RELATÓRIO - TESTES REALIZADOS PÓS SEGUNDA ENTREGA

* foram realizados testes com 3 redes diferentes: AlexNet, ResNet18 e ADDNet, para verificar como as 3 redes se comportariam
* além disso, foram estudados 3 datasets: Kaggle (que são dados do OASIS), OASIS e ADNI.
* os primeiros testes realizados foram para verificar o comportamento inicial das redes, sem nenhuma transformação nas imagens. Logo de cara, a acurácia já estava muito alta, acima de 0.80 e isso nos estranhou bastante. Realizamos esse primeiro treinamento com 20, 50 e 100 épocas.
* fizemos testes com diversas transformações: rotação em 10 graus, flip horizontal, resize, crop, normalização, adição de ruído gaussiano, adição de pad, grayscale. Verificamos que dessas transformações, uma ou outra prejudicava pouquíssimo o treinamento da rede. E todas as outras, não ajudavam a melhorar, o desempenho das redes ficava na mesma. Conferimos com as matrizes de confusão de cada transformação e o resultado era mantido.
* foram realizados testes com o dataset ADNI, foi realizado um teste com a rede ADDNet com apenas 10 épocas sem nenhuma transformação e o desempenho da rede já foi muito alto.
* realizamos testes também de classificação binária entre controle e muito leve. O foco aqui era tentar entender o motivo do desempenho da rede estar tão alta
* descobrimos o que estava causando esse alto desempenho: o modelo não aprendeu as características da patologia, mas sim, aprendeu a associar o cérebro do indivíduo ao diagnóstico, por conta das slices que foram necessárias para conseguirmos processar as imagens todas. Múltiplas slices de múltiplas imagens de um mesmo indivíduo, de forma que tinham slices do mesmo indivíduo tanto no treinamento quanto na validação/teste e com isso, a rede treinava com a imagem de um indivíduo que tinha aquele diagnóstico e tinha slice desse mesmo indivíduo no teste e aí, ficava fácil a classificação correta
* fizemos a separação das slices e pastas, garantindo que teriam indivíduos únicos em cada pasta e com isso, o desempenho da validação e do teste caíram (separação aleatória porém intencional pra não repetir o indivíduo)
* fizemos outro teste pegando somente a slice 01 de cada indivíduo e realizamos os testes, mas tivemos problemas pois ficamos com poucas amostradas, sendo difícil de realizar um bom treinamento, uma das classes ficou apenas com 2 amostras, bem desbalanceado. Esse teste fez o desempenho cair muito e as transformações nesse momento não estão ajudando muito, pois temos poucas amostras
* tentamos também um dataset misturado, tentamos multiplicar a quantidade de amostras, tentamos balanceamento feito a mão mesmo… e nenhuma dessas estratégias funcionou
* testes sendo realizados com o dataset do oasis completo, sem a divisão das slices, com 50, 60, 75 e 100 épocas -> sem muitas melhores, independente das transformações usadas
* testes com balanceamento de classes através da função de perda, tanto com o OASIS completo quanto com o OASIS com uma slice só de um indivíduo só
* nos testes com o OASIS com uma slide só de um indivíduo só, a acurácia está muito baixa, mesmo com o balanceamento de classes. As transformações que auxiliaram foram: resize, normalize, gaussian, pad - com 20 épocas. Com 50 épocas, o horizontal flip e grayscale ajudaram, mas o pad não ajudou, piorou.