UNIVERSIDADE FEDERAL DO ABC
Engenharia de Informação

GABRIEL MESQUITA DE SOUZA

CLASSIFICAÇÃO DE ALZHEIMER POR IMAGENS MRI: ABORDANDO DIFERENTES TÉCNICAS DE DEEP LEARNING

Santo André
2021

GABRIEL MESQUITA DE SOUZA

CLASSIFICAÇÃO DE ALZHEIMER POR MEIO DE IMAGENS MRI: ABORDANDO DIFERENTES TÉCNICAS DE DEEP LEARNING

Trabalho apresentado como requisito parcial para a conclusão do curso de Engenharia de Informação da Universidade Federal do ABC

Orientador: Prof Dr. Luneque Del Rio de Souza e Silva Junior

Santo André
2021

Dedico este trabalho especialmente às pessoas que enfrentam doenças psiquicas e a todos os seres sencientes que estão mergulhados em mares de sofrimento.
Espero que despertem para perceber a manifestação de sabedoria pura no surgir de cada fenômeno.

"Quando tudo tá perdido na vida
Só quando tudo tá perdido na vida
É que a gente descobre que na vida
Nunca tudo tá perdido."

Resumo

Com o envelhecimento da população mundial o progresso no número de indivíduos com a Doença de Alzheimer tem subido cada vez mais. Enquanto a cura ainda continua desconhecida, muitos pesquisadores e instituições de saúde tem se dedicado a investigar novas técnicas que possibilitem um diagnóstico mais precoce da doença, possibilitando ao infermo tomar medidas que remediem o avanço do quadro e dando tempo à sua família para planejar-se financeiramente para os cuidados especiais que deverão ser tomados.

Hoje as imagens de Ressonância Magnética são uma das principais formas de diagnóstico do Alzheimer e com o avanço da visão computacional na medicina, classificadores inteligentes conseguem extrair e reconhecer características nos pixels das imagens que muitas vezes são imperceptíveis ao sistema visual humano. Essa técnica também tem se popularizado em pesquisas feitas para o diagnóstico do Alzheimer nos últimos anos.

Nesse projeto será construindo uma nova abordagem introduzindo uma arquitetura já existente de um classificador no contexto do diagnóstico de pacientes com Comprometimento Cognitivo leve, um estágio intermediário de perda de memória e que pode levar ao estado de Alzheimer no enfermo. Focando na implementação e posterior análise da arquitetura e de seus parâmetros, o projeto também visa explorar a eficiência do modelo quando complementado com duas técnicas de Deep Learning frequentemente utilizadas.

O modelo foi construído utilizando a Rede Convolucional Densamente Conectada, uma arquitetura de classificador criada em 2017 que permitiu um avanço na perfomance do treinamento de Redes Convolucionais para imagens 2D, e que será introduzida no contexto do projeto, além de analisar o efeito da aplicação das técnicas de Data Augmentationm, Transferência de Aprendizado utilizando os parâmetros da DenseNet já estabelecidos pelo treinamento no projeto ImageNet e com o balanceamento de classes, na eficácia do modelo. Tais técnicas já amplamente utilizadas no contexto de classificação de imagens mas que ainda pouco foram discutidas no contexto específico de imagens médicas, especialmente no uso de modelos classificadores de Alzheimer por imagens de Ressonância Magnética.

Abstract

As the world population age increases the number of subjects diagnosed with Alzheimer Disease has also been growing. While the cure of the disease is still unknown, a large number of researches and health institutions are dedicated to explore new techniques that allow an earlier diagnosis of disease, giving patient the possibility to take some measures to slow up the disease's progress and also give more time to patient's family planning involved expenses with future special cares he will need.

Nowadays, the MRI imaging is one of the most important Alzheimer's diagnosis tools and with the advances of computer vision applied to medicine, smart classifiers can extract and recognize some features in the image's pixels that most part of the time are imperceptible to visual human system. This technique has become famous in Alzheimer's diagnosis papers in the last few years.

In this project it was built a classifier model for the diagnosis of patients with Mild Cognitive Impairment, an intermediate stage of memory loss which can lead to Alzheimer's Diseade. Focusing on the implementation and subsequent analysis of a new approach based on an existent classifier architecture as well as the adjustment of its parameters, in addition to two other Deep Learning techniques.

The model was built using the Densely Connected Convolutional Network, a classifier architecture created in 2017 that allowed advances in the performance of Convolutional Networks training for 2D images. It will be introduced in the context of the project, in addition to the Data Augmentation and Transfer Learning techniques using the parameters of DenseNet established in the training of the ImageNet project. Such techniques have already been frequently used in many applications but that has not yet been evaluated in the specific context of medical images, especially in the use of Alzheimer's Disease classifier models by MRI.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

Na última década a implementação de sistemas inteligentes tem sido utilizado em diversas áreas, vendo seu uso crescer amplamente, especialmente na área chamada de Aprendizado Profundo, ou *Deep Learning* (DL) do inglês. Tal área é referente à uma classe de técnicas e algoritmos de Aprendizado de Máquina que possibilitam descobrir inúmeros padrões principalmente em grandes conjuntos de dados em que o aprendizado é dividido em pequenas funções dentro de uma rede. Embora eficiente em muitas ocasiões, essa ampla implementação nos mais variados tipos de problemas só foi possível graças à crescente expansão na disponibilidade de dados armazenados e ao avanço no desenvolvimento de poder computacional nas últimas décadas, para treinar, analisar e manipular toda essa grande massa de informação (LUNDERVOLD; LUNDERVOLD, 2018, p. 102).

A técnica utilizada por algoritmos de DL para reconhecer padrões nos dados que nos sirvam com alguma informação útil para posteriores tomadas de decisão, é feita por meio do aprendizado distribuido em camadas, utilizando as famosas Redes Neurais. Esses algoritmos foram criados baseando sua arquitetura no funcionamento do cérebro humano e que normalmente obtém melhores resultados trabalhando com dados não estruturados (como imagens) se comparado a outros algoritmos de machine learning convencionais. As arquiteturas das redes foram se tornando cada vez mais profundas conforme o poder computacional possibilitou essa expansão, devido ao elevado número de parâmetros requeridos para treinar grande parte das arquiteturas de Redes Neurais (LECUN; BENGIO; HINTON, 2015; SKANSI, 2018).

A área que trabalha com o Aprendizado de Máquina baseando-se em imagens como fonte de dados é chamada de visão computacional. Devido ao grande volume de dados presente em uma imagem e por esses serem processados de forma não estruturada, sendo que em grande parte das vezes os algoritmos que melhor trabalham com esses dados são as Redes Neurais. Além disso, o aprendizado por imagem normalmente requer do sistema a resolução de problemas mais complexos do que aqueles demandados por dados estruturados, com um aprendizado distribuído em camadas e compartilhado pela rede até que a mesma identifique padrões mais complexos mostra-se mais eficiente para abstrair inúmeros padrões dentro de um volume de dados muito grande, por isso na maior parte das vezes as técnicas de DL obtem maiores acurácias do que modelo mais convencionais quando utilizados dentro do contexto de visão computacional (RAZZAK; NAZ; ZAIB, 2017, p. 2).

Um dos setores que tem se mostrado cada vez mais proeminente no uso

de técnicas de Aprendizado Profundo é a medicina, a qual também colheu frutos do rápido desenvolvimento do poder computacional nas últimas décadas. Hoje em diversas instituições do setor de saúde uma enorme quantidade de dados é gerada e armazenada, contando na maior parte das vezes com arquivos contendo um volume de dados maiores do que comumente vistos em formatos de dados tradicionais. Não somente a capacidade de armazenamento desses dados está diretamente relacionada com o avanço computacional, mas também a sua geração: (RAZZAK; NAZ; ZAIB, 2017) cita que exames como Tomografia Computadorizada(TC), Ressonância Magnética , ou do inglês *Magnetic Ressonance Imaging* (MRI), e Raios X aperfeiçoaram sua capacidade de resolução de forma crescente nos últimos anos. A melhora na qualidade das imagens coletadas significa maior informação disponível para o ser humano, assim como para o aprendizado das máquinas. Esse avanço paralelo entre as técnicas de DL e imagens médicas tem criado grandes esperanças no diagnóstico precoce de inúmeras doenças, como é o caso da Doença de Alzheimer (DA) que será abordada neste trabalho (LUNDERVOLD; LUNDERVOLD, 2018).

.

1.1 Alzheimer

De acordo com o Protocolo Clínico e Diretrizes Terapeutas do Ministério da Saúde estima-se que 1,1 milhão de pessoas no Brasil sofrem de demência, aliado a esse número, estudos realizados com brasileiros acima de 65 anos mostraram que em mais de 55 % dos casos a demência estava associada com a doença de Alzheimer (DA). Com o aumento da expectativa de vida e do número de idosos pelo qual a sociedade contemporânea está passando, passou a ser observado um consequente aumento no número de indivíduos com Alzheimer, sendo hoje a principal causa de demência no mundo (MOVEMENT, 2017). De acordo com Fernandes e Andrade (2017) a população dentro da faixa etária de maior risco ao desenvolvimento da DA será 22% da população mundial em 2050, sendo esse um dos principais fatores de preocupações da comunidade médica global especializada em saúde mental, no combate à DA.

Avanços iniciados nos dias atuais podem refletir no alcance de uma maior expectativa de vida para gerações futuras, ao passo de que é constatado que a DA reduz em 50% o tempo de vida do paciente (ALMEIDA; GOMES; NASCIMENTO, 2014).

Afetando inicialmente a formação hipocampal, o Alzheimer espalha danos para áreas vitais referentes ao funcionamento da memória, além do Hipocampus a doença afeta também as regiões do Subicolo e o Córtex Entorrinal. Exames neurológicos em pessoas diagnosticadas com Alzheimer indicaram também uma atrofia nas regiões corticais (NITZSCHE; MORAES; JÚNIOR, 2015, p. 228). Os danos causados nessas

regiões cerebrais são irreversíveis, gerando a perda de memória, confusão mental e raciocínio lógico prejudicado aos indivíduos acometidos pela doença (MOVEMENT, 2017).

Conforme os danos da DA no cérebro vão progredindo, o paciente se vê impedido de executar atividades básicas cotidianas, necessitando de cuidados especiais de outra pessoa na maior parte do tempo, mudanças no comportamento e outros sintomas neuropsiquiátricos também são observados. Um diagnóstico precoce da doença é benéfico para os pacientes e suas famílias, dando-lhes tempo para planejar financeiramente os tratamentos e futuros cuidados do enfermo, além de possibilitar a prática de atividades que ajudem no retardo do avanço da doença, além de intervenções medicinais(MOVEMENT, 2017). De acordo com Liu, Cheng e Yan (2018), os tratamentos iniciados mais cedo apresentam melhores resultados e é considerado a melhor forma de combate à doença, ao passo que o Alzheimer, dado os conhecimentos clínicos obtidos até os dias atuais, é considerada uma doença que ocasiona danos irreversíveis ao indivíduo, impossibilitando bons resultados quando submetidos a tratamentos tardios.

Long et al. (2012) salienta que nas últimas décadas inúmeros procedimentos foram testados para o diagnóstico de pacientes com Alzheimer, como testes neurofisiológicos, auditivos e eletrofisiológicos. Nos dias atuais o diagnóstico é feito com base em critérios clínicos, no entanto esse método torna-se eficaz somente quando a doença já está em um estado avançado, tornando-se um impeditivo no progresso médico sobre a doença, ao passo que os danos nas regiões do cérebro afetadas pela DA não podem ser revertidos, e uma cura para o avanço da mesma não é esperada à curto prazo. Diagnosticar a doença o mais cedo possível para remediar seus danos no paciente é o principal caminho que médicos tem enfretado a DA.

.

1.2 Justificativa

Ainda que embora não possa ser utilizado como único e determinante método para o diagnóstico, as neuroimagens tornaram-se o foco de estudo principal nas últimas décadas, motivado por avanços na visualização de como a doença progride, os efeitos da DA no cérebro e sobre o funcionamento de remédios. Dentro dessa nova perspectiva, o exame de Ressonância Magnética aparece como uma promissora forma de entendimento, diagnóstico e análise do avanço dessa doença. O diagnóstico por imagens MRI mostra-se viável já que a atrofia de regiões no cérebro é uma das principais características do progresso do Alzheimer no paciente, sendo a causa de tais degenerações causadas por perdas de neurônios e dendritos no sistema nervoso do paciente. A atrofia em regiões como o Hipocampo é uma característica importante no

diagnóstico da doença e que pode ser visualizada pela MRI. (JOHNSON et al., 2012)

De acordo com Bron et al. (2015) o exame além de ser um bom indicador do progressso da doença, também pode ser usado para treinamento de classificadores inteligentes que buscam reconhecer diferenças de padrões entre pacientes com Alzheimer e indivíduos saudáveis para classificarem a qual grupo pertence o indivíduo pela imagem MRI. O uso de visão computacional no diagnóstico de doenças mostra também a capacidade para suprir em alguns hospitais a falta de expertise médica, acelerar o processo de diganóstico, permitindo que o paciente comece o tratamento em uma fase precoce de danos. Outro fator importante para a evolução dos estudos com classificadores inteligentes é de que os exames de MRI contam com uma alta disponiblidade de datasets online, e que tende a crescer ainda mais conforme a comunidade contribui (BRON et al., 2015).

1.2.1 Classificadores inteligentes no diagnóstico da DA

Com o avanço da Inteligência Artificial (IA) em uma sociedade que produz e armazena uma quantidade enorme de dados possibilitou transformar grande parte desses dados em informação, refletindo também no âmbito da medicina nas últimas décadas (LOBO, 2018). Logo surgiram diversas propostas de classificadores para o diagnóstico de doenças utilizando técnicas de aprendizado de máquina (LECUN; BENGIO; HINTON, 2015; SKANSI, 2018).

Classificadores inteligentes são capazes de identificar padrões entre os dados abstraindo informações conforme esses padrões são compartilhados na rede (no caso de um modelo de DL), sendo que uma análise qualitativa feita por seres humanos não são capazes de reconhecê-las. Quando o dataset dispõe de grupos etiquetados o aprendizado é chamado de supervisionado. Modelos de Deep Learning são os mais utilizados e que possuem maior eficiência na classificação de imagens, devido a complexidade dos padrões entre os pixels de uma imagem, além do maior volume de dados presentes nesses arquivos, sendo esses fatores normalmente considerados adequados para o uso das Redes Neurais, na qual a identificação de cada padrão é feita por camadas e atribuída ao conhecimento geral da rede, transformando os padrões nos dados em informação útil para a classificação. As Redes Neurais Convolucionais (CNN) é a principal arquitetura utilizada para esse tipo de aprendizado, em que sua estrutura e funcionamento são inspirados no funcionamento entre o sistema óptico e nervoso do ser humano (LECUN; BENGIO; HINTON, 2015; WEN et al., 2020).

Dentre inúmeras arquiteturas criadas, embora as redes neurais convolucionais tenham se mostrado um dos métodos de maior sucesso para classificação de imagens (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), principalmente suas arquiteturas mais conhecidas (e.g LeNet, AlexNet, VGG, Inception Net), existem diversos casos

no qual o modelo registra um melhor desempenho trabalhando junto com outras técnicas de DL para complementar seu aprendizado (KLINGELHOEFER; ROUSSEAU, 2017; HOSSEINI-ASL; KEYNTON; EL-BAZ, 2016). Além da escolha do algoritmo de treinamento, a performance de um modelo para classificação de imagens médicas e as comparações entre diferentes modelos podem variar com: a segmentação dos dados que serão aplicados para treinamento; Técnicas de pré processamento de imagens utilizadas; Escolha dos tipos de imagens e particularidades dos pacientes no dataset usado para o treinamento e teste; Métricas de desempenho utilizadas (WEN et al., 2020).

Além da escolha do algoritmo, existem técnicas essenciais para a otimização de modelos como o *Data Augmentation* e a Transferência de Aprendizado. Embora sejam técnicas corriqueiramente utilizadas em projetos de visão computacional, esses processos no entanto necessitam de uma avaliação mais aprofundada quando aplicados no contexto de imagens médicas devido ao predomínio de padrões mais sutis presentes nesse tipo de imagens, os quais precisam ser identificados pelo classificador quando aplicados ao diagnóstico de uma doença (HUSSAIN et al., 2018a). O *Data Augmentation* visa suprir limitações de datasets pequenos e evitar o overfitting do modelo por meio da criação de novas imagens a partir da aplicação de transformações de pré processamentos no conjunto original das imagens. Já o método de Trasnferência de Aprendizado se beneficia de uma rede pré treinada com outro dataset em que os pesos do algoritmos são pré computados e passado seu aprendizado para uma nova arquitetura, diminuindo assim o tempo e custo computacional, além de também otimizar o treinamento quando o dataset contém poucas imagens (LECUN; BENGIO; HINTON, 2015; HUSSAIN et al., 2018a).

De acordo com Wen et al. (2020) mais de 30 artigos foram publicados referentes a classificadores usando CNN no diagnóstico de Alzheimer por meio de imagens sMRI . No entanto, sem uma padronização na metodologia dos modelos construídos torna-se difícil comparar os resultados. Variações nos tipos de neuroimagens utilizadas, métricas de desempenho, critérios de diagnóstico, técnicas de extração de características, dentre outros, impedem uma avaliação justa e transparente sobre quais as técnicas mais promissoras para cada abordagem no diagnóstico, além do fato de que a maior parte dos trabalhos concentram-se na análise da arquitetura da rede e dão pouca interpretabilidade sobre a influência de técnicas comumente usadas, como o Data Augmentation e a Transferência de Aprendizado, na performance de modelos para classificação da DA (BRON et al., 2015; WEN et al., 2020).

No mesmo estudo realizado por Wen et al. (2020) fica claro também que a maior parte dos modelos já conseguem níveis altamente satisfatórios de performance mas todos concentrando na classificação de pacientes com Alzheimer/saudáveis, sendo

escasso projetos que buscam atender a urgência da área em identificar pacientes em estágio inicial. Ao passo que perdas maiores na região do hipocampo apresentadas em graus avançados são mais fáceis de identificar pelo modelo, a classificação de níveis inicias de demência tornam-se mais desafiadores para o diagnóstico (LIU; CHENG; YAN, 2018; BRON et al., 2015).

No trabalho de revisão literária de classificadores da DA feito por (BRON et al., 2015), é mostrada a ocorrência de uma concentração de estudos feitos com arquiteturas CNN clássicas como VGG, ResNet, Alexnet (e.g Sahaf; Thofighi, 2016) ou adaptações das mesmas (e.g Simonyam; Zisserman, 2014; Korolev et al, 2017; Shmulev et al, 2018). Singh et al. (2020) numa exploração acerca de abordagens de DL em imagens médicas levanta uma visão otimista acerca da performance de novas abordagens na classificação de pacientes infermos/saudáveis, citando o modelo GoogleNet recentemente criado pela Google como uma arquitetura de sucesso nesse quesito.

A literatura mostra uma escolha promissora para um modelo que pode traballhar bem com dados pesados: em 2017 foi criada por Huang, Liu e Maaten (2018) as Redes Convolucionais Densamente Conectadas (DenseNet) que fazem uso de conexões compartilhadas por toda rede para classificação de imagens, este conceito permitiu aos autores obterem desempenho similares e até melhores comparada à outras arquiteturas consagradas, no entanto a DenseNet mostrou que por meio das conexões Densamente Conectadas entre os blocos foi possível treinar com muito menos parâmetros, o que cria uma promissora perspectiva para essa abordagem em imagens MRI (HUANG; LIU; MAATEN, 2018; LECUN; BENGIO; HINTON, 2015).

1.3 Objetivo

Este trabalho tem o objetivo de introduzir uma nova abordagem de classificador (Redes Convolucinais Densamente Conectadas - DenseNet - de Huang et al. (2018), analisando a influência dos parâmetros da rede no desempenho do modelo. A opção por essa técnica visa inserir uma nova abordagem na arquitetura de classificadores que contribuam no diagnóstico da Doença de Alzheimer em estágio inicial utilizando imagens MRI, com a análise focada no desempenho do classificador para diagnosticar pacientes em três grupos: saudáveis (CN), com Comprometimento cognitivo leve em estágio intermediário (MCI) e com Comprometimento Cognitivo Leve em estágio inicial (VMCI), sendo o MCI um estado inicial/intermediário da perda de memória que pode ou não se tornar um Alzheimer futuramente. Os testes focarão na avaliação de como as variáveis growth rate, profundidade de layers e uso de Transfer Learning do DenseNet afetam no desempenho do modelo, assim como a utilização de Data Augmentation e o

treinamento com dataset balanceado. O projeto também visa padronizar as etapas de pré processamento afim de permitir comparações mais justas e interpretável para a influência dos diferentes parâmetros na performances do modelo.

1.4 Objetivos Específicos

Introduzir uma arquitetura moderna de CNN no contexto de classificação de Alzheimer por imagens MRI

Focar na identificação de estágios inicias de perda de memória.

Comparação da perfomance entre o modelo utilizado no trabalho com projetos anteriores.

Avaliar a influência das técnicas de Data Augmentation e Transfer Learning no desempenho do modelo.

Avaliação do desempenho do modelo com base na mudança dos parâmetros growth rate e profundidade da rede classificadora.

Possibilitar à comunidade interessada na área um acesso mais fácil ao desenvolvimento do trabalho, com uma abordagem detalhada sobre as etapas de implementação e por meio da utilização de um dataset público, afim de facilitar comparações com outros trabalhos futuros.

2 Ressonância Magnética - MRI

Para compreender o funcionamento da imagem de Ressonância Magnética é necessário compreender sobre a Ressonância Magnética Nuclear (RMN), um fenômeno físico que compreende toda a base do funcionamento contido na coleta das imagens MRI. Em 1924, Wolfran Paule sugeriu que as partículas possuíam um momento angular, o qual seria um atributo correspondente a campos magnéticos e que hoje conhecemos como spin. Dando prosseguimento a linha de conhecimento sobre o assunto, em 1938 Rabi realizou um experimento capaz de medir o deslocamento do momento magnético de uma partícula e que usou pela primeira vez o nome de Ressonância Magnética Nuclear. Os mais importantes passos dados no estudo desse fenômeno deu-se na década de 40. Com experimentos realizados nessa época, passou a ser possível detectar a mudança no eixo de rotação de núcleos magnéticos dentro de um campo magnético, peça chave no entendimento da RMN. Os trabalhos dessa época que mais tiveram impacto no estudo da RMN foram os realizados por Bloch e Purcell, os quais permitiram por meio da ressonância registrar as propriedades dos núcleos atômicos (HAYDEN; NACHER, 2015; PATRALEKH; KALRA, 2012).

Com o experimento realizado por Purcell, Torrey e Pound, todos do Laboratório de Radiação do *Massachusetts Institute of Technology* (MIT), foi possível observar uma variação na amplitude do sinal de radiofrequência dentro de uma cavidade sintonizada para ressoar em 30 MHz. Ao preencherem-na com 1 litro de parafina sólida, foi constatado uma mudança de 0,4% na amplitude do sinal de Radiofrequência conforme o campo magnético que era estático foi varrido por uma ressonância intensa. Bloch constatou o mesmo fenômeno ao realizar um experimento similar utilizando 1,5 cm^3 de água com duas bobinas ortogonais. Na época esses experimentos não puderam ser explicados com exatidão sobre qual a causa que levava à essa variação no sinal RF, mas a constatação desse fenômeno foi um enorme passo para que estudos futuros pudessem nos levar ao uso da Ressonância Magnética dos tempos atuais (HAYDEN; NACHER, 2015).

2.1 A física da Ressonância Magnética

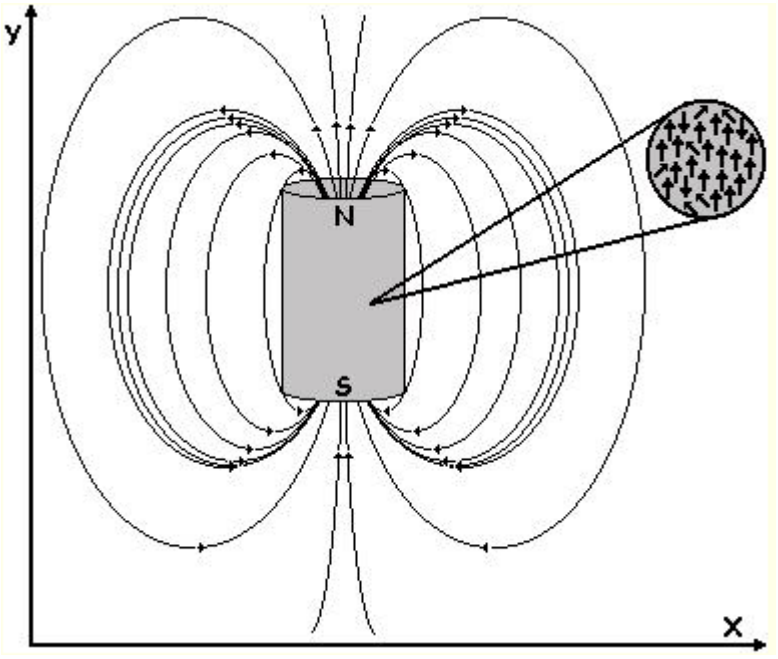
Compreender alguns conceitos de física básica é fundamental para entender como as imagens MRI são geradas. Assim como outras imagens médicas ,a Ressonância Magnética tem seu comportamento explicado em escala atômica, mas diferentemente de grande parte das outras técnicas a explicação do seu comportamento não tem foco nos elétrons, mas nos prótons, que são pequenas partículas de carga elétrica

positiva situadas no núcleo de um átomo. O número de prótons no núcleo varia entre cada elemento, no caso do átomo de Hidrogênio que é figura central na fundamentação teórica da MRI, seu núcleo é constituído por um único próton. O foco do estudo sobre o Hidrogênio no funcionamento da Ressonancia Magnética é justificado pelo fato de que o corpo humano é constituído predominantemente por água, ocasionando que haja uma abundância de átomos de Hidrogênio em nosso organismo (RIOS, 1998; HAYDEN; NACHER, 2015).

O experimento de Rabi citado anteriormente, contribuiu para a identificação do Spin, um atributo chave do comportamento das partículas atômicas. Os prótons e neutrons que constituem o núcleo dos átomos exercem constantemente um movimento em torno de seu eixo, em outras palavras, essas partículas possuem um movimento angular intríseco. Pela Lei de Faraday da Indução é constatado que cargas elétricas em movimento geram uma corrente elétrica e essa corrente cria ao seu redor um campo magnético, com tal conhecimento foi possível constatar que devido ao movimento em torno de seu próprio eixo, essas partículas produzem um momento magnético na direção do eixo do Spin (HAYDEN; NACHER, 2015; RIOS, 1998).

[1]

Figura 1 – Campo magnético gerado por uma carga elétrica em movimento



Fonte: UNESP/Bauru (2009)

Uma outra definição atribuída ao próton é de um dipolo magnético (DM), pois um DM além de produzir um campo magnético ele também responde a campos magnéticos gerados por outras fontes (FERRARINI; HAGE; IWASAKI, 2009). A resposta do próton a um campo magnético externo é semelhante a um imã, quando um átomo está sob

[1] Disponível em: <http://www2.fc.unesp.br/experimentosdefisica/ele13.htm>. Acesso em: 13/11/2020.

a presença de um campo magnético, os prótons irão alinhar-se de forma expontânea ao longo das linhas de força desse campo. Uma peculiaridade do próton é de que a orientação de seus momentos magnéticos podem estar em duas classes diferentes: paralela ou antiparalela ao campo magnético externo. Procurando a orientação com menor gasto de energia, a maior parte dos prótons irão orientar-se paralelamente ao campo(RIOS, 1998). De acordo com Rios (1998) é possível obter uma relação entre a quantidade de prótons paralelas e antiparalelas à direção do campo magnético externo: Para cada 10^N prótons na direção antiparalela, haverão $10^N + N$ paralelos. Sendo essa quantidade N de prótons paralelos livres que serão essenciais para o funcionamento físico da Ressonância Magnética (FERRARINI; HAGE; IWASAKI, 2009; RIOS, 1998).

Em linhas práticas, um paciente ao executar o exame, transforma-se numa espécie de imã, em que seu próprio corpo cria um campo magnético. A RM produz um sinal pela corrente elétrica induzida pelo momento magnético do próton, que age sobre uma bobina receptora. No entanto, o momento magnético de um próton possui intensidade insuficiente para ser detectável pelo corpo, para isso eles precisam estar alinhados. Os prótons no corpo orientam-se de forma aleatória, seus momentos se cancelam ao apontarem para todas as direções. Para resolver esse problema, é gerada a emissão de um pulso de onda de rádiofrequência sobre o paciente que está submetido a um campo magnético externo, tal onda perturba os prótons fazendo com que suas orientações sejam uniformizadas em três possíveis sentidos: no mesmo sentido do campo externo, no sentido de seu vetor de momentos magnéticos ou no sentido contrário. Como descrito anteriormente, o próton irá buscar pelo sentido de menor energia fazendo com que ocorra uma predominância de N prótons no mesmo sentido do campo magnético externo. Essa pequena quantidade N de prótons criará uma pequena Magnetização Resultante de Equilíbrio (MRE) no tecido do corpo, permitindo magnetizar o tecido do paciente, fator fundamental para a futura formação da imagem(FERRARINI; HAGE; IWASAKI, 2009; RIOS, 1998).

Uma constatação importante é que quando os spins estão alinhados na direção do campo externo - também chamado de eixo z - os prótons referentes a eles não se alinham de forma totalmente precisa sobre esse eixo, mas realizam movimento chamado de precessão, em que os prótons giram em torno do seu eixo gravitacional, um movimento parecido com o feito por um peão.

A Magnetização Resultante de Equilíbrio ou Magnetização Tissular intríseca citada anteriormente possui um valor muito baixo sendo praticamente impossível medi-lo quando alinhado ao Campo Magnético devido a desproporcionalidade escalar, mas quando a MO é desviada para o plano xy a mensuração e consequentemente a formação da imagem torna-se possível. O desvio é feito pelo pulso de radiofrequência, a sua inserção no sistema também contribui para a mudança de fase dos prótons de

modo a agrupá-los, isso é fundamental para que as bobinas receptoras detectem o sinal. Como dito anteriormente a onda RF também excita os prótons do núcleo com energia, quando esses por fim retornam ao estado de equilíbrio liberando energia, tal processo é chamado de relaxação e dá origem a duas classes de imagem: spin-spin, ou T2 e spin-lattice, ou T1, essa última será a classe abordada pelo classificador.

A definição da variável T1 depende do entendimento do processo após a inserção da onda RF: quando a manetização tissular é desviada do eixo z, parte do valor vertical de MO é perdido, sendo que o tempo que o núcleo do átomo leva para retornar ao valor de 63,2% de MO original é definido como o T1. Portanto, T1 é mutável com relação ao tipo de molécula que está sendo abordada no sistema. O comportamento de T1 que originará o grau de intensidade da imagem nos trazendo o significado intríseco que ela representa, é explicado da seguinte forma: (RIOS, 1998; FERRARINI; HAGE; IWASAKI, 2009)

> Quando um tecido com T1 curto é examinado usando uma seqüência com um tempo de repetição (TR) do pulso de RF de 90 graus relativamente mais longo, o sinal oriundo desse tecido é intenso. Se o tempo de repetição (TR) for mais curto do que o T1 do tecido, o núcleo não retornará ao equilíbrio antes do próximo pulso de RF, e o tecido é dito como estando saturado (sem sinal). Dessa forma, a intensidade do sinal aumenta à medida que o tempo de relaxação do tecido diminui.(FERRARINI; HAGE; IWASAKI, 2009, 1290)

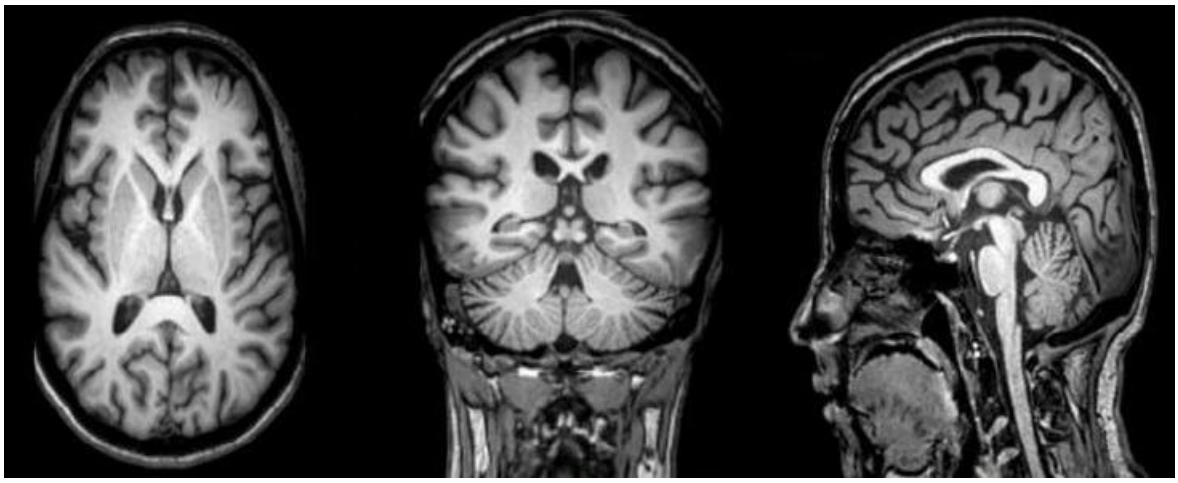
Com a ocorrência da relaxação dos prótons, é chegado ao ponto de ressonância, em que a frequência do pulso RF está na mesma frequência dos prótons, tal ponto ocasiona a precessão em fase dos prótons, o qual gera um sinal magnético que pode ser medido. A imagem final é obtida por um processo de *scanner* que varre o sinal resultante em cada ponto do pedaço do corpo a ser examinado. Para conseguir obter esse sinal é preciso que essa varredura seja feita com gradientes sucessivos em três dimensões . O sinal monitorado é a soma dos sinais de cada únicovelemento de volume dentro do corpo. Desta forma, o instrumento recebe milhares ou até mesmo milhões de dados, criando um conjunto de equações a partir das quais a magnitude do sinal de cada elemento de volume pode ser calculada (WRIGHT et al., 2008).

A reconstrução da imagem só é possível graças a um processo de análise de um sinal com frequência e fase específica, feita por um computador por meio da transformada de Fourier, permitindo a obtenção um espectro de frequência de um sinal composto. O espectro das frequências é a representação gráfica da amplitude e das frequências dos sinais primitivos usados para gerar o sinal composto. A amplitude do sinal codifica a quantidade dos prótons ressonantes, enquanto a frequência de cada onda permite que a posição de cada amostra possa ser conhecida. A reconstrução pode ser feita então, a partir da correlação desse sinal com um ponto definido no volume (RIOS, 1998; FERRARINI; HAGE; IWASAKI, 2009).

As imagens MRI são adquiridas em um formato digital denominado DICOM, o qual além de armazenar a imagem, também pode salvar as informações do paciente relacionadas à ela. Com o intuito de facilitar o armazenamento e a comunicação de diagnósticos médicos, o formato DICOM pode ser definido como um conjunto de normas que permitiu unificar o formato de exames de diagnóstico, sendo também utilizados para outros exames além da MRI, como a TC (WRIGHT et al., 2008).

[2]

Figura 2 – Exemplo de uma imagem MRI



Fonte: Padmanaban et al. (2020)

2.1.1 MRI no contexto atual de Alzheimer

Na Doença de Alzheimer células neurológicas relacionadas à memória são mortas, gerando uma perda de volume no Hipocampo (principal região afetada pela DA), a Ressonância Magnética Estrutural é capaz de fornecer a visualização e a medição volumétrica de diferentes regiões do cérebro, como o Hipocampo que é a região mais afetada pela doença. Em tais regiões em um scanner T1 o tempo de relaxação é menor do que áreas não afetadas, ao passo que células neuronais estão mortas nessas áreas (BARROS, 2017).

A Ressonância Magnética tem sido uma das técnicas preferidas por médicos e radiologistas para a visualização de órgãos, graças ao fato de ser um método não invasivo e por não utilizar material radioativo como a Tomografia Computadorizada (TC). As estimativas estruturais de MRI de dano ou perda de tecido em regiões cerebrais caracteristicamente vulneráveis, como o hipocampo e o córtex entorrinal, são preditivas de progressão de MCI para DA. Além disso, foi estabelecida a utilidade clínica da MRI na diferenciação entre a DA e outras patologias, como a neurodegeneração vascular, ou não-Alzheimer (KLINGELHOEFER; ROUSSEAU, 2017; FRISONI et al., 2010).

Os scanners atuais de ressonância magnética permitem a visualização de imagens do cérebro de altíssima resolução, com detalhes estruturais refinados, excelente contraste de tecido e resolução espacial de até 1 mm. Escalas de avaliação visual permitem identificar atrofia medial estrutural do lobo temporal categorizadas em níveis discretos. Estruturas com limites definidos podem ser rotuladas por esboço manual e os volumes podem ser computados. Todas essas capacidades mostram a esperança que a comunidade médica-tecnológica tem depositado na utilização do exame de Ressonância Magnética como classificador no diagnóstico da DA (FRISONI et al., 2010).

De acordo com Frisoni et al. (2010), com as novas descobertas sobre o comportamento das proteínas tau e A no progresso do Alzheimer, criou uma perspectiva positiva na criação de drogas que retardem ou previnam a progressão da doença, tornando ainda mais necessário o diagnóstico da doença em fases iniciais, como o Comprometimento Cognitivo Leve , ou do inglês *Mild Cognitive Impairment* (MCI), um estágio em que o paciente começa a ter déficits cognitivos e que pode ou não vir a progredir para Alzheimer. No mesmo estudo, o autor mostra alguns pontos chaves do exame de Ressonância Magnética Estrutural como potencial catalisador desse diagnóstico precoce:

- A atrofia cerebral detectada pelos exames MRI de alta resolução está correlacionada tanto com os depósitos de proteínas tau como déficits neuropsicológicos, sendo hoje já considerado um marcador válido no diagnóstico de Alzheimer e em sua progressão
- O grau de atrofia das estruturas temporais mediais, como o hipocampo, também é um marcador diagnóstico para AD no estágio de MCI.
- As taxas de atrofia do cérebro inteiro e do hipocampo são marcadores sensíveis de progressão de neurodegeneração, e são cada vez mais usados como resultados substitutos em ensaios de drogas potencialmente modificadoras da doença.

3 Redes Neurais e Deep Learning

3.1 Sistema Nervoso

A partir do trabalho de Camillo Golgi em 1875, que possibilitou a visualização de neurônios de forma isolada e individual, levou Santiago Ramón y Cajal a introduzir o conceito de sistema nervoso, postulando sobre a comunicação entre as células por meio da região chamada de sinapse. O estudo da conexão entre os neurônios feito por Cajal, possibilitou a consolidação em estudos subsequentes do conceito de rede neural.

O neurônio é constituído por um corpo celular chamado de *soma*, e com diversas ramificações denominadas *dendritos*, as quais conduzem os sinais das extremidades do neurônio para seu corpo celular. Dentre as ramificações existentes, uma delas é conhecida como axônio, e que faz a função inversa das outras ramificações: transmitir sinais do corpo celular para as extremidades. Por meio da sinapse é possível conectar o axônio com dendritos de outros neurônios, ou até mesmo diretamente com outro axônio ou corpo celular de outro neurônio.

As sinapses possuem papel fundamental no sistema nervoso, pois permitem além da comunicação de sinais neuronais, a memorização da informação. De acordo com Barreto (2002), em cada sinapse, a quantidade de neurotransmissores que podem ser liberados para uma mesma frequência de pulsos do axônio representa a informação armazenada nesta sinapse. Conforme cada ativação de sinapse que ocorre e encontra ou consegue ativar outro neurônio, o número de neurotransmissores liberados aumenta na próxima vez que o neurônio for ativado (BARRETO, 2002). Esse processo conhecido como facilitação, permitiu a postulação da conhecida Lei de Hebb, a qual inspirou a base conceitual de muitos algoritmos de redes neurais artificiais: *"A intensidade de uma conexão sináptica entre dois neurônios aumenta quando os dois neurônios estão excitados simultaneamente."*.

3.2 Perceptrons

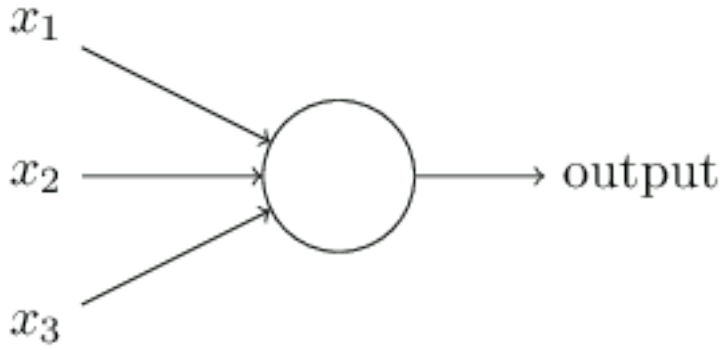
Baseando-se no comportamento dos neurônios biológicos em nosso sistema nervoso, as redes neurais artificiais (RNA) tem sido uma das principais técnicas de aprendizado de máquina nos últimos anos devido a evolução do hardware, principalmente do desenvolvimento das unidades de processamento gráfico, do inglês *Graphics Processing Unit* (GPU), que possibilitou a construção de modelos mais complexos e

que demandam maior poder computacional do que outras técnicas de aprendizado de máquina com menos parâmetros e menor complexibilidade (KLINGELHOEFER; ROUSSEAU, 2017).

Os primeiros passos na construção de um algoritmo de Rede Neural Artificial foi dado por Frank Rosenblatt, o qual introduziu a primeira e mais simples arquitetura de uma RNA, a qual foi chamada de Perceptron. Essa estrutura foi inspirada no trabalho de McCulloch e Pitts (1943) e fuciona da seguinte forma: um nó recebe um número *n* de entradas, que são computadas na única camada da rede para gerar uma saída binária com valor 0 ou 1. A estrutura dessa rede pode ser vista na Figura 3. A computação do valor é definida então com base nos chamados pesos *(w_i)*, de cada conexão. Essas variáveis representam a importância que cada entrada tem para a computação da saída. Com base nesses valores, o nó irá calcular a saída que será condicionada em um valor de threshold, com o valor calculado pelo somatório dos produtos de cada peso w_i com sua respectiva entrada x_i (NIELSEN, 2018; BISHOP, 1994). Para um dado número *N* de entradas temos:

$$f(x) = \begin{cases} 1 & \text{, se } \sum_{i=0}^{N-1} x_i\, w_i \leq \textit{threshold} \\ 0 & \text{, se } \sum_{i=0}^{i=N} \mathrm{x}_i\, w_i \geq \textit{threshold} \end{cases} \tag{3.1}$$

Figura 3 – Perceptron com N = 3



[1]

Fonte: Deep Learning Book (2021).

Os pesos **W**, são peça chave na construção de um Perceptron ou uma RNA, o foco do aprendizado está em descobrir quais os melhores valores de cada peso para se obter a saida mais próxima da desejada. Apesar de notóriamente importante na construção básica do conhecimento de Deep Learning, os perceptrons são limitados na resolução de diversos problemas do mundo real, em que são exigidos o reconhecimento de padrões mais complexos ou problemas que a saída precisa ser contínua

[1] Disponível em: <https://www.deeplearningbook.com.br/o-perceptron-parte-1/>. Acesso em: 15/12/2020.

e não binária. Dada essas limitações mas baseando-se na importância conceitual da introdução de perceptrons, diversas arquiteturas foram aprimoradas para chegarmos aos modelos presentes nos dias atuais (NIELSEN, 2018).

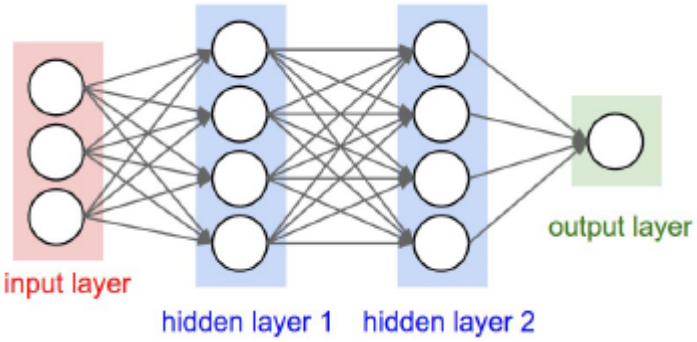
3.3 Conceitos básicos das arquiteturas modernas

O funcionamento das redes neurais naturais está baseado na conexão entre milhões de neurônios por meio das chamadas sinapses, constituindo um sistema dinâmico que permite a comunicação entre diferentes partes do cérebro, ou entre o cérebro e diferentes partes do corpo. De forma análoga, uma RNA também é um sistema composto por vários neurônios e que atua dinamicamente no tráfego de informação entre as estruturas que constituem a rede. Um neurônio de uma RNA pode se comunicar com o exterior, recebendo excitações vindas de fora da rede, sendo essas estruturas conhecidas como neurônios de entrada, e que atuam de forma similar aos neurônios dos órgãos do sentido presentes em um sistema nervoso natural. Já os chamados neurônios de saída, aplicam excitações no mundo exterior afim de modificá-los de alguma forma, essas respostas em forma de excitação é baseada no conteúdo da informação presente e compartilhada na rede, tais estruturas atuam como os neurônios biológicos que excitam o músculo. A classe restante de estruturas conhecidas como neurônios internos, formam as chamadas camadas intermediárias, as quais podem interconectar centenas de neurônios, os quais trafegam a informação distribuindo-a por toda a rede (BISHOP, 1994; BARRETO, 2002).

Conforme a informação circula na rede, o conhecimento é compartilhado e possibilita, com a agregação das características aprendidas em cada camada, que se chegue a um reconhecimento de padrões de mais alto nível de complexidade nas camadas finais, permitindo uma tomada de decisão com base no valor dado pelo neurônio de saída. É dessa perspectiva que as arquiteturas foram incorporando cada vez mais neurônios intermediários e aprofundando a rede, daí o nome de aprendizado profundo, ou do inglês *Deep Learning*. Muitas dessas arquiteturas só foram possíveis de implementá-las na prática graças ao desenvolvimento computacional nas décadas mais recentes, em que o computador passou a ser capaz de realizar tarefas paralelamente, e os processadores gráficos, peça essencial na computação de cálculos tensoriais, tornaram-se mais potentes e viáveis economicamente. De forma geral, quanto maior o número de camadas escondidas, e consequentemente mais profunda torna-se a rede, possibilitando o conhecimento de padrões mais complexos, mais também é exigido do poder computacional e maior o tempo de treinamento da rede. (LECUN; BENGIO; HINTON, 2015; BISHOP, 1994)

Cada neurônio da rede está conectado com outro por meio de pesos atribuídos

Figura 4 – Arquitetura de uma RNA



[2]

Fonte: Página CS231n.

a essa conexão, assim como visto nos perceptrons (LECUN; BENGIO; HINTON, 2015). Outros parâmetros importantes presentes em uma Rede Neural além dos pesos de cada nó, e que também são dados para cada conexão entre neurônios, são os chamados *bias* (ou viés) e as funções de ativação. A atribuição dos *bias* com os pesos, são aplicados aos dados para obter uma combinação linear de saída, dada por $ax + b$. Tal função calcula o valor de saída de cada neurônio. O *bias* recebe seus valores de outros neurônios, ou de uma fonte externa (dados de um dataset) caso o neurônio seja de entrada, e então é incorporado visando flexibilizar o ajuste da rede permitindo o aprendizado em partes invariantes da RNA. Diferentes *bias* são atribuídos a cada nó da rede, sofrendo influência do peso atribuído a este dado nó (PONTI; COSTA, 2017). O termo *bias*, junto com seu respectivo peso, produz uma combinação linear (soma ponderada de suas entradas) em que a saída do neurônio é dada pelo resultado de uma função f dessa combinação linear..

$$f(\sum_{i=0}^{N-1} w_i \times x_i) \tag{3.2}$$

A equação (3.2) demonstra a saída de cada neurônio, sendo os dados de entrada do nó representados por **x**, e **w** os pesos atribuídos a esse nó. A função representada nessa equação é a chamada função de ativação, a qual é aplicada na saída do nó para gerar determinado tipo de saída desejada. De uma forma geral, as funções de ativações podem ser dividas entre lineares e não lineares, sendo esse último grupo o mais utilizado.

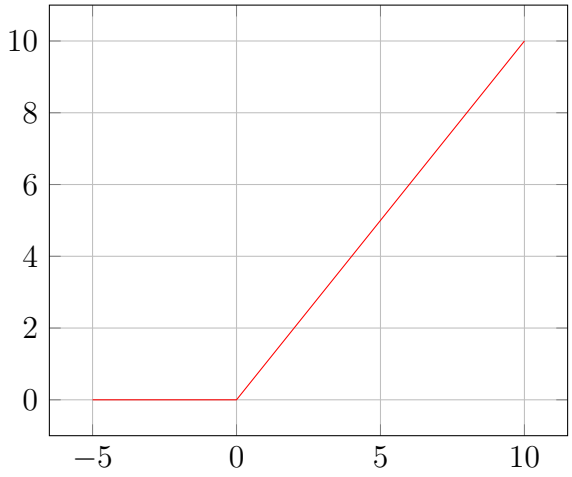
A ativação linear retificada, do inglês *Retified Linear Unit* (ReLU), é normalmente utilizada nos nós intermediários devido a facilidade para o treinamento dessa função, principalmente em um modelo com muitos nós, que exigirá custo computacional e consequentemente mais tempo para o treinamento (LAI, 2015; AGGARWAL, 2018). A

função ReLU está representada na equação 3.4

$$\phi(v) = max(0, v) \tag{3.3}$$

A função ReLU produz uma saída linear para número maiores que zero, como pode ser visto no gráfico da figura 4.

Figura 5 – Função ReLU



Fonte: O autor.

As funções de ativação não são incorporadas para resolver problemas complexos a partir dela, mas para definir as relações entre os varios nós da rede. Como o próprio nome diz ela decidirá se irá ativar ou não cada conexão, por meio da atribuição de valores positivos ou negativos/zerados na saída do neurônio. Há também um grupo de funções de ativação que são responsáveis por decidirem a resposta final do modelo, tais funções são utilizadas na última camada da rede e variam conforme o tipo de problema. Para problemas de classificação binária as funções de ativação são baseadas na função *Sigmoide*. Assumindo valores entre 0 e 1, essa função é utilizada quando é desejado obter um valor de saída que represente uma probabilidade, com base em um dado valor de threshold será escolhido o valor representativo de uma das classes assumida na classificação.

Em muitos problemas é desejado uma classificação que possui mais do que duas classes possíveis, chamada também de classificação categórica, a função de ativação mais utilizada para esses casos é a *Softmax*, a qual atua na última camada, que possui o número de neurônios de saída dado por N, o qual representa o número de classes do problema. A função *Softmax* foi criada como uma extensão da *Sigmoide* quando a rede está trabalhando com um problema de classificação multiclasse, a aplicação dessa função gera uma saída com valores decimais que podem ser interpretados como as probabilidades do exemplo dado ser referente a cada classe do problema de

acordo com o classificador. A função *Softmax* é definida pela equação (3.3), em que $\vec{z}$ representa o vetor de entrada, e K o número de classes. (BISHOP, 1994; LECUN; BENGIO; HINTON, 2015).

$$\sigma(\vec{z})_i = \frac{e^i}{\sum_{j=1}^{K} e^{z_j}} \tag{3.4}$$

3.3.1 Gradiente Descendente

O grande responsável pelo potencial do aprendizado em Redes Neurais profundas está no fato de que seu treinamento é baseado nos próprios dados de entrada da rede. Uma RNA promove seu aprendizado por meio da atualização de seus parâmetros (pesos, *bias*) através de um algoritmo de otimização, até que se encontre os valores considerados ótimos, tendo como referencial atingir a menor diferença entre os valores previstos pelo modelo (saída da rede) e os valores reais. A métrica escolhida para calcular essa diferença é chamada de função de custo, e tem papel fundamental na avaliação da performance do modelo, pois é a função de custo que mensura a capacidade de previsão da rede, comparando a saída da camada final com a resposta correta quando um certo dado de entrada alimenta a rede (JANOCHA; CZARNECKI, 2017).

A escolha da função de custo é um parâmetro a ser escolhido pelo autor, e que de acordo com Patterson e Gibson (2017), a escolha dessa função depende de variáveis como o tipo de problema a ser resolvido, o algoritmo de aprendizado de máquina que está sendo utilizado, a facilidade de calcular as derivadas e a porcentagem de outliers presente no conjunto de dados. De uma forma geral as funções de custo podem ser divididas em dois grupos de problema: funções de custo de regressão e funções de custo de classificação, sendo esse último o conjunto de interesse para um problema de classificação como o abordado neste trabalho. A função de erro de Entropia Cruzada tem sido a principal métrica para problemas de classificação, devido ao fato de se adaptarem bem tanto para problemas de classificação binária como também para multiclasses, além de permitir que erros na previsão do modelo alterem os pesos da rede mesmo quando houver nós saturados (quando a derivada já está próxima de 0). A Entropia Cruzada é de forma geral usada para quantificar a diferença entre duas distribuições de probabilidade, quando aplicada a um problema de classificação essa métrica tende a aumentar o seu valor conforme as previsões estejam mais distantes da resposta correta. Considerando um dado exemplo do conjunto de dados, sendo *y* o valor correto desse dado exemplo, e $\hat{y}$ o valor previsto pela rede nesse mesmo exemplo, a função de erro de Entropia Cruzada de um problema de multiclassificação, para um dado conjunto de dados de tamanho *K*, em que *k* indica qual a classe correta desse dado exemplo, pode ser vista na equação (3.5) (PATTERSON; GIBSON, 2017;

JANOCHA; CZARNECKI, 2017).

$$L(y, \hat{y}) = -\sum_{k}^{K} y^{(k)} \log \hat{y}^{(k)} \tag{3.5}$$

Focando na minimização da função de custo, os parâmetros da rede são modificados por meio de uma função de otimização, a qual busca atualizar os pesos da rede afim de achar a combinação que diminua ao máximo possível, a diferença entre as previsões da rede e os valores reais separados para teste. O ponto onde a função de custo assume o menor valor possível é chamado de ponto de convergência (PATTERSON; GIBSON, 2017).

A função de otimização mais utilizada no aprendizado profundo e que serviu como base de muitas outras funções de otimização é a chamada Gradiente Descendente (GD). Muitas vezes utilizada junto com a técnica de retropropagação, o GD inicia os pesos da rede de forma aleatória, permitindo que seja computado a função de custo da rede. Sendo orientada por essa função, o Gradiente Descendente atualiza os parâmetros na rede afim de encontrar a direção das regiões de diminuição da função de custo, o mais rápido possível. Supondo que a função de custo seja uma função convexa, o funcionamento do GD pode ser explicado de uma forma visual, e intuitiva: como uma função convexa, a função de custo pode ser imaginado como uma tigela em que o único ponto mínimo é chamado de mínimo global, o gradiente dessa função apontará a direção de descida mais íngreme a partir do estado inicial, para que seja possível chegar no fundo da tigela, o local de menor erro das previsões da rede (PATTERSON; GIBSON, 2017).

O Gradiente Descendente atualiza simultaneamente os parâmetros da rede (e), subtraindo dos mesmos as derivadas parciais da função de custo para cada respectivo parâmetro. A derivada informa a taxa de variação de uma função, nesse caso, ela é responsável por mensurar a diferença dos resultados da função custo conforme a atualização dos parâmetros, permitindo ao algoritmo identificar a direção correta ao mínimo global,valor em que a derivada é igual a 0 (PATTERSON; GIBSON, 2017; LECUN; BENGIO; HINTON, 2015). A atualização dos parâmetros está representado naś equações 3.6a e 3.6b.

$$\hat{b} := \hat{b} - \alpha \frac{\partial}{\partial \hat{b}} L(\hat{b}, \hat{w}) \tag{3.6a}$$

$$\hat{w} = \hat{w} - \alpha \frac{\partial}{\partial \hat{b}} L(\hat{b}, \hat{w}) \tag{3.6b}$$

O processo de atualização dos parâmetros na direção de menor erro é repetido até que tais parâmetros cheguem a um ponto no qual a função de custo não pode

mais ser minimizada. Em funções convexas há apenas um único ponto de mínimo global, no entanto, para muitas funções ocorre a existência de vários vales na função de custo, chamados de mínimos locais, os quais podem induzir erroneamente ao algoritmo, a chegada do valor mínimo da função. O problema de mínimos locais pode ser superado por meio da hiperparametrização do algoritmo, como a mudança na taxa de aprendizagem (PATTERSON; GIBSON, 2017).

O funcionamento até aqui descrito das redes neurais artificias são conceitos base das arquiteturas atuais sendo essas aperfeiçoadas ao longos dos anos para resolver os mais diversos tipos de problemas, dando origem a novas arquiteturas e abordagens mais modernas, das quais três serão abordadas nesse trabalho, aplicando-as para construir um modelo de classificação da Doença de Alzheimer por imagens MRI(LECUN; BENGIO; HINTON, 2015).

3.4 Redes Neurais Convolucionais

Os modelos de Redes Neurais Convolucionais, do inglês *Convolutional Neural Networks* (CNN), foram motivados a partir de limitações dos algoritmos de Redes Neurais Densas para algumas tarefas de classificação de imagens, além do seu seu potencial para uso em imagens gerais (KLINGELHOEFER; ROUSSEAU, 2017). Com o avanço da inteligência artificial (IA) na medicina, a arquitetura também ganhou espaço na classificação de imagens médicas. De acordo com Ravì et al. (2016), a CNN foi a principal arquitetura de *Deep Learning* na Informática Médica desde 2010.

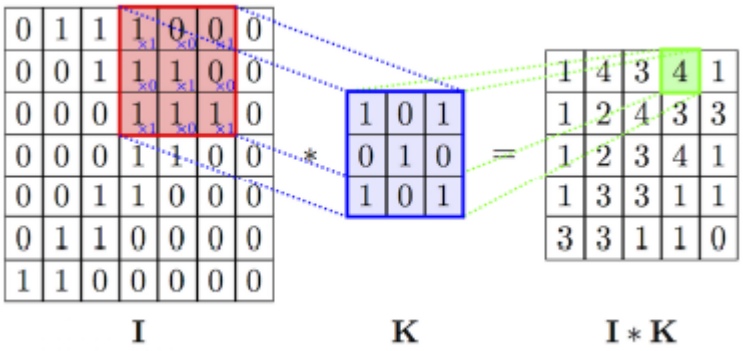
Com uma arquitetura desenhada para receber entrandas na forma de vetores, as CNNs trabalham muito bem com sinais naturais como: espectogramas de rádio e imagens que são processadas como vetores bidimensionais, ou vídeos e imagens volumétricas que são vetores tridimensionais, como é o caso das imagens resultantes de um exame de Ressonância Magnética. A arquitetura de uma Rede Neural Convolucional conta com uma série de camadas específicas que são integrados a rede conforme a necessidade que o problema, ao qual o autor da arquitetura está abordando, exige. O estágio inicial da CNN começa com as camadas denominadas *Pool*, e camadas convolucionais, operação principal da arquitetura, e que dão origem ao nome da classe de redes. (LECUN; BENGIO; HINTON, 2015)

Uma camada convolucional possui as mesmas dimensões dos seus dados de entrada, e seus neurônios são separados por um mapa de características que são compartilhadas entre as camadas da rede por meio dos pesos dos nós. Cada respectivo pixel da imagem de entrada é ativado por um conjunto de pixels ao redor que estão

[3] Disponível em: <https://cambridgespark.com/content/tutorials/convolutionalneural-networks-with-keras/index.html>. Acesso em: 13/05/2020.

3

Figura 6 – Processo de convolução aplicado por um kernel 3x3.



Fonte: Cambridge Spark (2020)

posicionados na mesma região dos dados de entrada. Esse processo é chamado de convolução, e percorre todos os pixels da imagem por um determinado *kernel* (Figura 5), uma matriz utilizada como filtro do processo e que permite ao neurônio aprender certo mapa de características da região em específico. O processo de convolução entre dois sinais, descrito na equação 3.7, junto com o compartilhamento dos pesos entre as camadas da rede, permitem que as características sejam aprendidas em níveis graduais de complexidade e que diferentes padrões na imagem sejam reconhecidos ao longo da rede.

A operação de convolução consistirá na varredura da imagem de entrada, por meio do *kernel* de tamanho a ser escolhido pelo autor, assim como o tamanho do passo de varredura. Durante o percurso será calculado o produto escalar entre o *kernel* e a imagem de entrada, gerando um valor escalar na saída do mapa de características. A varredura é feita até que toda a imagem tenha sido percorrida. A operação matemática da convolução pode ser vista na equação 3.7, sendo o resultado da convolução s[t], também chamado de mapa de características, é obtido no contexto da CNN aplicando a operação entre um *kernel* representado por **w**, e uma imagem de entrada **x** (GHOSH et al., 2020).

Além da dimensão do filtro (*kernel*) e o passo dado na convolução, o autor da rede também define a variável chamada *padding*, a qual refere-se ao preenchimento de bordas laterais na imagem, adicionadas ou retiradas do processo de convolução. A escolha consiste em duas possíveis opções: *padding* igual, em que a saída da convolução terá as mesmas dimensões que foi utilizado na entrada, ou zero *pad*, em que é adicionada uma nova colunas na borda da imagem, a qual é preechida com valores iguais à 0. Esse hiperparâmetro auxilia no controle de dimensão dos dados treinados para que eles não diminuam de forma acelerada durante o aprendizado, além de dar maior importância as informações presentes nas bordas da imagem, pois

com a adição do *padding* evita-se que as informações laterais da imagem acabem perdendo relevância devido a diminuição de dimensionalidade durante as consecutivas convoluções (ALVES, 2018; GHOSH et al., 2020).

$$s[t] = (x * w)[t] = \sum_{a=-\infty}^{\infty} x[a]w[a+t] \tag{3.7}$$

As camadas *Pooling* realizam a técnica chamada *Downsample*, na qual reduzem a imagem em dimensionalidade, acarretando em um modelo com melhor performance pois reduzem o tempo de treinamento, além de gerarem invariâncias rotacional e translacional na rede que agregam às características aprendidas nas camadas convolucionais (LAI, 2015).

Nas camadas finais normalmente são incluídos os *layers* totalmente conectados, do inglês *Fully Connected* (FC), em que todo neurônio da camada anterior possui uma conexão com todo neurônio da camada seguinte. Nesse estágio as informações dependentes da posição e padrões mais gerais são codificados para serem passados à saida, a importância de que todos os padrões apreendidos no mapa de características seja passado aos nós finais, torna importante a presença do FC nesse estágio da arquitetura (GHOSH et al., 2020).

3.5 Transferência de Aprendizado

Embora para o ser humano seja uma tarefa fácil reconhecer objetos pelo sistema visual, para a máquina, tal trabalho não é algo tão simples. As imagens quando processadas e armazenadas de forma digital, constituem um dos formatos de dados mais volumosos, e quanto melhor a resolução, mais bits serão requeridos. Para as CNNs aplicarem o processo de convolução descrito anteriormente, e atualizar o parâmetros da rede durante o treinamento, é demandado um alto custo computacional e consequentemente um maior tempo para o treinamento (LECUN; BENGIO; HINTON, 2015; SAHA, 2018).

Para superar o desafios citado, diferentes arquiteturas foram criadas e outras técnicas complementares foram desenvolvidas e testadas na área. Nesse último grupo uma aplicação comumente usada é a técnica de Transferência de Aprendizado (TA), a qual visa prover uma melhora no aprendizado do modelo por meio da transferência de conhecimento de uma tarefa já aprendida por outra rede, tais tarefas são muitas vezes a identificação de padrões comuns em diferentes tipos de imagens, como estruturas e formas que são compartilhadas nos mais diferentes domínios, e normalmente identificados pelas primeiras camadas da rede (SAHA, 2018; SARKAR, 2018).

Além de lidar com a redução de custo computacional, já que uma parte do aprendizado é herdada, a Transferência de Aprendizado também permite otimizar a performance do modelo quando o aprendizado não conta com um conjunto de dados extenso. Com o conhecimento transferido, é possível que a parte do modelo a ser treinada concentre-se apenas na identificação de padrões mais complexos do domínio em específico ao qual está sendo aplicado. Como o número de *datasets* públicos de imagens médicas são escassos e não tão extensos, a utilização dessa técnica pode suprir os desafios da aplicação de *Deep Learning* nesse contexto. (SAHA, 2018; SARKAR, 2018)

Embora a TA possa ser efetiva, em muitos casos há também aplicações que podem prejudicar a eficiência do modelo, especialmente quando os domínios de aplicação dos modelos são bastante distintos, como é o caso de imagens médicas, em que a disponibilidade de modelos treinados com esse conjunto de imagens é escasso, e a herança de aprendizado por uma rede treinada em diferente domínio pode não ter sucesso, já que a diferença entre as características dos padrões a serem reconhecidos são muito distintas. No entanto, mesmo com as diferenças de estruturas e padrões entre imagens médicas comparadas à imagens tradicionais, alguns trabalhos mostraram-se eficazes (RAGHU et al., 2019; MORID; BORJALI; FIOL, 2021) no uso da Transferência de Aprendizado utilizando modelos treinados com imagens tradicionais para a transferência de seus parâmetros a redes aplicadas na classificação de imagens médicas. Embora promissor, é importante ressaltar que na revisão literária feita por Morid, Borjali e Fiol (2021), é notado uma grande variância entre os resultados após a aplicação da TA, em que o autor constata que as diferenças entre os tipos de imagens médicas e o objetivo da classificação podem resultar em diferentes performances. Nenhuma aplicação na classificação de pacientes com MCI por imagens MRI foi encontrada na revisão, sendo o tema proposto nesse trabalho uma abordagem inicial da técnica dentro desse contexto.

3.6 Data Augmentation

Outra técnica corriqueiramente utilizada para superar o desafio de treinar um modelo de classificação com um conjunto de dados pequenos, e também afim de evitar o *overfitting*, é a chamada *Data Augmentation* (DAG). O método consiste na aplicação de transformações nos arquivos do conjunto de dados original durante o treinamento do algoritmo, as variações na orientação, coloração e outras características da imagem permite que o modelo generalize melhor seu aprendizado, especialmente quando é contado apenas com um pequeno conjunto de imagens levando o modelo ao *overfitting* rapidamente (GLICKMAN, 2020).

O *Data Augmentation* por ser amplamente usado na classificação de imagens naturais, há um número elevado de diferentes técnicas que podem ser aplicadas nessa etapa. No entanto, ainda há um certo desconhecimento das técnicas ótimas no domínio de imagens médicas. No trabalho feito por Hussain et al. (2018b) o autor explora a influência de diferentes técnicas de DAG, na eficácia do modelo classificador aplicado a imagens de mamografia. Um ponto interessante desse trabalho está na divisão da análise feita por um parâmetro chamado de informação mútua (MI), o qual mensura a similaridade entre as imagens pré-aumentadas e após a aplicação da transformação. Ao obter as acurácias do modelo aplicando individualmente várias técnicas diferentes de DA, o autor chegou a um resultado surpreendente: as melhores performances foram obtidas com as técnicas que possuíam os maiores valores de MI, em outras palavras as transformações que preservam maior informação do dataset original foram as mais indicadas para a aplicação do DA (HUSSAIN et al., 2018b). Os métodos que obtiveram melhores resultados e que serão utilizados nesse trabalho são os seguintes:

- Rotação: transformação que realiza voltas em relação a um ponto fixo da imagem chamado de centro de rotação. Tal transformação preserva o tamanho e formato da imagem, mas a figura é rotacionada numa dada direção θ. A rotação de uma imagem é realizada por um cálculo matricial, na equação logo abaixo está demonstrado a aplicação desse cálculo para um dado ponto x,y no espaço cartesiano, em que θ é o ângulo de rotação aplicado:

$$Rv = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta & -y\sin\theta \\ x\sin\theta & y\cos\theta \end{bmatrix} \tag{3.8}$$

- Zoom: operação de escalamento na imagem feito por uma transformação afim, que aumenta ou reduz a imagem por um dado fator de escalonamento. De forma mais detalhada o operador de escala executa uma transformação geométrica que pode ser usada para ampliar o tamanho de uma imagem. O zoom é obtido por replicação de pixels ou por interpolação dos valores de pixels vizinhos na imagem original pode ser realizada a fim de substituir cada pixel por um grupo expandido de pixels.

- Deslocamento: transformação que desloca cada ponto para uma dada distância escolhida. Em matemática, uma matriz de deslocamento é uma matriz binária com apenas 1 na diagonal e 0 nos demais índices. Assim como uma transformação linear, uma matriz de deslocamento inferior (subdiagonal com 1) trabalha deslocando os componentes de um vetor coluna uma posição para baixo, com um zero aparecendo na primeira posição. Já a matriz de deslocamento superior

(superdiagonal com 1) desloca os componentes de um vetor coluna uma posição para cima, com um zero aparecendo na última posição. A operação de deslocamento pode ser vista matematicamente no exemplo da equação 3.9, onde S é uma matriz de deslocamento e A a matriz a ser deslocada.

$$S \times A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 3 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 3 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \end{bmatrix} \tag{3.9}$$

4 Metodologia

A construção do modelo classificador proposto se dividirá em dois tipos de abordagem, que serão explicadas mais detalhadamente na seção 4.3. A construção da primeira abordagem seguirá a ordem das seguintes etapas: Pré-processamento; Segmentação; Transferência de aprendizado e treinamento do classificador. A segunda etapa segue a mesma ordem, no entanto será excluída a etapa de segmentação.

Os modelos foram desenvolvidos utilizando o ambiente da plataforma *Kaggle*, contando com uma GPU NVidia K80 para o treinamento dos modelos. A placa de vídeo é capaz de executar cálculos tensoriais de forma muito mais veloz do que uma CPU, sendo vital para o treinamento de modelos de Deep Learning (LECUN; BENGIO; HINTON, 2015; SKANSI, 2018).

4.1 Dados

O dataset utilizado foi retirado do projeto *Open Access Series of Imaging Studies* (OASIS), o qual visa disponibilizar de forma pública e gratuita um conjunto de neuroimagens para a comunidade científica. O projeto conta por volta de 5000 imagens MRI, todas focadas na análise de pacientes com Alzheimer nos mais variados estágios da doença. A disponibilização do dataset vem ajudando a comunidade a realizar tarefas como análises de dados orientadas por hipóteses, desenvolvimento de atlas neuroanatômicos e desenvolvimento de algoritmos de segmentação e classificação (MARCUS et al., 2009).

Os dados foram coletados por vários projetos em andamento por meio do instrumento WUSTL Knight ADRC, ao longo de 30 anos. Os participantes incluem 609 adultos cognitivamente normais e 489 indivíduos em vários estágios de declínio cognitivo, com idades entre 42 e 95 anos. . O conjunto de dados contém mais de 2.000 sessões dos quais foram coletados apenas os volumes de peso T1, os quais estavam acompanhadas por arquivos de segmentação volumétrica produzidos por meio do software de processamento *Freesurfer*, permitindo assim o trabalho ser realizado por meio de imagens 2D, baseando a classificação em patches (MARCUS et al., 2009).

Para a determinação do grau de demência foi utilizado como parâmetro o *Clinical Dementia Rating* (CDR), que é uma métrica amplamente utilizada para determinar o grau de demência. O CDR trabalha em uma escala de 5 pontos usada para caracterizar seis diferentes tipos de domínios de desempenho cognitivo e funcional, aplicáveis à doença de Alzheimer e outras demências relacionadas. Os domínios cognitivos usados

para avaliação são: memória, orientação, julgamento e solução de problemas, assuntos comunitários, tarefas de casa e hobbies e cuidado pessoal. As informações necessárias para fazer cada classificação, foram obtidas por meio de uma entrevista semiestruturada com o paciente e um informante confiável, ou fonte colateral (por exemplo, um membro da família).

Na tabela 1 pode ser visto o mapa de CDR relativo à cada grau de demência atribuido. O classificador desse projeto trabalhará no diagnóstico de pacientes apenas com os três primeiros graus da tabela (0, 0.5 e 1).

Tabela 1 – Rating CDR.

Score	Nível de demência
0	Não demente (ND)
0.5	Demência muito leve (VMD)
1	Levemente demente (MD)
2	Demência moderada
3	Demência severa

Fonte: (MARCUS et al., 2009)

As imagens do projeto foram adquiridas por um scanner 1.5-T Vision (Siemens, Erlangen, Alemanha) em uma única sessão de imagem. Os parametros de MP-RAGE foram empiricamente otimizados para contraste cinza-branco. O scanner e as sequências foram mantidos ao longo da duração do estudo, de modo que os dados presentes não são influenciado por atualizações de hardware ou outro instrumento diferentes. (MARCUS et al., 2009)

Figura 7 – Imagens de cada classe.

Fonte: O autor

4.2 Pré Processamento

Para cada paciente, os arquivos de digitalização individuais foram convertidos do formato IMA (proprietário da Siemens) em NiFTI1 de 16 bits, as imagens também foram corrigidas para obter o movimento da cabeça entre scans e espacialmente distorcido com base no atlas de Talairach e Tournoux (1988). A imagem foi primeiramente segmentada para classificar o tecido cerebral como fluido cérebro-espinhal, matéria cinzenta ou branca. O procedimento de segmentação atribuiu iterativamente voxels às suas respectivas classes de tecido com base em estimativas de máxima verossimilhança baseadas no modelo de Campo Aleatório de Markov. O modelo usa a proximidade espacial para restringir a probabilidade com a qual voxels de uma determinada intensidade são atribuídos a cada classe de tecido. Finalmente, o volume de todo o cérebro normalizado (nWBV) foi calculado como a proporção de todos os voxels dentro da máscara do cérebro classificada como tecido (matéria cinza ou branca) (MARCUS et al., 2009).

4.2.1 Métodos adicionados

De acordo com Wen et al. (2020), é constatado que "*CNNs requerem apenas um mínimo pré-processamento devido a sua habilidade em extrair características dos mais baixos até os mais altos níveis*". Revisando a literatura de modelos classificadores de imagens MRI (BRON et al., 2015; SARRAF, 2016; WEN et al., 2020; HOSSEINI-ASL; KEYNTON; EL-BAZ, 2016; ZOU et al., 2017) em busca de um padrão de pré-processamento para imagens MRI, foi visto que alguns passos são comuns a maior parrte dos trabalhos para o pré-processamento das imagens MRI antes da classificação. Apenas Bron et al. (2015) realizou uma análise sobre a influência do pré processamento nos resultados do modelo, e obteve uma pequena diferença no desempenho comparando duas diferentes abordagens: uma extensiva e a outra básica. O pré processamento também é importante para a redução de dimensionalidade das imagens, ao passo que embora quanto maior informação disponível, melhor seja o aprendizado em termos de acurácia, o custo operacional também tende a aumentar drasticamente com arquivos muito grandes, como é o caso das imagens MRI (BRON et al., 2015; SARRAF, 2016). Com base nessas informações fornecidas pela literatura, foi realizado apenas as etapas básicas que estão presentes em Bron et al. (2015), Sarraf (2016), ZOU et al. (2017) com foco em reduzir a dimensionalidade da imagem, levando em consideração a pequena diferença em termos de acurácia encontrada por (BRON et al., 2015).

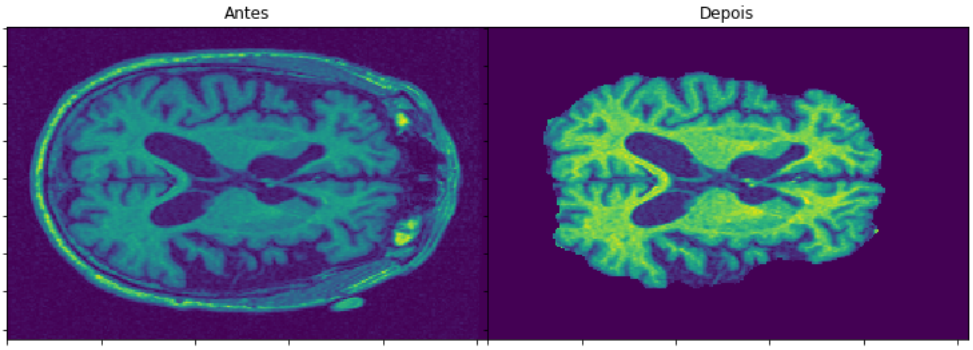
- *Skull Stripping*: Eliminação dos ossos da imagem por uso de máscaras, uma espécie de filtro para intensidades de píxels. Essa tarefa permite focar o treinamento

apenas nas regiões que possam armazenar padrões de interesse. Realizado pelo software *FreeSurfer*, o resultado desse processo poder ser visto na Figura 7.

- Normalização dos valores de intensidade dos pixels, para que a variação das escalas de valores não interfira no aprendizado. Para essa etapa foi utilizado o método denominado *MinMax*, que foca a nova escala dos píxels baseada nos valores mínimos e máximos, de forma que o novo conjunto de valores situem-se entre 0 e 1. A equação (4.1) mostra a aplicação da função *MinMax*, considerando que x' representa o novo valor após o escalonamento, x o valor na escala atual, x_{min} o menor valor da variável x, e x_{max} o maior valor dessa variável.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.1}$$

Figura 8 – Skull Stripping.



Fonte: O autor

4.3 Classificação

4.3.1 Rede Convolucional Densamente Conectada

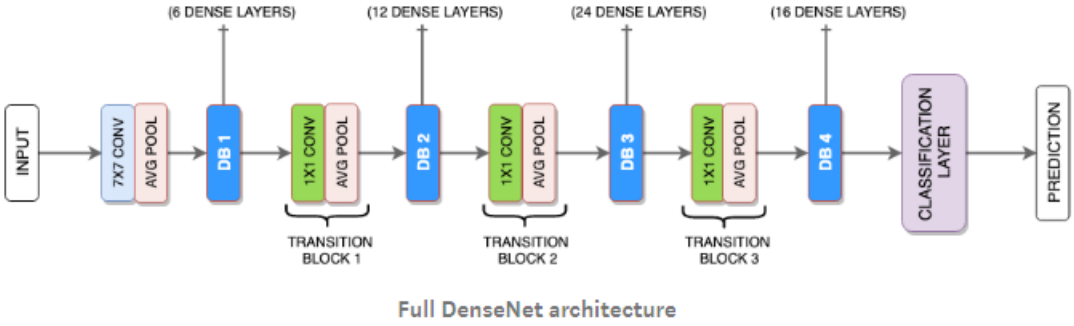
A arquitetura CNN escolhida como classificador nesse projeto foi a *Rede Convolucional Densamente Conectada*, do inglês *Densely Connected Convolutional Network* (DenseNet), e será o principal foco na análise de desempenho. A arquitetura foi criada e publicada no artigo de Huang, Liu e Maaten (2018) por quatro pesquisadores, sendo o autor principal Gao Huang da Cornell University, e contando também com a participação do diretor de pesquisa em Inteligência Artificial do Facebook, Laurens Van der Maaten.

O modelo foi ganhador do prêmio de melhor artigo do IEEE *Conference on Computer Vision and Pattern Recognition* (CPVR), e focou seu desenvolvimento na melhoria da qualidade da imagem por meio de reconstrução, otimizando a perfomance de classificação para imagens com super resolução, caso de grande parte das imagens

médicas. A arquitetura contribuiu na fuga do problema da Dissipação de Gradiente, obtendo o sucesso na propagação de muitas características pela rede com a redução do número de parâmetros e reutilizando características (HUANG; LIU; MAATEN, 2018).

O trabalho foi baseado na utilização de conexões curtas entre os layers, que permitem a construção de modelos mais profundos (com maior número de camadas), resultando em maior acurácia sem que seja perdido a eficiência no treinamento. A arquitetura completa do modelo é visto na Figura 8.

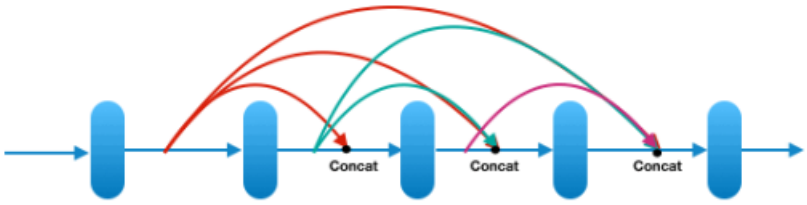
Figura 9 – Arquitetura da DenseNet.



Fonte: Khannel (2019)

A grande característica desse modelo é de que cada camada se conecta com todos as outras da rede por meio da retro-propagação, esse tipo de abordagem que foi chamada de conexões densas, permite um melhor fluxo do gradiente, já que ocorre um aprendizado compartilhado na rede, necessitando de menos *layers* ao passo que evita a redundância no aprendizado das características, acarretando em um menor número de parâmetros para o aprendizado da rede (KHANNEL, 2019).

Figura 10 – Abstração das conexões densas presentes na DenseNet.



Fonte: Khannel (2019)

Em seu trabalho, Huang, Liu e Maaten (2018) comparou a perfomance de seu modelo com a famosa *Residual Netowrk* (ResNet), e um fato interessante foi notado que reflete o aprimoramento da DenseNet: embora as duas redes tenham apresentado um desempenho semelhante, a última registrou apenas 1/3 do número

de parâmetros necessários para treinamento da ResNet, tal fator acarreta um menor custo computacional para o treinamento do modelo (HUANG; LIU; MAATEN, 2018).

Arquiteturas de modelo retro-alimentada como a ResNet, conectando a saída do *layer* l como a entrada do *layer* l+1 que dá origem ao layer de transição representado por: $x_l = H_{l-1} + x_{l-1}$, sendo a função H uma transformação não linear aplicada no *layer* a que se refere, uma das grandes qualidades dessas redes que foi inspirado para criação das DenseNets, foi o fato de que as ResNets podem conectar a função identidade das últimas camadas com a entrada das primeiras, isso ajudaria no melhor fluxo de aprendizado do modelo. No entanto, nessas redes a função H soma-se à função identidade impedindo que a informação se distribua na rede. A forma encontrada por Huang, Liu e Maaten (2018) de contornar essa situação, foi criar as camadas de conexão densa, em que cada *layer* se conecta com todos os *layers* que estão depois dele, possibilitando que o gradiente de informação se distribua entre camadas não vizinhas.

Assim como Huang, Liu e Maaten (2018) realizou em seu artigo, esse trabalho visa analisar a influência de duas variáveis dessa rede no aprendizado: o *growth rate* (k) e a profundidade (L), no entanto aplicando a rede a um diferente domínio de imagem do qual o trabalho em questão foi realizado.

Parâmetros importantes da rede que serão avaliados :

- Profundidade (L): Número de layers presente na rede, cada um implementa uma função $H_l(\times)$ composta. Conforme o número aumenta mais complexa e com mais parâmetros a rede é formada (HUANG; LIU; MAATEN, 2018).
- Growth Rate (k): Considerando que a cada passagem de função composta são produzidas uma quantidade k de mapas de características então o layer l^{th} terá uma quantidade de mapa de características em sua entrada de $k_0 + k * (l - 1)$. Sendo k_0 é o número de canais no layer l. A importância dessa variável no trabalho de (HUANG; LIU; MAATEN, 2018) é que ela possibilita demonstrar que a DenseNet pode obter ótimas performances utilizando menos parâmetros, ao passo de que essa variável retrata como a rede irá crescer em número de parâmetros. O presente trabalho analisará a perfomance do modelo treinando com 2 valores de k distintos (12 e 24) para constatar como essa variável do modelo se comporta trabalhando com arquivos que contém um volume maior de informação (HUANG; LIU; MAATEN, 2018; LECUN; BENGIO; HINTON, 2015).

4.3.2 Xception

Devido a falta de padronização das abordagens feitas na revisão literária, tornou-se difícil e inapropriado comparar o presente projeto com classificadores apresentados

em outros trabalhos. As abordagens variam em: dimensionalidade, com classificações em volumes (YANG; RANGARAJAN; RANKA, 2018; KOROLEV et al., 2017) e imagens 2D (QIU et al., 2018; SHI et al., 2017); Pacientes com Alzheimer (WANG et al., 2018) e/ou MCI (QIU et al., 2018); com imagens MRI (LIU; CHENG; YAN, 2018) e/ou multimodais (SUK; LEE; SHEN, 2014).

Dado que o objetivo nesse projeto está na inserção de uma nova arquitetura (DenseNet), avaliação de técnicas já conhecidas (Transfer Learning e Data Augmentation) mas pouco debatidas sobre a influência no presente contexto, além do foco no diagnóstico de pacientes em estágio inicial (MCI), foi decidido reproduzir as mesmas abordagens com uma arquitetura já conhecida em outros trabalhos conhecido como *Xception*, como ponto de comparação, a qual deriva da tradicional rede *Inception* (SZEGEDY et al., 2016).

Assim como a maior parte das arquiteturas CNN, a *Xception* também foi lançada no desafio ImageNet, desenvolvido por (CHOLLET, 2017) e pertencente à Google o nome dessa rede deriva de "Versão Extrema da Inception", já que tem como ponto de referência a rede Inception. Ambas as arquiteturas tem como ponto base a operação feita pelos layers chamada de Convolução Separável em Profundidade, do inglês *Dense Separated Convolution* (DSC), a qual é constituída por dois elementos centrais:

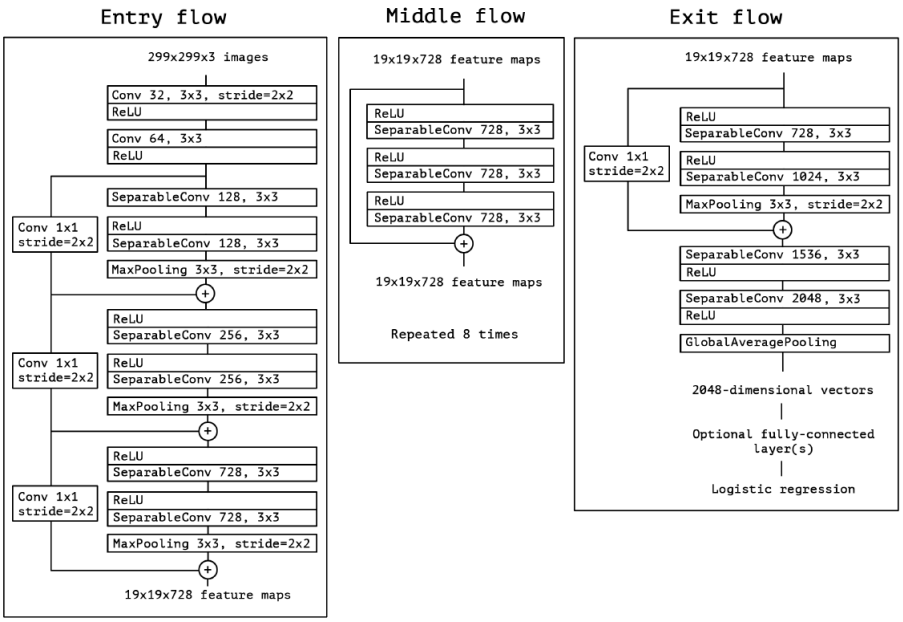
Convolução pontual: Convolução com kernel 1x1 que busca identificar as correlações entre layers diminuindo a dimensão inicial.

Convolução profunda: Camadas com convolução espacial nxn, a qual mapeia todas as correlações nos espaços dimensionados pela convolução pontual.

Uma DSC original é composta por uma convolução profunda seguida de uma convolução pontual, essa abordagem permite que cada operação seja dedicada independentemente a correlações espacias e outra para correlações entre canais. Separando essas duas tarefas foi possível evitar a necessidade de operar convolução em todos os canais da rede, diminuindo o número de parâmetros (CHOLLET, 2017; SZEGEDY et al., 2016).

A rede *Xception* trabalha com uma Convolução Separável em Profundidade Modificada, que foi introduzida no modelo Inception V3. Essa operação muda a ordem da DSC, iniciando com uma convolução pontual e seguida por uma convolução profunda. Como o próprio nome diz a *Xception* buscou uma versão extrema da Inception, separando totalmente a independência na identificação das correlações entre canais e correlações espaciais.

Figura 11 – Arquitetura da rede Xception



Fonte: Chollet (2017)

4.4 Transferência de Aprendizado

Visando superar a falta de um dataset extenso e poder computacional limitado, foi utilizado a técnica de Transfer Learning utilizando a rede DenseNet169 com a primeira metade dos layers pré-treinados com os pesos do projeto ImageNet, a rede e seus parâmetros estão disponíveis no framework TensorFlow. A outra metade restante será treinada com o dataset específico utilizado nesse trabalho, essa abordagem pretende concentrar o aprendizado da rede apenas nos detalhes específicos do contexto de identificação de MCI, essas características mais complexas são identificadas nos layers finais da rede (LECUN; BENGIO; HINTON, 2015; SAHA, 2018).

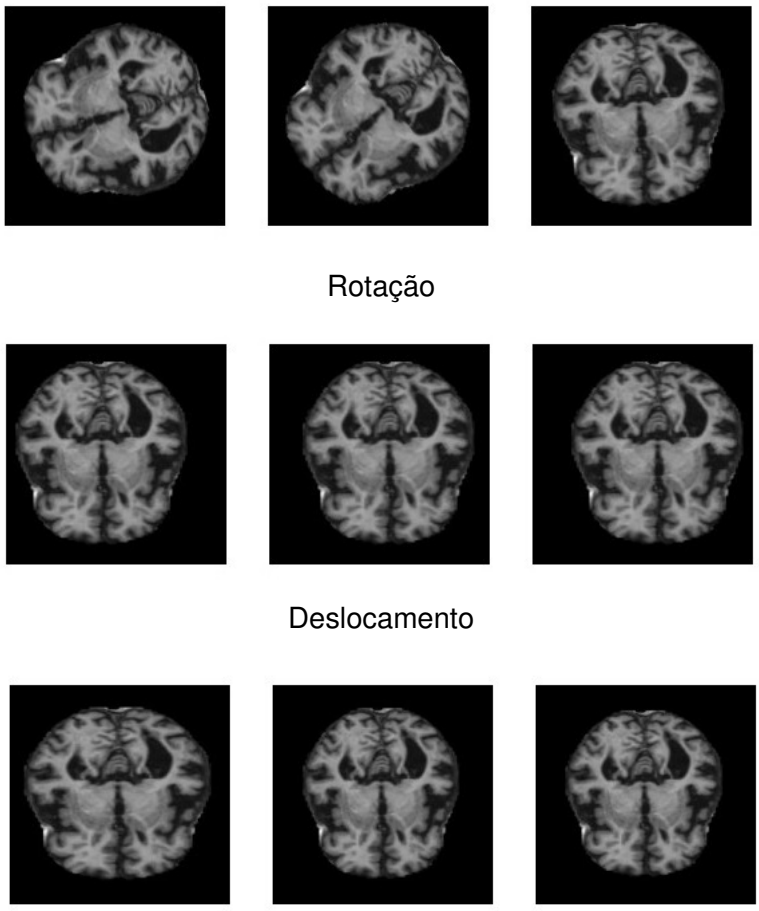
4.4.1 ImageNet

Grande parte dos maiores feitos alcançados na área de visão computacional nos últimos anos estiveram presentes no *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC), uma competição anual de visão computacional desenvolvida sob um enorme dataset disponível publicamente denominado de ImageNet, nesta competição foram desenvolvidas as principais redes CNN (BROWNIEE, 2019). Além de contribuir com novas arquiteturas para a comunidade, o ILSVRC ajudou em muitos projetos com a disponibilização dos pesos dessas redes após serem treinadas com o ImageNet no desafio. Esse dataset contém mais de 14 milhões de imagens com mais de 21 mil classes anotadas, esse volume gigantesco de dados permite que parte do aprendizado sobre esse dataset possa ser compartilhado em domínios que

não possuem essa quantidade de imagens para treinamento (BROWNIEE, 2019; KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

4.5 Data Augmentation

Figura 12 – Data Augmentation.



Fonte: O autor

As transformações foram aplicadas no conjunto de imagens que alimentou o treinamento, onde foram geradas cópias transformadas das imagens originais pra cada técnica durante o treinamento do modelo.

Conforme visto no capítulo 3, as técnicas escolhidas para o processo de *Data Augmentation* foram baseadas no trabalho feito por Hussain et al. (2018a) em que foi constatado uma melhor adaptabilidade dos classificadores em imagens, quando a técnica de *Data Augmentation* preserva a maior parte da informação. Dentro dessa perspectiva, foram escolhidas as técnicas que apresentavam maior valor de Informação Mútua: rotação, zoom e deslocamento. A aplicação dessas técnicas apliacadas ao

dataset estão exemplificadas na Figura 10, enquanto na Tabela 2 estão listados os *ranges* utilizados para cada transformação

Tabela 2 – *Ranges* utilizados no DAG.

Transformação	Range
Rotação	90°
Deslocameto	0.10
Zoom	0.10

4.6 Detalhes do procedimento

Para o treinamento e posterior análise de desempenho as imagens foram divididas em três grupos: treinamento, teste e validação. O primeiro refere-se ao conjunto de dados que foi utilizado para treinar o modelo, o segundo para testar a eficácia do modelo durante o treinamento com dados não vistos e a terceira consiste no conjunto de imagens também não vistas durante o treinamento e que servirá como validação final da performance de cada etapa (LIU; CHENG; YAN, 2018; KHVOSTIKOVA et al., 2018; WEN et al., 2020).

O treinamento dividiu-se em seis processos de variação e que foram treinados e analisados conforme o esquema que pode ser visto logo abaixo:

- Não balanceado: treinamento com o dataset desbalanceado (número de imagens por classes de ordens diferentes), conforme foram coletadas.
- Balanceado: treinamento com o número de imagens por classe balanceados por meio do *undersample*.
- *Augmentation*: foi treinado o conjunto de dados utilizando a técnica de augmentation.
- Sem *augmentation*: treinamento do conjunto de dados sem nenhuma adição de augmentation.
- Variação de parâmetros: treinamento com a DenseNet variando a profundidade (L) em 7, 24 e 40 layers, e a taxa de crescimento (k) em 12 e 24. Para cada variação a tabela explica o nome adotado de cada modelo.
- Transferência de aprendizado: treinamento com a DenseNet atualizada com os parâmetros treinados no projeto ImageNet.

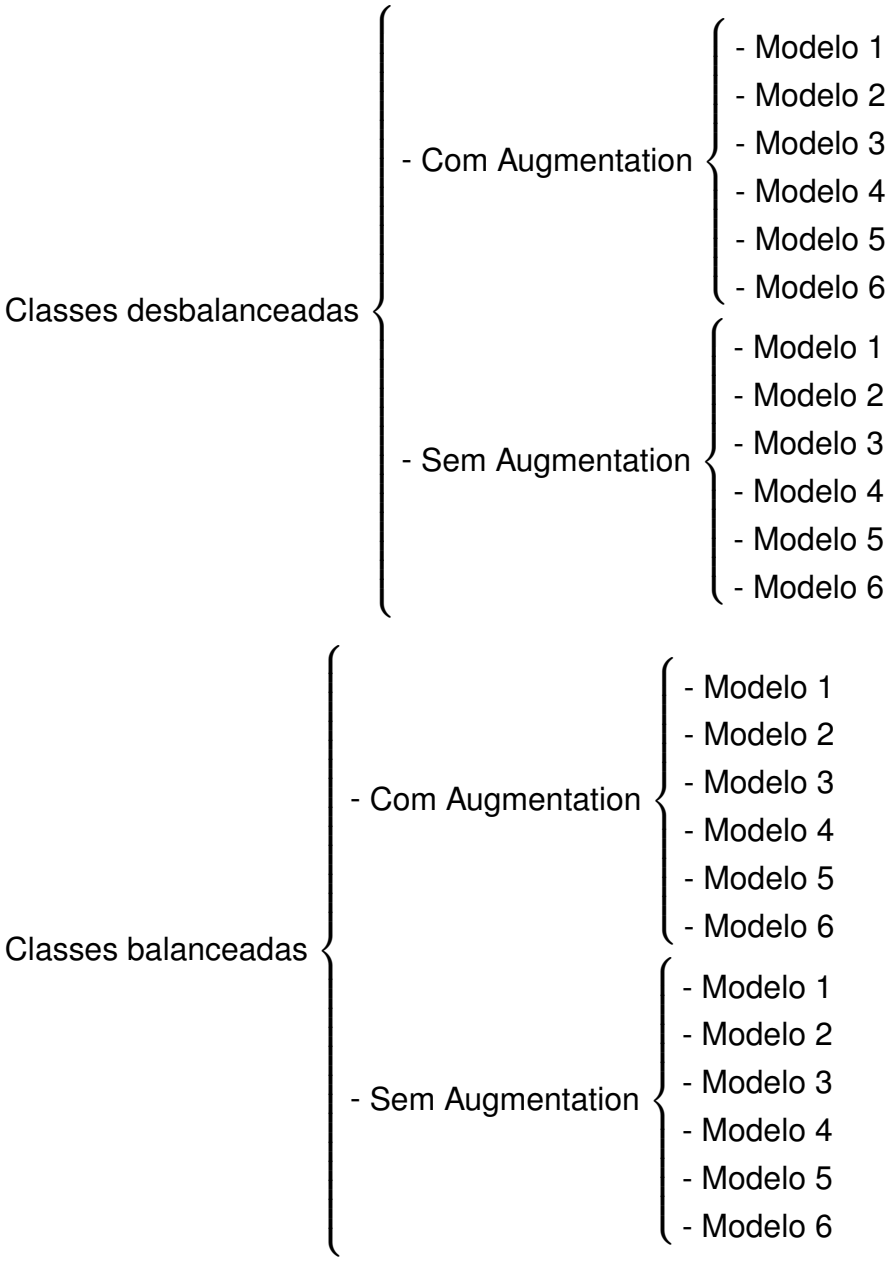


Tabela 3 – Imagens por classe em cada abordagem.

Abordagem	Normal	VMD	MD
Não balanceado	3200	2240	896
Balanceado	950	950	896

Para a avaliação de desempenho foram utilizadas as seguintes métricas:

- Acurácia de classificação (ACC): definida como a proporção das imagens classificadas corretamento pelo número total de imagens.
- Área abaixo da curva ROC (AUC): A curva ROC é obtida por meio da razão entre a taxa de Positivos verdadeiros (TPR) versus a taxa de falsos positivos (FPR), quanto maior a área abaixo dessa curva melhor a perfomance do modelo dentro de cada limite de classificação.

Tabela 4 – Dicionário de modelos.

	L	k
Modelo 1	7	12
Modelo 2	22	12
Modelo 3	40	12
Modelo 4	7	24
Modelo 5	40	24
Modelo 6	DenseNet com TA	
Modelo 7	*Xception* com TA	

- Tempo de execução (TE): Tempo que o computador leva para executar o treinamento do modelo.

Para cada métrica foi computada seu valor durante o treinamento e teste, mas na validação será a principal forma de análise, a qual foi obtida da seguinte forma: o modelo após treinado, foi validado por meio da predição com o modelo treinado para o dataset de visualização, sendo todas as métricas computadas três vezes, e então a média dessas amostras ficaram definidas como as acurácias e AUCs definitivas.

5 Resultados

Utilizando o dataset conforme obtido na fonte, houve um desbalanceamento entre as classes conforme foi apresentado na Tabela 3. Para todas as abordagens dividiu-se os dados em treinamento, validação e teste: 64% para treino, 20% validação e 16% teste. Por meio da linguagem Python foram criados os modelos conforme orientado pelos parâmetros mostrados na Tabela 4 e por Huang, Liu e Maaten (2018). Fazendo uso do ambiente disponível na plataforma *Kaggle*, foi utilizado a GPU para treinamento dos modelos. Na abordagem de balanceamento entre as classes, foi feito o *undersample* por meio da redução de imagens das classes majoritárias, ND e VMD. Dividindo de forma equivalente cada classe, com o número de arquivos para treinamento iguais entre elas. A divisão após o *undersample* pode ser vista com mais detalhes na Tabela 5.

Tabela 5 – Balanceamento do conjunto de dados após balanceamento

	ND	VMD	MD
Treino	650	650	650
Teste	150	150	100
Validação	150	150	146

Afim de reproduzir o modelo de maneira semelhante ao trabalho de (HUANG; LIU; MAATEN, 2018), foi adotado como função de perda o Gradiente Descendente Estocástico, com decaimento de peso de 10^{-4} e momentum Nesterov de 0.9 sem amortecimento. Utilizando 40 épocas e um lote de tamanho 32, foram obtidos os resultados vistos na Tabela 6.

Considerando inicialmente apenas as DenseNets sem aprendizado prévio (M1 até M5), foi visto que de maneira geral os resultados não foram tão satisfatórios, tendo a melhor acurácia obtida utilizando o Modelo 4 (57,14%) no conjunto desbalanceado de imagens e sem a aplicação de *Data Augmentation*. No entanto, avaliar somente essa métrica para classes desbalanceadas pode distorcer a interpretação dos resultados (o modelo pode ter previsto todas as instâncias como a classe majoritária), portanto avaliar a métrica AUC e as matrizes de confusão serão necessárias para uma melhor explicabilidade do resultado. Já na tabela 6 é possível notar que o Modelo 6, a DenseNet com transferência de aprendizado, saiu-se ótima para todas as abordagens, indicando que os padrões aprendidos no ImageNet *Challenge* também ajudaram a rede a diferenciar as classes de pacientes MCI e saudáveis. A diferença dos resultados entre a abordagem com transferência de aprendizado e os modelos sem pré aprendizado foram bem significantes, mostrando o enfraquecimento da DenseNet quando é treinada

Tabela 6 – Resultados obtidos com a DenseNet.

	Desbalanceado											
	Sem *augmentation*						Com *augmentation*					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
ACC	53,35	48,07	55,41	57,14	50,83	66,77	54,12	51,54	36,49	52,09	50,51	61,59
AUC	76,16	67,67	77,35	76,77	74,90	80,63	75,78	74,91	52,92	74,08	74,24	80,66
TE (min)	6,35	18,05	54,79	6,02	104,9	23,23	36,68	42,05	60,2	36,82	108,6	36,02

	Balanceado											
	Sem augmentation						Com augmentation					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
ACC	41,00	37,5	38,75	48,25	51,75	66,25	49,08	52,75	55,67	46,33	43,75	64,75
AUC	64,99	63,38	57,39	70,57	73,78	79,00	70,21	71,48	73,89	69,28	68,00	77,21
TE (min)	1,71	6,90	21,05	1,78	40,17	3,68	11,22	13,70	22,44	11,22	41,09	3,59

com um conjunto de dados pequenos. Mesmo utilizando o *Data Augmentation*, que acrescenta mais imagens ao treinamento, não foi possível notar alguma diferença expressiva nos modelos sem a TA. Apenas na abordagem utilizando o conjunto de dados balanceado, houve uma leve melhora nos modelos após o DAG.(HUANG; LIU; MAATEN, 2018; LECUN; BENGIO; HINTON, 2015) A avaliação foi feita por influência de cada técnica aplicada.

5.1 **Balanceamento**

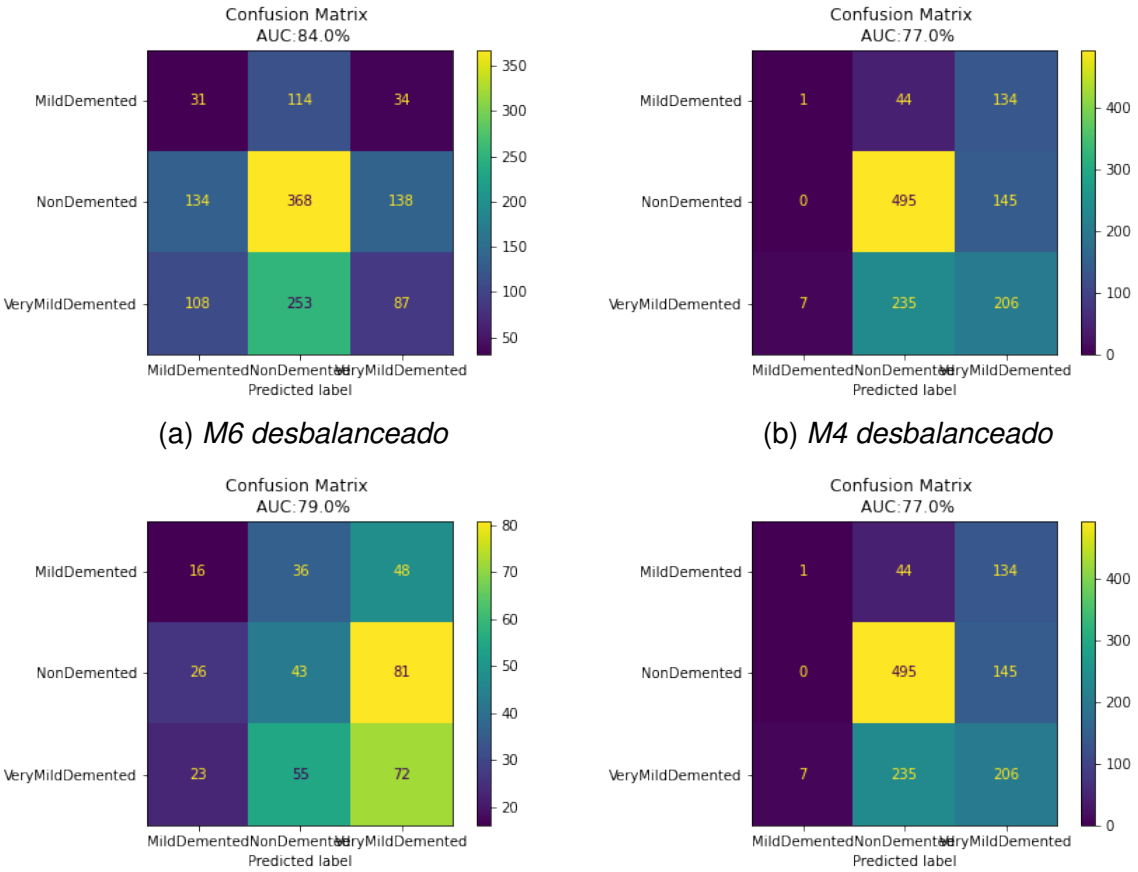
Comparando somente a tabela 6, é visto que a utilização do conjunto de dados desbalanceado aparentou obter resultados levemente superiores do que aplicando o *undersample*, avaliando tanto pela métrica AUC como para ACC . Já na Figura 13 é possível comparar as matrizes de confusão dos modelos com melhores resultados para o conjunto de dados desbalanceado (M3 e M4 sem *augmentation*), em relação aos mesmos modelos com o os dados balanceados. Pelas matrizes é possível ter uma visão mais detalhada de como o modelo trabalhou nas previsões dos dados de validação, e por ela é possível concluir que de forma geral o desbalanceamento entre as classes não afetou tanto o desempenho dos melhores modelos como o esperado, embora a classe minoritária, representada pelos pacientes com MCI moderado, tivesse sido ignorada nas previsões dos modelos em algumas abordagens.

Mesmo trabalhando com o conjunto de dados desbalanceado, os modelos com os melhores resultados não tenderam a escolher somente a classe majoritária, como

ocorre na maior parte das vezes em conjuntos de imagens desbalanceadas entre as classes. A pequena perda na performance para o conjunto balanceado, pode ser explicado pela redução do número de exemplos para treinamento após a aplicação do *undersample*, indicando inicialmente uma sensibilidade maior da arquitetura à falta de dados do que ao desbalanceamento (LECUN; BENGIO; HINTON, 2015) .

No modelo 6, o qual foi aplicado a transferência de aprendizado e que obteve os melhores resultados, a influência do balanceamento foi mínima, não refletindo nas métricas finais. De uma forma geral, como o balanceamento por *undersample* não teve impacto significativo na performance do modelo, foi optado em não utilizá-lo na busca pelo modelo ideal, já que nessa abordagem um conjunto de imagens é perdida, diminuindo a quantidade informação disponível ao aprendizado do modelo. Somado ao fato de que o *Data Augmentation* também tem utilidade na redução do impacto de classes majoritárias. Potanto, foi considerado o Balanceamento por *undersample* uma técnica descartável nesse projeto.

Figura 13 – Matrizes de confusão (análise do balanceamento)



(a) *M6 desbalanceado*

(b) *M4 desbalanceado*

(c) *M6 balanceado*

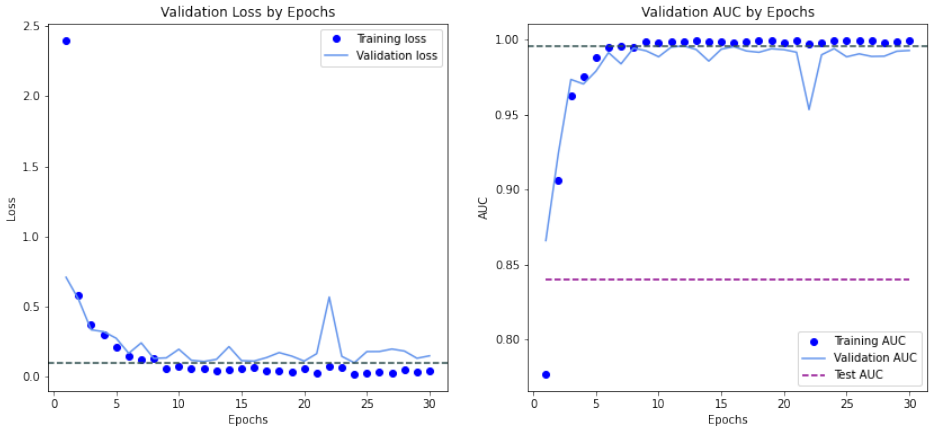
(d) *M4 balanceado*

Fonte: O autor.

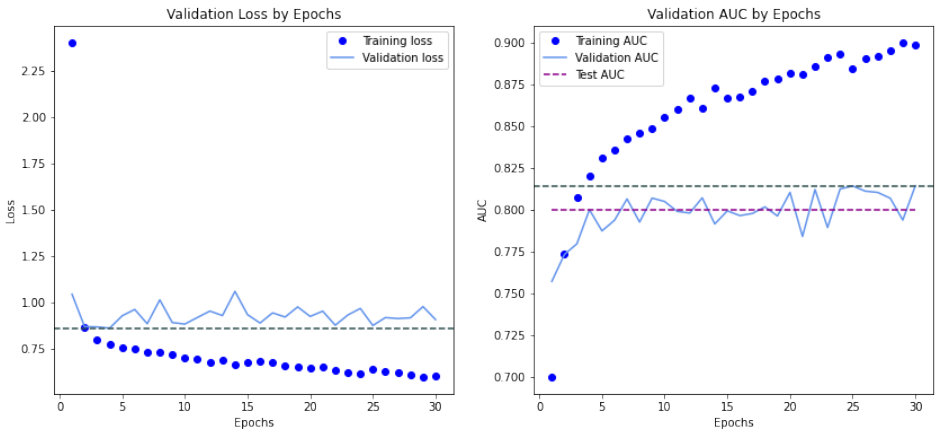
5.2 Augmentation

Com relação ao *Data Augmentation* para a DenseNet, é possível notar que para o conjunto de classes balanceadas, o aumento da peformance do modelo foi mais notável, isso devido ao fato de que o DAG promove um aumento do número de imagens a serem treinadas, especialmente após ter sido feito o *undersample*, que tornou o dataset ainda menor. Tal resultado reforça mais a hipótese do modelo ser sensível negativamente quando treinado com um menor volume de dados (HUSSAIN et al., 2018a).

Figura 14 – Métricas durante treinamento (Análise do *Data Augmentation*).



(a) Treinamento sem DAG.



(b) Treinamento com DAG.

Fonte: O autor.

Analisando a DenseNet submetida a transferência de aprendizado (M6), apenas pela acurácia aparenta que o *Data Augmentation* tenha piorado sua performance. No entanto, ao comparar a curva de AUC ao longo das *epochs*, e a função de perda vista

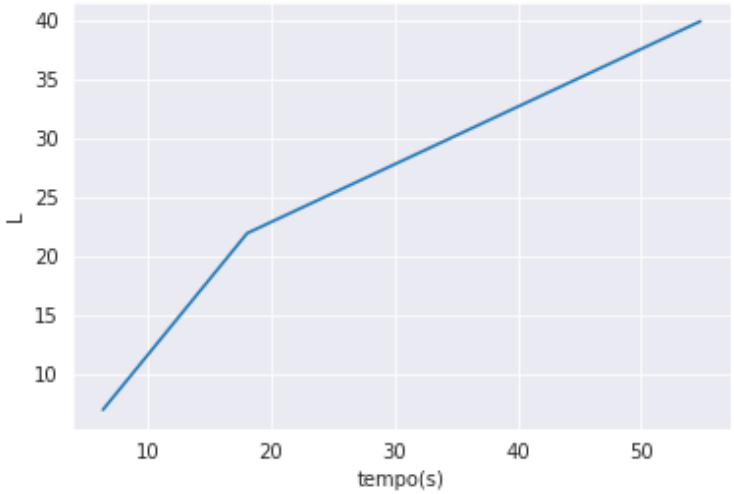
na Figura 14, para os dados de treino representados pelos pontos azuis, a validação durante o treinamento pela curva em azul claro e a validação final na reta roxa, fica evidente uma diferença entre as abordagens. Embora tenha até conseguido uma acurácia melhor, as métricas durante o treinamento e teste sem o DAG diferiu muito da validação final, indicando que possivelmente o modelo pode melhorar seu resultado com mais épocas sendo acrescentadas, já que as curvas das métricas para treino e validação ainda estão próximas (LECUN; BENGIO; HINTON, 2015; SKANSI, 2018).

Já na abordagem com o DAG, embora exista também a diferença entre as métricas de treinamento e validação, a distância entre elas é menor, indicando um maior poder de generalização para dados não vistos. Também considerando a curva ainda ascendente no treinamento, possivelmente o modelo pode melhorar seu resultado com mais épocas sendo acrescentadas, já que a curva AUC e a função de perda ainda não apresentaram uma estabilização. Como o *Data Augmentation* realiza transformações nas imagens afim de permitir uma maior generalização na descriminação entre as classes, é necessário um tempo maior de treino até o modelo atingir o estado ótimo. No entanto a diferença entre as métricas de treino e teste, podem indicar o começo de um enviezamento do modelo (LECUN; BENGIO; HINTON, 2015; SKANSI, 2018).

5.3 Parâmetro L

O parâmetro L, referente a profundidade da rede, é destacado no trabalho de Huang, Liu e Maaten (2018) como positivamente correlacionado com a performance de seu modelo, ou seja, quanto maior a profundidade da rede, maiores foram as acurácias constatadas. Esse fator só seguiu a mesma lógica de forma categórica no conjunto balanceado e com *Data Augmentation*, nos quais os modelos M1, M2 e M3 aumentaram os valores de suas métricas conforme o número de L aumenta (sendo M3 com o maior L, igual à 40). Embora no conjunto desbalanceado e sem aplicação de DAG o modelo 3 tenha performado com o melhor AUC, não fica claro apenas pelas tabelas de resultados, a influência positiva desse parâmetro de forma geral. O efeito de L no aprendizado pode ter sido ínfimo nesse caso, devido a influência das outras variáveis (volume de dados especialmente), que podem ter sido tão impactantes ao ponto de anular qualquer efeito da profundidade da rede sobre o resultado final (SKANSI, 2018). No entanto foi possível notar que o parâmetro L estava correlacionado com o tempo de treinamento, ao passo que quanto mais camadas foram adicionados à rede, mais parâmetros foram necessários passar por treinamento. O gráfico de L em relação ao tempo pode ser visto na Figura 15.

Figura 15 – L x tempo(min).

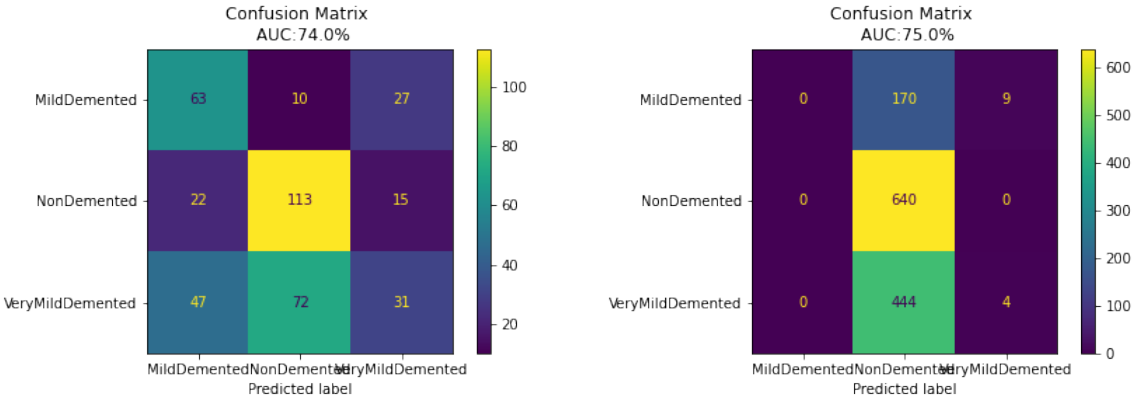


Fonte: O autor.

5.4 Parâmetro k

Os modelos 4 e 5 apresentam um k igual à 24, dobrando o valor em relação a M1, M2 e M3. Conforme a taxa de crescimento aumenta, é possível notar que o tempo de treinamento também é acrescido, ao passo que mais parâmetros são inseridos na rede. No entanto tal complexidade não refletiu nos resultados, em todas as abordagens ao aumentar o número de k as métricas conquistadas foram menores do que os modelos com k menor, se comparado ao seu par de mesmo valor L. Ao observar as matrizes de confusão em quase todos os casos os modelos de k elevados apresentaram a previsão de todas, ou quase todas as instância como a classe majoritária na abordagem

Figura 16 – Matrizes de confusão (análise de k).



(a) *M5 desbalanceado* (b) *M5 balanceado*

Fonte: O autor.

sem balanceamento entre as classes, mostrando o reflexo do desbalanceamento nos resultados quando a rede possui k elevado. Na figura 15 é possível comparar a diferença entre o modelo 4 para a abordagem com o conjunto de dados balanceados, em relação ao mesmo modelo com o conjunto de dados desbalanceado, ambas sem *Data Augmentation*.

5.5 Transferência de Aprendizado

O uso da transferência de aprendizado possibilitou à DenseNet melhorar sua performance comparado ao aprendizado sem pré conhecimento. O uso dos pesos treinados no dataset do ImageNet possibilitou o acesso do modelo a uma quantidade maior de imagens no treinamento, mesmo sendo de domínios difierentes, os padrões transferidos do ImageNet ajudou no diagnóstico das classes de cada paciente.

Mesmo não chegando a alcançar resultados perfeitos na validação, o aumento de desempenho representa uma conquista simbólica dado ao fato de poucos trabalhos terem analisado a influência da transferência de aprendizado no contexto desse projeto. Parte da explicação desse sucesso tem origem na natureza do problema: no exame de MRI estruturual o principal indicador diferenciador entre diferentes graus de perda cognitiva está na atrofia cerebral detectada por esses exames, apresentando-se como um "espaço"na imagem com intensidade de pixel diferente do restante da massa encefálica. Em outras palavras, o modelo procura por "buracos"em que seu tamanho influencia na escolha da classe, portanto o reconhecimento de estruturas como essa podem ser aprendidos em domínios de imagens diferentes (FRISONI et al., 2010).

Tabela 7 – Resultados obtidos com os modelos de TA

	Desbalanceado			
	Sem *augmentation*		Com *augmentation*	
	M6	M7	M6	M7
ACC	66,77	58,96	61,59	59,01
AUC	80,63	74,96	80,66	78,53
TE (min)	23,23	8,18	36,02	45,5

	Balanceado			
	Sem *augmentation*		Com *augmentation*	
	M6	M7	M6	M7
ACC	66,25	51,25	64,75	50,42
AUC	79,00	67,11	77,21	71,46
TE (min)	3,68	4,03	3,59	16,52

Comparando os dois modelos, foi possível notar que os resultados obtidos pela DenseNet foram superiores ao da Xception, no entanto, ao analisar as curvas das métricas de treinamento para ambas as redes foi notado que a convergência entre as métricas de validação e treino não foram atingidas, assim como em nenhum caso ocorreu um sinal claro de *overfitting*, sugerindo aumentar o número de epochs do treinamento até chegar ao modelo ótimo(LECUN; BENGIO; HINTON, 2015; SKANSI, 2018). Na Figura 14 é possível visualizar as curvas de aprendizados da rede Xception (M7) sem *Data Augmentation*, a qual é similar à curva de mesma abordagem da DenseNet na Figura 13b discutida anteriormente.

A melhor performance da DenseNet pode ser explicada pela sua arquitetura contribuir com o compartilhamento de características aprendidas. De acordo com Huang, Liu e Maaten (2018), o fato da DenseNet estabalecer conexões com todos seus layers subsequentes reutilizando as características aprendidas pode ter se beneficiado mais do processo de transferência de aprendizado do que a Xception, dado que a metade inicial pré-treinada com o enorme dataset do ImageNet pôde compartilhar os papdrões aprendidos com toda a rede, tendo mais impacto na decisão dos layers finais (HUANG; LIU; MAATEN, 2018; LECUN; BENGIO; HINTON, 2015; SKANSI, 2018).

Outro ponto interessante de ser notado foi a diferença entre as velocidades das arquiteturas, além de obter melhores resultados a DenseNet também conseguiu completar seus treinamentos na maior parte das vezes de forma mais rápida do que a Xception, e diferente dessa última, ao ser submetida a um maior número de dados de treinamento com a aplicação do *Data Augmentation*, não cresceu seu tempo de treinamento de forma consistente. Tal fato pode ser explicado pela reutilização de features dita anteriormente, que contribui na criação de um modelo mais compacto, além da utilização das chamadas camadas *bottleneck*, que são implementadas no modelo treinado com o ImageNet, disponível para a transferência de aprendizado, essas camadas reduzem dimensionalmente os mapas de características na saída de cada *layer* (HUANG; LIU; MAATEN, 2018).

5.6 Aumento no número de epochs

Devido aos gráficos de aprendizado para os modelos utilizando TA, os quais sairam-se como os melhores e aparentavam não ter chego ao estado ótimo de treinamento, especialmente nas abordagens com *Data Augmentation*. Foi acrescentando um número de 80 epochs para treinamento, obtendo os resultados mostrados na Tabela 8.

Após a inserção de mais epochs para treinamento, foi possível aumentar a acurácia e a AUC, tanto para a abordagem com e sem DAG. No entanto, o aumento de performance e resultado final para o treinamento sem *Data Augmentation* mostrou-se

Tabela 8 – Resultados obtidos com os modelos de TL com 80 epochs

	Desbalanceado	
	Sem *augmentation*	Com *augmentation*
	M6	M6
ACC	71,82	65,17
AUC	84,46	83,56
TE (min)	22,28	96,03

mais eficiente do que utilizando essa técnica. O aumento no tempo de treinamento também mostrou-se um limitante no uso de DAG.

De acordo com Huang, Liu e Maaten (2018), o fato de ser uma rede que utiliza parâmetros de forma mais otimizada do que outras redes tradicionais, o que foi demonstrado em seu trabalho, torna a DenseNet menos propensa a *overfitting*. Esse fator responde a piora da abordagem com *Data Augmentation*, a qual visa transformar as imagens de modo a induzir a rede paraa novos exemplos de treino, diminuindo assim o risco de overfitting. No entanto, como a DenseNet é menos propensa ao enviezamento do modelo, a abordagem acaba apenas "atrasando"o treinamento, sem ter impacto na performance do modelo final (HUANG; LIU; MAATEN, 2018; SKANSI, 2018).

Com esses últimos resultados, foi visto que, embora não tenha conseguido um resultado satisfatório para aplicação no diagnóstico de Comprometimento Cognitivo Leve, o modelo ótimo reproduzido nesse trabalho dá-se pela utilização da rede DenseNet, com transferência de aprendizado, utilizando os dados sem balanceamento e sem *Data Augmentation*. Os detalhes da rede podem ser vistos na Figura 17.

Figura 17 – Rede final

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
densenet169 (Functional)     (None, 7, 7, 1664)        12642880
_________________________________________________________________
dropout (Dropout)            (None, 7, 7, 1664)        0
_________________________________________________________________
batch_normalization (BatchNo (None, 7, 7, 1664)        6656
_________________________________________________________________
flatten (Flatten)            (None, 81536)             0
_________________________________________________________________
dense (Dense)                (None, 512)               41746944
_________________________________________________________________
dense_1 (Dense)              (None, 3)                 1539
=================================================================
Total params: 54,398,019
Trainable params: 41,751,811
Non-trainable params: 12,646,208
_________________________________________________________________
Epoch 1/80
```

Fonte: O autor.

6 Conclusão

Motivado por avanços no diagnóstico automatizado de Alzheimer por Inteligência Artificial, esse trabalho visou explorar diferentes técnicas de Deep Learning utilizadas na área de classificação de imagens procurando entender a influência que tais técnicas teriam no diagnóstico precoce de Alzheimer (Comprometimento Cognitivo Leve) por meio de imagens MRI estruturais. Com os resultados e discuessões levantadas foi possível concluir que muitas técnicas comuns na utilização de classificação de imagens podem atuar de forma danosa quando aplicadas ao contexto de saúde, além de evidenciar a necessidade de que cada uma dessas técnicas merecem ser consideradas e analisadas no contexto de identificação da classificação em específico, ao passo que o diagnóstico de cada doença difere em estrutura, forma e padrão. (SINGH et al., 2020; MORID; BORJALI; FIOL, 2021)

Embora sem ter alcançado grandes valores de métricas que permitam dizer que algum dos modelos criados seja plenamente satisfatório no diagnóstico das classes, o trabalho permitiu mostrar as influência negativas e positivas de cada técnica dentro do contexto de imagens médicas até que se chegasse num modelo otimizado. Tal modelo evidenciou a importância do método de Transfer Learning que pôde melhorar a performance do modelo ao transferir aprendizado de um dataset de diferente domínio mas com um volume maior de dados disponíveis. Esse resultado é de grande valia ao passo que foi possível constatar a eficácia de um método que pôde suprir o gargalo de escassez de imagens nos datasets de imagens médicas.

Quando iniciado o aprendizado do zero, a DenseNet mostrou-se grandemente impactada pela quantidade de dados disponíveis no dataset, ocasionando em um mal desempenho que tornou até a análise de seus parâmetros inconclusiva, após a transferência de aprendizado houve uma evolução notável de performance, o que evidenciou a sensibilidade dessa rede em relação ao volume de dados, tornando um potencial fator de limitação na aplicação da DenseNet no diagnóstico de pacientes com Comprometimento Cognitivo Leve.

A técnica de Data Augmentation mostrou-se não tão efetiva para o uso em imagens médicas no contexto do trabalho, o que pode ser explicado pela diferença de características ao ser submetida na classificação de imagens médicas, nas quais ínfimos padrões podem ser a fronteira de decisão no diagnóstico da doença, ao aplicar a técnica de DA tais padrões podem não ser alimentados no treinamento do modelo ou até mesmo confundi-lo ao aplicar as transformações. (HOSSEINI-ASL; KEYNTON; EL-BAZ, 2016; SUK; LEE; SHEN, 2014)

6.1 Considerações

Devido à sensibilidade do modelo ao tamanho do dataset, aumentar o número de imagens para treinamento do modelo seria um passo fundamental para buscar a otimização da performance. A dificuldade de encontrar datasets públicos disponíveis de imagens médicas que sejam extensos, prejudicou a realização da tarefa de forma plenamente satisfatória, no entanto existem soluções possíveis qie possam suprir esse déficit. Consultando a literatura é possível notar um grande número de trabalhos em que o autor tinha domínio sobre a aquisição dos dados (HUSSAIN et al., 2018a; KLINGELHOEFER; ROUSSEAU, 2017; HOSSEINI-ASL; KEYNTON; EL-BAZ, 2016), podendo construir o dataset de forma otimizada para o modelo em questão, escolhendo características e quantidade de imagens necessárias. Embora tenha sido inviável para a realização edsse projeto, tal abordagem seria de grande valia para a otimização do mesmo. (BRON et al., 2015)

Além da coleta de dados outra forma de suprir a falta de imagens seria por meio da geração de imagens sintéticas a partir das já existentes, especialmente na classe minoritária, o que lidaria também com o desbalancemento entre as classes. O método de balanceamento por meio do aumento de exemplos da classe minoritária é conhecido como Oversample e conta com várias formas de realização dessa tarefa, algumas mais simples como duplicação dos dados já existentes e outras formas mais complexas que fazem até usos de algoritmos focados totalmente na criação de dados sintéticos, esse último grupo tem sido representado e conseguido grandes resultados por meio da aplicação das chamadas Redes Adversárias Generativas (LECUN; BENGIO; HINTON, 2015). No trabalho proposto por Mariani et al. (2018), foi utilizado uma rede GAN como uma ferramenta de aumento para restaurar o balanceamento no conjuntos de dados. Um dos pontos desafiantes que o autor cita para essa abordagem está no fato de que as poucas imagens de classes minoritárias podem não ser suficientes para treinar uma GAN. No entanto é constatado que foi possível superar tal problema incluindo durante o treinamento adversário todas as imagens disponíveis das classes majoritárias e minoritárias. Nessa abordagem o modelo gerador aprende recursos de classes majoritárias e que são úteis para gerar imagens das classes menos frequentes. (MARIANI et al., 2018)

A falta de disponibilidade a uma infraestrutura melhor, especialmente o acesso a melhores GPUs tornou o progresso do modelo tarefa mais difícil, isso porque a criação de redes com mais parâmetros, aumentando os valores de L e k ficou impossibilitado por esse déficit, tais parâmetros de acordo com Huang, Liu e Maaten (2018) está correlacionado com a performance do modelo para reconhecimento de padrões mais complexos. A infraestrutura também impossibilitou uma primeira abordagem pretendida nesse projeto que seria utilizar volumes MRI ao invés de transformá-los em slices

2D como foi feito, a GPU limitada não permitiu o treinamento de volumes de forma minimamente viável. De acordo com Hosseini-Asl, Keynton e El-Baz (2016), trabalhar na classificação de volumes por pacientes agrega mais informação que pode ser útil e são perdidas quando ocorre a segmentação e classificação por patches em imagens 2D como foi feito no presente trabalho. No entanto isso demandaria um elevado custo computacional. (HOSSEINI-ASL; KEYNTON; EL-BAZ, 2016)

Referências

AGGARWAL, Charu C. *Neural Networks and Deep Learning*. Yorktown Heights: Springer, 2018. 512 p. Acesso em: 17 mai. 2020.

ALMEIDA, M.C; GOMES, M.C; NASCIMENTO, L.F. Spatial distribution of deaths due to alzheimer's disease in the state of são paulo, brazil. *Medical Journal*, p. 199–204, 2014. Acesso em: 12 mai. 2020.

ALVES, Gisely. Understanding convnets (cnn). Medium, p. 1, 2018. Acesso em: 05 dez. 2020.

BARRETO, Jorge M. Indrodução às redes neurais artificiais. p. 1–57, 2002.

BISHOP, Chris M. Neural networks and their applications. p. 30, 1994. Acesso em: 23 nov. 2020.

BRON, Esther E. et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural mri: the caddementia challenge. p. 1–51, 2015. Acesso em: 17 mai. 2020.

BROWNIEE, Jason. How do convolutional layers work in deep learning neural networks? 2019. Disponível em: <https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/>. Acesso em: 05 mar. 2021.

CHOLLET, Francois. Xception: Deep learning with depthwise separable convolutions. p. 1800–1807, 07 2017.

FERNANDES, Janaína da Silva Gonçalves; ANDRADE, Márcia Siqueira de. Revisão sobre a doença de alzheimer: Diagnóstico, evolução e cuidados. *Sociedade Portuguesa de Psicologia da Saúde - SPPS*, p. 10, 2017. Acesso em: 12 mai. 2020.

FERRARINI, Maria Cristina; HAGE, Nunes Soares; IWASAKI, Masao. Imagem por ressonância magnética: princípios básicos. *Ciência Rural*, p. 1288–1294, 2009. Acesso em: 16 nov. 2020.

FRISONI, Giovanni B. et al. The clinical use of structural mri in alzheimer disease. *National Institutes of Health*, p. 24, 2010. Acesso em: 29 nov. 2020.

GHOSH, Anirudha et al. Fundamental concepts of convolutional neural network. In: ____. [S.l.: s.n.], 2020. p. 519–567.

GLICKMAN, Cody. Data augmentation in medical images. 2020. Disponível em: <https://towardsdatascience.com/data-augmentation-in-medical-images-95c774e6eaae>. Acesso em: 12 out. 2020.

HAYDEN, Michael; NACHER, Pierre-Jean. History and physical principles of mri. *CRC press*, p. 12, 2015. Acesso em: 26 nov. 2020.

HOSSEINI-ASL, Ehsan; KEYNTON, Robert; EL-BAZ, Ayman. Alzheimer's disease diagnostics by adaptation of 3d convolutional network. *IEEE*, Louisville, p. 5, 2016. Acesso em: 17 mai. 2020.

HUANG, Gao; LIU, Zhuang; MAATEN, Laurens van der. Densely connected convolutional networks. p. 9, 2018. Acesso em: 10 mai. 2020.

HUSSAIN, Zeshan et al. Differential data augmentation techniques for medical imaging classification tasks. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, p. 7, 04 2018.

____. ____. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, p. 979–984, 04 2018.

JANOCHA, Katarzyna; CZARNECKI, Wojciech. On loss functions for deep neural networks in classification. *Schedae Informaticae*, v. 25, p. 1–10, 2017.

JOHNSON, Keith A. et al. Brain imaging in alzheimer disease. *Springer*, p. 1–10, 2012. Acesso em: 25 nov. 2020.

KHANNEL, Mukul. Paper review: Densenet - densely connected convolutional networks. 2019. Disponível em: <https://towardsdatascience.com/paper-review-densenet-densely-connected-convolutional-networks-acf9065dfefb>. Acesso em: 12 mai. 2020.

KHVOSTIKOVA, Alexander et al. 3d inception-based cnn with smri and md-dti data fusion for alzheimer's disease diagnostics. p. 13, 2018. Acesso em: 14 mai. 2020.

KLINGELHOEFER, Felix; ROUSSEAU, François. Convolutional autoencoder for mri modality synthesis. *INTERNSHIP REPORT*, p. 14, 2017. Acesso em: 17 mai. 2020.

KOROLEV, Sergey et al. Residual and plain convolutional neural networks for 3d brain mri classification. *IEEE*, p. 835–838, 2017. Acesso em: 14 mai. 2020.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. Toronto, p. 9, 2012. Acesso em: 13 mai. 2020.

LAI, Matthew. À margem da lei: o programa comunidade solidária. p. 1–23, 2015. Acesso em: 14 mai. 2020.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *Nature*, p. 10, 2015. Acesso em: 14 mai. 2020.

LIU, Manhua; CHENG, Danni; YAN, Weiwu. Classification of alzheimer's disease by combination of convolutional and recurrent neural networks using fdg-pet images. p. 12, 2018. Acesso em: 12 mai. 2020.

LOBO, Luiz Carlos. Inteligência artificial, o futuro da medicina e a educação médica. Brasília, p. 6, 2018. Acesso em: 15 mai. 2020.

LONG, Dan et al. Automatic classification of early parkinson's disease with multi-modal mr imaging. *PLoS ONE*, p. 9, 2012. Acesso em: 12 mai. 2020.

LUNDERVOLD, Alexander Selvikvåg; LUNDERVOLD, Arvid. An overview of deep learning in medical imaging focusing on mri. *Elsevier*, p. 102–119, 2018. Acesso em: 22 nov. 2020.

MARCUS, Daniel et al. Open access series of imaging studies: Longitudinal mri data in nondemented and demented older adults. *Journal of cognitive neuroscience*, p. 2677–84, 11 2009.

MARIANI, Giovanni et al. Bagan: Data augmentation with balancing gan. 03 2018.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous. Springer, p. 115–133, 1943. Acesso em: 24 nov. 2020.

MORID, M. A.; BORJALI, A.; FIOL, Guilherme Del. A scoping review of transfer learning research on medical image analysis using imagenet. *Computers in biology and medicine*, v. 128, p. 104115, 2021.

MOVEMENT, ALZHEIMERS IMPACT. Early detection and diagnosis of alzheimer's dementia. p. 1–6, 2017. Acesso em: 12 mai. 2020.

NIELSEN, Michael. Neural networks and deep learning. Determination Press, p. 224, 2018. Acesso em: 23 nov. 2020.

NITZSCHE, Bárbara Oliveira; MORAES, Helena Providelli de; JÚNIOR, Almir Ribeiro Tavares. Alzheimer's disease: new guidelines for diagnosis. p. 7, 2015. Acesso em: 13 mai. 2020.

PATRALEKH, Mohit Kumar; KALRA, Mukesh. Basics of magnetic resonance imaging. *Springer*, p. 17, 2012. Acesso em: 26 nov. 2020.

PATTERSON, Josh; GIBSON, Adam. Deep learning: A practitioner's approach. *O'Reilly*, p. 27–32, 2017.

PONTI, Moacir A.; COSTA, Gabriel B. Paranhos da. Como funciona o deep learning. São Bernardo do Campo, p. 31, 2017. Acesso em: 15 mai. 2020.

QIU, Shangran et al. Fusion of deep learning models of mri scans, mini–mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimer's Dementia: Diagnosis, Assessment Disease Monitoring*, 09 2018.

RAGHU, Maithra et al. Transfusion: Understanding transfer learning with applications to medical imaging. *CoRR*, 2019.

RAVì, Daniele et al. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 2016.

RAZZAK, Muhammad Imran; NAZ, Saeeda; ZAIB, Ahmad. Deep learning for medical image processing: Overview, challenges and future. *Springer*, p. 1–30, 2017. Acesso em: 25 nov. 2020.

RIOS, Eduardo Diaz. TÉcnica de diagnÓstico por imagens: RessonÂncia magnÉtica nuclear. *UFRGS*, p. 73, 1998. Acesso em: 15 nov. 2020.

SAHA, Rohan. Transfer learning - a comparative analysis. p. 5–15, 12 2018.

SARKAR, Dipanjan. A comprehensive hands-on guide to transfer learning with real-world applications in deep learning. 2018. Disponível em: <https://towardsdatascience.com/a-comprehensive-hands-on-guide\-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>. Acesso em: 13 fev. 2021.

SARRAF, Ghassem Tofighi Saman. Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. p. 5, 2016. Acesso em: 17 mai. 2020.

SHI, Jun et al. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, p. 1–1, 01 2017.

SINGH, Satya P. et al. 3d deep learning on medical images: A review. p. 13, 2020. Acesso em: 17 mai. 2020.

SKANSI, Sandro. *Introduction to Deep Learning*: From logical calculus to artificial intelligence. Zagreb: Springer, 2018. 196 p. Acesso em: 17 mai. 2020.

SUK, Heung-Il; LEE, Seong-Whan; SHEN, Dinggang. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 07 2014.

SZEGEDY, Christian et al. Rethinking the inception architecture for computer vision. 06 2016.

WANG, Shuihua et al. Classification of alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *Journal of Medical Systems*, 03 2018.

WEN, Junhao et al. Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. Paris, p. 68, 2020. Acesso em: 15 mai. 2020.

WRIGHT, M. et al. Introduction to dicom for the practicing veterinarian. *Veterinary radiology & ultrasound : the official journal of the American College of Veterinary Radiology and the International Veterinary Radiology Association*, 2008.

YANG, Chengliang; RANGARAJAN, Anand; RANKA, Sanjay. Visual explanations from deep 3d convolutional neural networks for alzheimer's disease classification. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, v. 2018, 03 2018.

ZOU, LIANG et al. 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. *IEEE*, p. 11, 2017. Acesso em: 17 mai. 2020.