mais. Desta forma, estudos sobre doenças da retina através da análise de imagens de fundo de olho não é uma exceção. Em 2016 foi publicado um trabalho validando o uso de técnicas de *deep learning* para detecção de retinopatia diabética por meio de imagens de fundo de olho utilizando-se de um dataset privado (Gulshan et al., 2016). Em 2019 foi publicada uma reprodução desse mesmo trabalho, porém avaliando a performance do algoritmo proposto anteriormente com um dataset de domínio público (Voets et al., 2019).

Apesar de ter sido provado que redes neurais profundas conseguem desempenhar eficientemente a tarefa de classificação de imagens médicas bidimensionais, a grande limitação é encontrar dataset grande, diversificado e anotado para a realização do treinamento dos modelos. Além disso, o problema de reconhecimento e classificação de imagens tridimensionais ainda era um limitante até então, contudo em 2018 foi publicado um trabalho que propõe uma nova arquitetura de aprendizado profundo para lidar com esse obstáculo (De Fauw et al., 2018).

### 3. Metodologia

Como o objetivo principal do trabalho era classificar e interpretar a retinopatia diabética por meio de modelos de aprendizado de máquina e validar em datasets de distintas proveniências, o trabalho consistiu de uma sequência de etapas.

A Figura 1 exibe um esquema ilustrativo sobre a metodologia utilizada no desenvolvimento do projeto. Ela consistiu em 4 etapas, sendo (i) Treinamento, (ii) Classificação, (iii) Interpretabilidade e (iv) Validação com especialista.

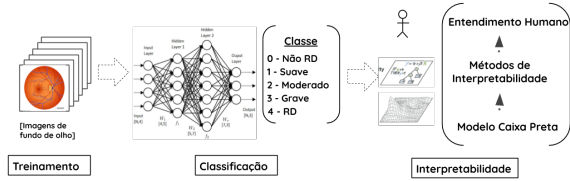


Figure 1. Fluxograma da metodologia proposta

A seguir, será detalhado o propósito de cada uma das etapas.

#### 3.1. Treinamento

##### 3.1.1. DATASET

Este trabalho fez uso de dois datasets de domínio público, disponibilizados pela plataforma do [Kaggle](#).

O primeiro dataset utilizado foi fornecido por *EyePACS* para a competição *Diabetic Retinopathy Detection, 2015* realizada com o apoio do California Healthcare Foundation.

O segundo dataset foi disponibilizado para a competição *Blindness Detection, 2019*, com o apoio do *4th Asia Pacific Tele-Ophthalmology Society (APTOS) Symposium*.

##### 3.1.2. PRÉ-PROCESSAMENTO DOS DADOS

Com o objetivo de melhorar a performance do algoritmo, os dados foram divididos em conjunto de *treino*, *validação* e *teste*.

O conjunto de treino é uma parcela dos dados utilizada para que o modelo aprenda a identificar características importantes. Já o conjunto de validação é utilizado para avaliar a performance, fornecendo ao desenvolvedor parâmetros para identificar se o algoritmo está aprendendo a classificar corretamente as instâncias ou se ainda precisa de melhorias. E por último, o conjunto de teste é utilizado somente no final como uma última prova do desempenho do modelo.

Como este é um problema que envolve várias classes, um primeiro desafio era lidar com o desbalanceamento das mesmas. Após a divisão dos dados em treino e validação, foi feita um balanceamento das classes no conjunto de treino

de modo a evitar que o algoritmo apresentasse um viés para alguma delas.

Uma outra técnica bastante utilizada em algoritmos de aprendizagem profunda é conhecida como *data augmentation*. Essa técnica possui o propósito de ampliar artificialmente o dataset de treino com variações das imagens originais de modo a melhorar a performance do algoritmo ao fornecer mais imagens para que o modelo possa aprender.

#### 3.2. Classificação

Para a classificação das imagens de fundo de olho, foi utilizado algoritmos de aprendizado profundo devido à facilidade e robustez que tais modelos oferecem no processamento de imagens.

Outra técnica que permite ganho de eficiência é denominada *transfer learning*, isto é, por meio dela é possível utilizar algum modelo já previamente treinado para algum problema similar, porém em escalar maior, de modo a aproveitar os pesos configurados anteriormente. Utilizar esse tipo de técnica apresenta a vantagem de economizar tempo na etapa de treinamento, pois o algoritmo não precisará configurar os pesos desde o início, fazendo com que ele convirja mais rapidamente ([Kornblith et al., 2019](#)).

A tarefa de classificação consistia em dada uma determinada entrada de uma imagem de fundo de olho, classificá-la em uma das 5 classes de acordo com ausência ou não da retinopatia diabética (RD), segundo o grau da doença:

##### Níveis de severidade da doença

0	Não apresenta
1	Suave
2	Moderado
3	Grave
4	Proliferativo

As Figuras 2, 3 e 4 ilustram exemplos de imagens de fundo de olho para cada um dos níveis.

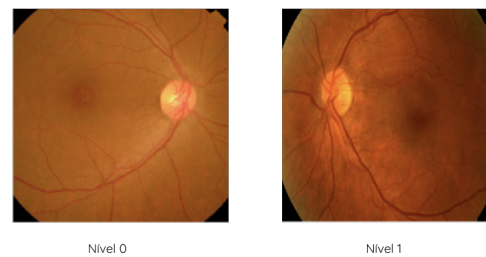


Figure 2. Exemplos de imagens de fundo de olho, níveis 0 e 1.

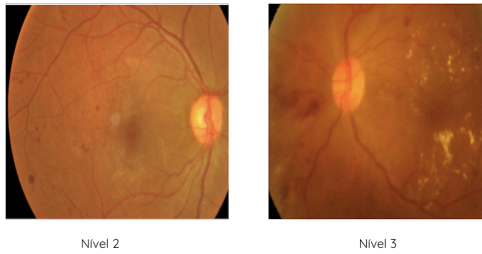


Figure 3. Exemplos de imagens de fundo de olho, níveis 2 e 3.

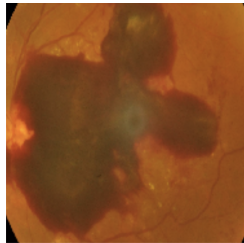


Figure 4. Exemplo de imagem de fundo de olho, nível 4.

### 3.2.1. EXPERIMENTOS

Com o intuito de saber se datasets provenientes de diferentes lugares apresentam impactos quanto ao desempenho do modelo, um mesmo algoritmo foi avaliado em dois conjuntos de dados (i) dataset advindo de uma instituição americana (ii) dataset advindo de uma instituição indiana.

Foram gerados 4 combinações de treino e teste dos datasets, conforme esquematizado na Tabela 1:

Esquema de treino/teste dos datasets

Treino	Teste
$D_{Americano}$	$D_{Americano}$
$D_{Americano}$	$D_{Indiano}$
$D_{Indiano}$	$D_{Indiano}$
$D_{Indiano}$	$D_{Americano}$

Table 1. Para um mesmo modelo, foram utilizadas distintas combinações dos dataset para treino e teste.

### 3.2.2. AVALIAÇÃO DOS RESULTADOS

Para avaliação dos resultados, foram utilizados os métodos do pacote `accuracy_score` disponíveis na biblioteca do `scikit-learn`. Para comparação dos resultados, foi avaliada a curva ROC (Característica de Operação do Receptor), que seria uma representação gráfica dos Verdadeiros Positivos versus os Verdadeiros Negativos, isto é, o quão bem o algoritmo pode prever que alguém tem a retinopatia diabética dado que a

pessoa realmente tem, versus o quão bem ele pode prever que a pessoa não tem a doença dado que ela realmente não tem.

### 3.3. Interpretabilidade

Apesar de muitas vezes apresentarem bons resultados, a grande limitação dos modelos de aprendizado de máquina é que eles não conseguem lidar com o problema da “caixa preta”, isto é, não é muito claro, os pontos que o algoritmo leva em consideração para a conclusão de tal diagnóstico.

Desta forma, os *mapas de atenção* são utilizados para entender quais os pontos que o algoritmo identificou como sendo de maior relevância no momento de classificação de cada uma das imagens, ou seja, qual foi a característica marcante que fez com que o modelo fornecesse determinado diagnóstico (Fukui et al., 2019).

A Figura 5 ilustra alguns exemplos de como os mapas de atenção foram utilizados para compreender melhor as características que o modelo julgou mais relevante para a classificação. As partes em amarelo foram as que receberam maior atenção.

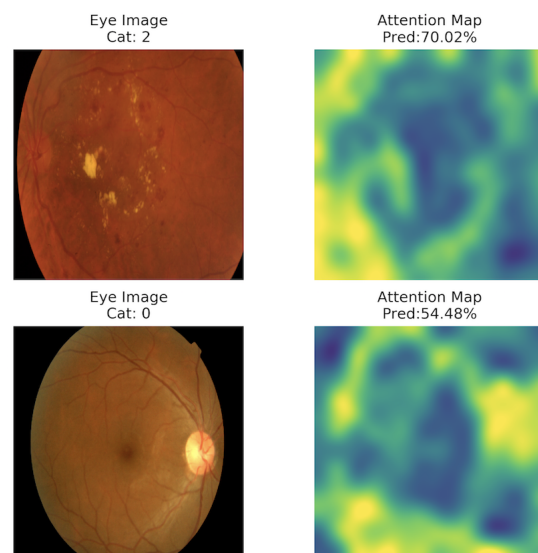


Figure 5. Exemplos do resultado gerado pelos maps de atenção.

## 4. Resultados Obtidos

Para cada um dos teste que se segue, o mesmo modelo de classificação de retinopatia diabética foi treinado e testado para dataset de distintas proveniências, origem americana ou origem indiana.

### 4.1. Treino dataset americano, teste dataset americano

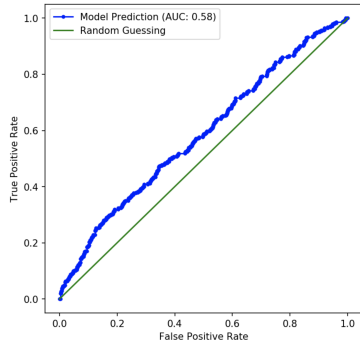


Figure 6. Acurária obtida ao treinar e avaliar em dataset de proveniência americana.

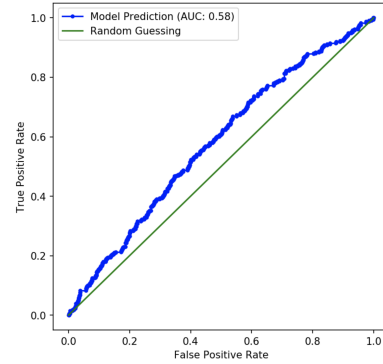


Figure 9. Acurária obtida ao treinar em dataset de proveniência indiana e avaliar em dataset de proveniência americana.

#### 4.2. Treino dataset americano, teste dataset indiano

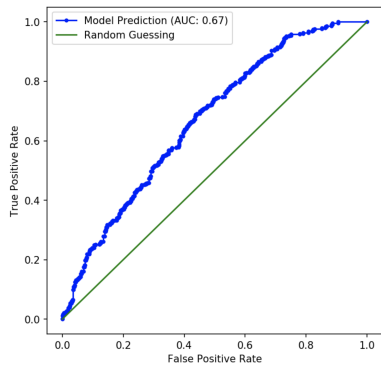


Figure 7. Acurária obtida ao treinar em dataset de proveniência americana e avaliar em dataset de proveniência indiana.

#### 4.3. Treino dataset indiano, teste dataset indiano

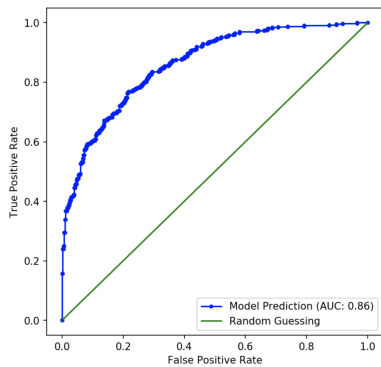


Figure 8. Acurária obtida ao treinar e avaliar em dataset de proveniência indiana.

#### 4.4. Treino dataset indiano, teste dataset americano

A Tabela 2 faz um comparativo dos resultados obtidos em cada um dos casos:

##### Acurária para os diferentes conjuntos de dados

Treino	Teste	Acurácia
$D_{Americano}$	$D_{Americano}$	0.58
$D_{Americano}$	$D_{Indiano}$	0.67
$D_{Indiano}$	$D_{Indiano}$	0.86
$D_{Indiano}$	$D_{Americano}$	0.58

Table 2. Acurária obtida ao treinar e avaliar com cada um dos datasets.

### 5. Discussões e Conclusões

Ao final dos experimentos verificou-se que houve uma diferença entre as acurácias ao avaliar datasets de proveniências distintas em um mesmo modelo. Apenas com esses resultados não se pode concluir que de fato a rede apresenta divergência quanto à origem e qualidade dos datasets, uma vez que os testes realizados não foram suficientes para tal conclusão.

Um ponto a ser considerado como um limitante foi o fato de que os conjuntos de dados possuíam uma relativa diferença de tamanho, o que pode ter influenciado nos resultados. Sendo assim, uma nova avaliação será feita utilizando-se de estratégias que proponham maior homogeneidade quanto ao tamanho e proporção dos datasets.

Por fim, concluiu-se que para tipos de problemas onde detalhes pontuais fazem toda a diferença no diagnóstico, o uso de técnicas de aprendizado profundo e de interpretabilidade se fazem fundamentais na qualidade e eficiência da predição

do diagnóstico. Além disso, o uso de mapas de atenção ajuda a identificar possíveis falhas e limitações dos modelos, bem como a compreender melhor como os algoritmos aprendem por meio das imagens.

Como trabalhos futuros pretende-se utilizar Funções de Influência para identificar as imagens do treino que mais tiveram impacto na predição do modelo.

## Referências

- Carson Lam, D. Y., Guo, M., and Lindsey, T. Automated detection of diabetic retinopathy using deep learning. *AMIA Summits on Translational Science Proceedings*, 2018:147, 2018.
- Cho, N., Shaw, J., Karuranga, S., Huang, Y., da Rocha Fernandes, J., Ohlrogge, A., and Malanda, B. Idf diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes research and clinical practice*, 138:271–281, 2018.
- De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O’Donoghue, B., Visentin, D., et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, 24(9):1342, 2018.
- Doshi-Velez, F. and Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.
- Erickson, B. J., Korfiatis, P., Akkus, Z., and Kline, T. L. Machine learning for medical imaging. *Radiographics*, 37(2):505–515, 2017.
- Fukui, H., Hirakawa, T., Yamashita, T., and Fujiyoshi, H. Attention branch network: Learning of attention mechanism for visual explanation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10705–10714, 2019.
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.
- Kornblith, S., Shlens, J., and Le, Q. V. Do better imagenet models transfer better? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2661–2671, 2019.
- Lobo, L. C. Artificial intelligence, the future of medicine and medical education. *Revista Brasileira de Educação Médica*, 42(3):3–8, 2018.
- Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B. S., Waldstein, S. M., and Bogunović, H. Artificial intelligence in retina. *Progress in retinal and eye research*, 67:1–29, 2018.
- Stewart JM, Coassin M, S. D. Diabetic retinopathy. In: *Feingold KR, Anawalt B, Boyce A, et al., editors. Endotext [Internet]*, South Dartmouth (MA): MDText.com, Inc.; 2000:1–44, 2017.
- Voets, M., Møllersen, K., and Bongo, L. A. Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PloS one*, 14(6):e0217541, 2019.