

Detectando retinopatia diabética utilizando Deep Learning

João Carlos Pandolfi Santana

Novembro 2018

1 Introdução

A retinopatia diabética é a principal causa de cegueira na população em idade ativa do mundo desenvolvido. Estima-se que a condição afete mais de 93 milhões de pessoas.

A necessidade de um método abrangente e automatizado de triagem de retinopatia diabética tem sido reconhecida há muito tempo, e os esforços anteriores fizeram um bom progresso usando classificação de imagem, reconhecimento de padrões e aprendizado de máquina. Com fotos de olhos como entrada, o objetivo deste trabalho é criar um novo modelo, idealmente resultando em potencial clínico realista.

As motivações para este projeto são duas:

- O processamento de imagem e classificação eficiente de imagens tem sido um interesse pessoal e objeto de estudo há anos, além de classificação a análise de dados em grande escala e auxílio a exames médicos.
- O processo normal de avaliação e diagnóstico é relativamente lento, consistindo em análise de imagens pelos médicos e agendamento da consulta de acompanhamento. Ao processar imagens em tempo real, o sistema permitiria uma redução de custos e agilização do diagnóstico e tratamento do paciente.

2 Projeto

A pesquisa consiste em desenvolver um sistema, utilizando Redes Neurais (*Deep Learning*), que classifique imagens de fundo de olho e detecte **retinopatia diabética**. De forma que consiga agilizar o diagnóstico do paciente e iniciar o tratamento mais rápido possível.

Outra vantagem do produto gerado pela pesquisa, é a redução de custos e aumento de eficiência em diagnósticos. Com isto, melhorando o atendimento do Hospital e/ou Clínica que utilize o sistema.

2.1 Produto

O produto gerado poderá ser comercializado em forma de serviço, onde a imagem será aplicada para diagnóstico e receberá o resultado em poucos segundos. Podendo assim, ser cobrado por imagem ou diagnóstico efetuado, aumentando a renda do hospital e instituição.

Como o sistema utilizará o conceito de *aprendizado contínuo supervisionado*, a cada diagnóstico bem sucedido, aumentará progressivamente a sua base de dados e ficará mais assertivo.

A interface de utilização será disponibilizada por um serviço web acessível por uma página de internet, onde o usuário terá que efetuar um cadastro (a ser discutido o processo) para utilizar o sistema. Após o cadastro, poderá carregar as imagens do exame para serem analisadas pelo sistema. Após o processamento, o resultado é exibido em uma página de internet e pode ser feito o *download* no formato *pdf*. Provavelmente não será abordada neste trabalho em questão, em vista deste requisito fugir do propósito da disciplina.

2.1.1 Ferramentas

Ferramentas necessárias para desenvolvimento:

- Rede Neural na estrutura *Deep Learning* (Aprendizado Profundo)
- Processamento de imagem
 - Normalização de cores
 - Redimensionamento da imagem
 - Corte do centro ¹
 - Geração de novas imagens
 - * Rotacionando imagens a 90°
 - * Rotacionando imagens a 180°
 - * Rotacionando imagens a 270°
 - * Espelhando imagem na *Horizontal*
- Análise de dados
 - Calcular Taxa de acerto
 - Calcular Perda

3 Metodologia

Metodologia aplicada na pesquisa

¹Remover bordas da imagem redonda e dar ênfase no fundo do olho

3.1 Obtenção dos dados

A obtenção da massa crítica de dados será feita utilizando bases de dados abertas disponíveis na internet, assim como bases de dados cedidas por instituições de pesquisa.

Os dados de validação e adaptação a realidade atual, irão ser obtidos pelos alunos da *Emescam* em parceria com o professor doutor em oftalmologia *Bruno Valbon*, onde serão utilizados para validação e adaptação da ferramenta a realidade do *Hospital Santa Casa de Misericórdia do Espírito Santo*.

Estes dados se encontram em posse da instituição mencionada acima e não tenho permissão para divulgação do mesmo. A base de dados passada para o professor na etapa de processamento é somente um pedaço da base original, contendo *35 GB* de treino como pode ser visto na Figura 1 e *53,3 GB* de teste Figura 2

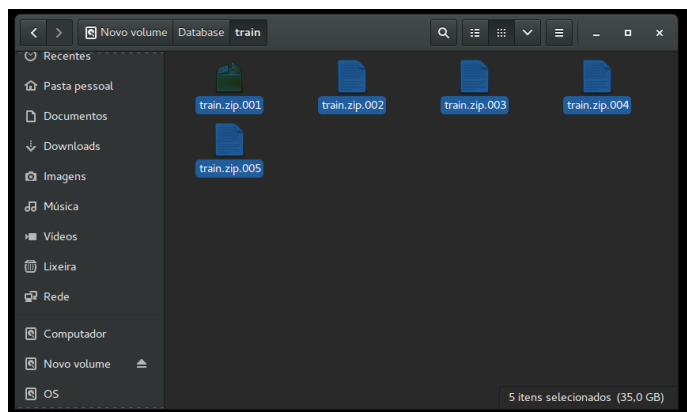


Figure 1: Tamanho da pasta de treino

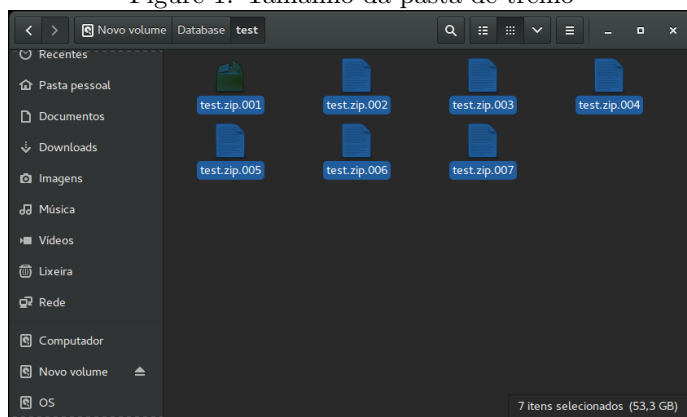


Figure 2: Tamanho da pasta de teste

3.1.1 Estruturação dos dados

Os dados coletados, serão categorizados por idade e diagnóstico real feito por especialista. Assim como por doenças oculares prévias que o paciente desenvolveu durante a vida.

Esta etapa foi pulada devido ao fato da base de dados obtida externamente não possuir categorização por idade e não conter dados passados, ou seja, somente uma parte da amostra contém essa informação, portanto deixando esta etapa para um trabalho futuro.

3.2 Tratamento dos dados

Após a categorização dos dados obtidos, a separação em dois pacotes será aplicada da seguinte forma: *Treino* e *Teste*. Onde serão utilizados para treinamento e regulação dos resultados obtidos.

Os dados de treino contém imagens coletadas pela instituição de apoio e imagens obtidas através da base de dados cedida por outras instituições de pesquisa parceiras e bases de dados abertas. As imagens usadas para teste também possuem a mesma mistura de fontes, porém, possuem uma proporção maior de imagens coletadas do que a de treino, sendo 70% deles para treino e 30% para teste.

3.2.1 Normalização das cores

Diferentes aparelhos podem gerar tons e escalas de cores variados, isso faz com que dificulte a curva de aprendizado do algoritmo, assim a normalização da escala e tons será feita, potencializando assim o resultado.

Após estudar melhor o comportamento da rede neural aplicada (Rede neural convolucional), cheguei a conclusão que esta etapa não seria necessária, em vista que a própria rede já equaliza os dados de entrada por sua própria característica

3.2.2 Geração de novas imagens

Para evitar o *overfitting* da rede e que ela aprendesse somente a classificar as imagens dependendo da posição e rotação inserida, foram geradas imagens plásticas para reforçar o treinamento. Estas são modificação simples nas imagens originais, sendo elas rotação nos ângulos 90°, 180° e 270°. Assim como espelhamento *Horizontal*, aumentando a base de dados em quatro vezes, ganhando nos pontos:

- Aumento na quantidade de amostras de teste e treino
- Ampliamento do espectro de classificação ²

²Garantindo que qualquer posição da imagem inserida seja suficiente

3.2.3 Segmentação

Seleção automática das regiões de interesse do algoritmo. Essas regiões serão utilizadas para treinamento e direcionamento do aprendizado.

Após uma análise por amostragem nas imagens, observei que todas possuem o mesmo padrão, são imagens quadradas com bordas circulares pretas 3. Para garantir que somente a parte importante seja inserida na Rede, recortei o centro como tamanho padrão de $1800 \times 1080 \text{ px}$, logo após redimensionei para $256 \times 256 \text{ px}$ isso faz com que as imagens geradas sejam menores e consumam menos recurso computacional Figura 4 e Figura 5. O motivador principal para redimensionamento foi o tamanho de entrada da Rede Neural, quanto maior a entrada, muito maior o custo computacional para processamento e maior a chance de erro.

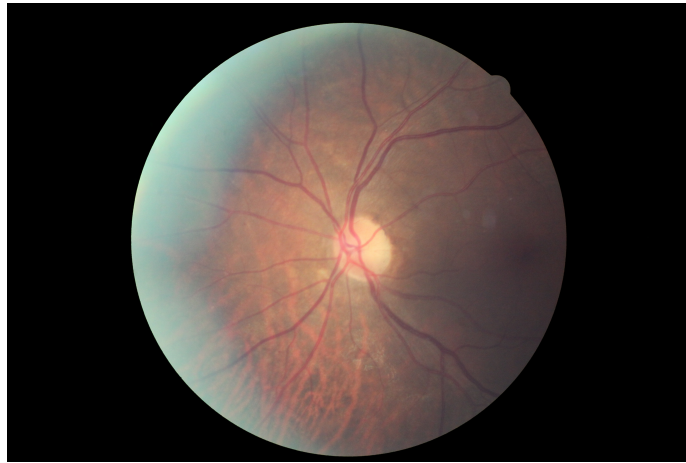


Figure 3: Amostra de imagem original

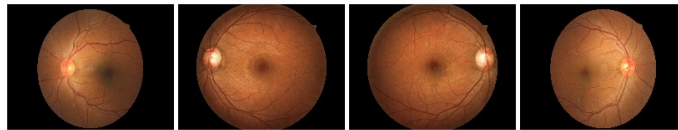


Figure 4: Imagem antes do processamento.



Figure 5: Imagem cortada e redimensionada após o processamento

O fator físico foi um grande limitante neste trabalho, em vista que a Rede

Neural utilizada se vale de *CUDA* ou seja processamento em placa gráfica, um recurso caro e limitado, por este motivo tive que redimensionar a rede e o tamanho da entrada para caber na memória da placa utilizada (*GTX 1070 TI* com 8Gb de RAM dedicada).

3.3 Modelagem da rede

Para resolução do problema, modeliei uma rede neural com a característica convolucional e de aprendizado profundo, como pode ser visto na Figura 7.

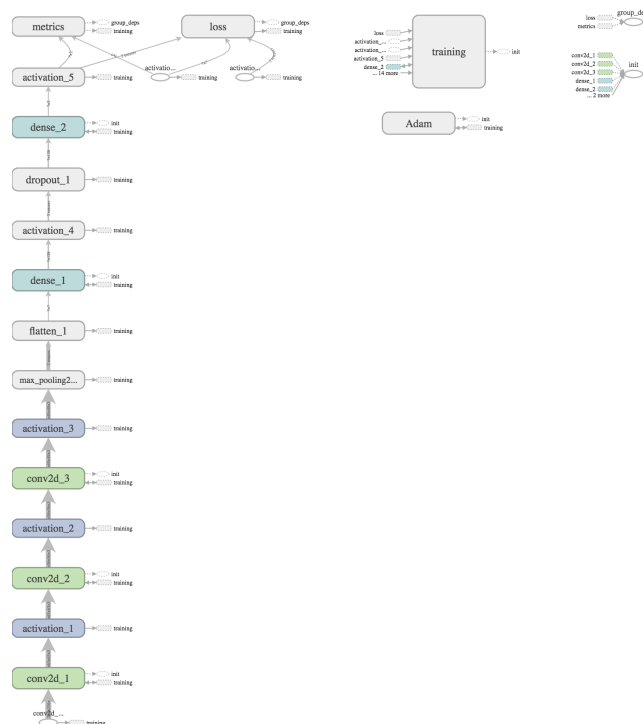


Figure 6: Modelo da rede neural utilizada

Foi utilizado este modelo, pelo fato da primeira camada, a parte convolucional, ser especializada em retirar *features* da imagem. Essa característica se dá pela rede executar várias convoluções com *kernels* inicialmente aleatórios de tamanho previamente definido (8×8).

A segunda etapa da rede segue a característica de *Max Pooling*, onde visa recuperar somente as *features*³ que passarem por um limiar de importância.

A terceira e ultima camada segue o conceito de rede densa totalmente conectada, onde visa encontrar os padrões necessários para resolver o problema. Uma

³Features identificadas na primeira etapa da rede (convolucional)

Rede Neural é uma forma de regressão matemática, onde dado uma entrada, encontrar a equação matemática para a saída esperada. Esta camada é responsável por informar o resultado final da classificação, variando de 0 a 4 (0 não tem retinopatia e 4 tem retinopatia grave).

3.4 Treinamento e validação

Nesta etapa foram executados os experimentos nas bases definidas previamente (treino e teste). Para retirada das métricas, foram utilizadas funções da própria biblioteca de Redes Neurais utilizada ⁴, sendo estas responsáveis por calcular *Precisão*, *Recall*, *Acurácia na etapa de testes* e *Taxa de erro na etapa de testes*. Estas métricas são utilizadas para definir a qualidade e efetividade da ferramenta. Os valores obtidos podem ser observados na Tabela 3.4 e no *print* do terminal ao fim do processamento Figura ??.

Métrica	Valor
Accuracy (Train)	82%
Accuracy (Test)	80%
Precision	88%
Recall	77%

Table 1: Tabela de resultados

```
Test score:81.97431125
Test accuracy:80.00902146
Precision: 88.38404853
Recall:77.00794624
```

Figure 7: Resultado exibido no terminal após execução.

4 Conclusão

Foi enviado uma massa de testes para o professor executar com as características que podem ser vistas na Figura 8, por este motivo os resultados apresentados aqui podem ser diferentes com o encontrado ao executar com a base enviada.

```
Test score:81.97431125
Test accuracy:80.00902146
Precision: 88.38404853
Recall:77.00794624
```

Figure 8: Terminal de execução enviado ao professor.

⁴Keras

Apesar da execução demorada e da grande quantidade de peculiaridades que a ferramenta *Keras* apresenta, o desenvolvimento da solução não se demonstrou muito complexa. A metodologia utilizada foi abordada na maioria dos problemas de visão computacional, obtendo assim um resultado esperado.

As métricas obtidas se apresentaram abaixo do esperado (esperava obter 95% de precisão), porém não são valores ruins, dada a possibilidade de adiantamento do diagnóstico do paciente e agilizar o tratamento, impedindo assim a cegueira causada pela doença. Problemas de classificação, geralmente se apresentam na categoria de problemas de fácil resolução, acredito eu que o resultado encontrado não foi melhor, devido a limitação computacional determinada pela placa de vídeo, impedindo assim o crescimento da rede aplicada.

Para trabalhos futuros, utilizarei redes mais complexas e maiores para melhoramento do resultado assim que obtiver acesso a um hardware de maior poder computacional.