



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Victor Hugo Rocha de Oliveira

**Aprendizado Profundo para Predição da Idade Cerebral utilizando Imagens de
Ressonância Magnética Estrutural**

Florianópolis
2023

Victor Hugo Rocha de Oliveira

Aprendizado Profundo para Predição da Idade Cerebral utilizando Imagens de Ressonância Magnética Estrutural

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de mestre em Engenharia Elétrica.

Orientador: Prof. Danilo Silva, Ph.D.

Florianópolis
2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Oliveira, Victor

Aprendizado Profundo para Predição da Idade Cerebral
utilizando Imagens de Ressonância Magnética Estrutural /
Victor Oliveira ; orientador, Danilo Silva, 2023.

73 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia Elétrica, Florianópolis, 2023.

Inclui referências.

1. Engenharia Elétrica. 2. Aprendizado Profundo. 3.
Inteligência Artificial. 4. Neuroimagem. 5. Redes Neurais.
I. Silva, Danilo. II. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Engenharia Elétrica.
III. Título.

Victor Hugo Rocha de Oliveira

Aprendizado Profundo para Predição da Idade Cerebral utilizando Imagens de Ressonância Magnética Estrutural

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Raimes Moraes, Ph.D.
Universidade Federal de Santa Catarina

Prof Jônata Tyska Carvalho, Ph.D.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Engenharia Elétrica.

Telles Brunelli Lazzarin, Ph. D.
Coordenador do Programa de
Pós-Graduação em Engenharia Elétrica

Prof. Danilo Silva, Ph.D.
Orientador

Florianópolis, 2023.

Este trabalho é dedicado aos meus queridos pais.

AGRADECIMENTOS

A Deus por toda força, saúde, bênçãos e proporcionar momentos especiais durante essa jornada muito importante em minha vida.

Aos meus pais, por todo amor, apoio, conselhos e conforto em momentos difíceis. Vocês são meus heróis.

Ao professor Danilo Silva, por toda paciência, atenção e orientações importantes que proporcionaram formação de habilidades profissionais e pessoais durante essa jornada.

Ao Grupo 3778 e Grupo Alliar, especialmente Augusto Antunes, Ricardo Giglio, Saulo Pedro, pelo incentivo à pesquisa, divulgação científica e participação no projeto.

Agradeço a Alexandre Soares e Bruno Pacheco, que desempenharam papéis fundamentais para o desenvolvimento do trabalho.

A Karine por toda paciência, companheirismo e amor durante essa longa jornada.

Aos amigos da UFSC que também vieram de outras regiões e compartilharam momentos de união, amizade e diversas noites em claro estudando e batalhando para atingir nosso propósito: Adriele Carvalho, Marcus Branch e Raul Nascimento.

Aos amigos que sempre me apoiaram e me fizeram acreditar: Breno Guss, Caio César Caldeira, Guilherme Castro Oliveira, Igor Araújo, Mateus Miranda, Mateus Wenceslau, Miller Oliveira, Pedro Miranda, Vinicius Abreu.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pelo apoio financeiro.

*“Tudo que eu posso te dar é o estudo”
(Maria Helena e José Ladir)*

RESUMO

Estudos relacionados ao envelhecimento natural do cérebro tem demonstrado diversas causas que acometem o sistema nervoso central, causando declínio cognitivo e possíveis doenças relacionadas à idade cerebral. Atualmente, radiologistas estimam a idade cerebral por meio de diversas análises e estudos do paciente, sendo uma tarefa complexa e que demanda expertise. Técnicas de Aprendizado Profundo podem auxiliar equipes médicas na solução deste problema. Modelos de Redes Neurais Convolucionais, baseados em reconhecimento de padrões em imagens, são amplamente utilizados na literatura, tendo mostrado promissores para estimar a idade cerebral de um paciente por meio de neuroimagens. O objetivo deste trabalho é desenvolver modelos de Redes Neurais Convolucionais para predição da idade cerebral por meio de Imagens Cerebrais de Ressonância Magnética Estrutural. Foram utilizadas imagens de Ressonância Magnética Estrutural de pacientes saudáveis pertencentes ao conjunto de dados ADNI (*Alzheimer's Disease Neuroimaging Initiative*). A ADNI é uma iniciativa composta por diversos pesquisadores que coletam, validam e utilizam dados para desenvolver estudos relacionados à progressão da doença de Alzheimer. Duas abordagens foram utilizadas: a primeira utiliza modelos pré-treinados de classificação, onde a camada classificadora é substituída com o propósito de criar um modelo de regressão que faça a predição da idade cerebral. A segunda abordagem utiliza um modelo pré-treinado na tarefa de segmentação de tumor cerebral. Um modelo de regressão é treinado e inicializado com os pesos do modelo de segmentação. Várias arquiteturas foram experimentadas na primeira abordagem; a que sobressaiu apresentou Erro Absoluto Médio de 3,57 anos e Erro Percentual Absoluto Medio de 4,72%. Na segunda abordagem, foi alcançado o Erro Absoluto Médio de 3,39 e Erro Percentual Absoluto Medio de 4,46% utilizando uma arquitetura computacionalmente mais leve. Os resultados demonstram ser promissores quando comparados a trabalhos atuais.

Palavras-chave: Aprendizado Profundo. Inteligência Artificial. Neuroimagem. Redes Neurais.

ABSTRACT

Studies related to the natural aging of the brain have demonstrated several causes that affect the central nervous system, causing cognitive decline and possible diseases related to brain age. Currently, radiologists estimate brain age through various analyses and patient studies, a complex task that demands expertise. Deep Learning techniques can help medical teams to solve this problem. Convolutional Neural Network models, based on image pattern recognition, are widely used in the literature and have demonstrated good performance for estimating the brain age of a patient through neuroimaging. The objective of this work is to develop Convolutional Neural Network models for predicting brain age using Structural Magnetic Resonance Imaging. Structural MRI images from healthy patients belonging to the ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset were used. ADNI is an initiative composed of several researchers who collect, validate and use data to develop studies related to the progression of Alzheimer's disease. Two approaches were used: the first uses pre-trained classification models, where the classifier layer is replaced for the purpose of creating a regression model that makes a prediction of brain age. The second approach uses a pre-trained model in the brain tumor segmentation task. A regression model is trained and initialized with the weights from the segmentation model. Several architectures were tried in the first approach; the one that excelled showed a Mean Absolute Error of 3,57 and a Mean Absolute Percentage Error of 4,72%. In the second approach, the Mean Absolute Error of 3,39 and Mean Absolute Percentage Error of 4,46% was achieved using a computationally lighter architecture. The results show promise when compared to current works.

Keywords: Deep Learning. Artificial Intelligence. Neuroimaging. Neural Networks.

LISTA DE FIGURAS

Figura 1 – Arquitetura de um Neurônio.	20
Figura 2 – Rede Neural FeedForward com duas camadas ocultas.	22
Figura 3 – Efeito da escolha do passo na convergência na descida do gradiente. Em cima - passo muito grande, o que pode ultrapassar o mínimo e, possivelmente, não convergir; em baixo - um passo muito pequeno que faz com que a descida do gradiente venha a convergir para o mínimo lentamente.	24
Figura 4 – Operação de correlação cruzada. As partes sombreadas são o primeiro elemento de saída, bem como os elementos tensores de entrada e kernel usados para o cálculo de saída:	25
Figura 5 – Efeito do filtro detector de bordas e ativação linear	25
Figura 6 – Operação de max pooling	26
Figura 7 – Aplicando max pooling em uma imagem	26
Figura 8 – Típica Arquitetura CNN contendo uma sequência de blocos convolucionais, compostas por operações de convolução e pooling; e camadas totalmente conectadas	27
Figura 9 – As saídas de cada camada de uma arquitetura de rede convolucional aplicada à imagem de um cachorro. A imagem demonstra a extração de recursos realizada pelos blocos convolucionais e a probabilidade do cachorro em destaque pertencer a cada classe.	27
Figura 10 – Erro de treinamento (esquerda) e erro de teste (direita) no dataset CIFAR-10 com redes planas de 20 e 56 camadas	28
Figura 11 – Bloco residual contendo o somatório de $f(x)$ com camadas de pesos unitários.	29
Figura 12 – Representação do Problema de Vanish Gradient, destacando a região em que o gradiente de f é quase nulo	29
Figura 13 – Representação de uma rede neural comum (esquerda) e uma rede neural com blocos residuais (direita). Representa dados que passam por toda a rede, eliminando o problema de <i>vanish gradient</i> , devido as <i>skip connections</i>	30
Figura 14 – Representação da Arquitetura InceptionResnetV2	31
Figura 15 – Esquema para módulos de grade 35×35 da rede Inception-v4 pura.	32

Figura 16 – (a) modelo sem nenhum método de dimensionamento composto aplicado; (b)-(d) dimensionamentos convencionais contendo técnicas que redimensionam a largura, profundidade ou resolução da rede. (e) é o método de dimensionamento composto que redimensiona uniformemente todas as três dimensões (largura, profundidade e resolução).	33
Figura 17 – Performance da Arquitetura EfficientNet com demais Redes na ImageNet	33
Figura 18 – Exemplos de Técnicas de Segmentação	34
Figura 19 – Arquitetura U-Net	36
Figura 20 – Ilustração dos 3 planos padrão de uma imagem de Ressonância Magnética Estrutural. Plano sagital (vista superior-inferior e anterior-posterior), axial (vista anterior-posterior e esquerda-direita), coronal (vista superior-inferior e esquerda-direita).	38
Figura 21 – Imagens de ressonância magnética estrutural em diferentes modalidades, demonstrando a diferença de contraste entre as regiões da estrutura do cérebro.	38
Figura 22 – Ilustração do processo de extração do cérebro (<i>skull stripping</i>). Imagem original (esquerda); extração do cérebro mal feita (meio); boa extração do cérebro (direita).	39
Figura 23 – Representação do atlas MNI152 - apenas cérebro	40
Figura 24 – Distribuição de idades (N=2751)	44
Figura 25 – Ilustração do Processo de Corte	47
Figura 26 – Ilustração do Processo de Transformação em Fatias 2D	48
Figura 27 – Representação da Arquitetura do Modelo	49
Figura 28 – Representação da Arquitetura proposta	50
Figura 29 – Processo de Treinamento	51
Figura 30 – Ajuste dos Hiperparâmetros em Grade. Os blocos destacados representam uma das 9 diferentes combinações. a) Representação da <i>learning rate</i> de 10^{-3} e <i>batch size</i> de tamanho 16; b) Learning rate de 10^{-3} e batch size de tamanho 32	52
Figura 31 – Processo de Inferência	53
Figura 32 – Processo de Correção do Viés das Predições, representando o Fold 1.	55
Figura 33 – Comparação entre um modelo ResNet50 com inicialização de pesos pré-treinados na ImageNet e um modelo com inicialização de pesos aleatórios	56
Figura 34 – Loss de treino e de validação - ResNet50	57
Figura 35 – Resultado da Inferência no Conjunto de Teste - Arquitetura ResNet50 retreinada.	59

Figura 36 – Erros residuais no Conjunto de Teste - ResNet50	60
Figura 37 – Loss de treino e de validação - UNet	61
Figura 38 – Resultado do teste utilizando a arquitetura U-Net	62
Figura 39 – Erros residuais - U-Net	62
Figura 40 – Resultado da Inferência no Conjunto de Dados AIBL - Arquitetura U-Net retreinada.	63
Figura 41 – Erros residuais da Inferência no Conjunto de Dados AIBL - Arquite- tura U-Net retreinada.	63

LISTA DE TABELAS

Tabela 1 – Tuning da Taxa de aprendizado x Tamanho do Lote - ResNet-50 . . .	57
Tabela 2 – Resultados do Teste de Correção de Viés aplicado em 5 Folds. . . .	58
Tabela 3 – Comparação de resultados obtidos no conjunto de validação: ResNet-50, ResNet-101, InceptionResnetV2 e EfficientNet-B4.	58
Tabela 4 – Resultado do modelo ResNet-50 retreinado, obtido no conjunto de teste.	59
Tabela 5 – Tuning da Taxa de aprendizado x Tamanho do Lote - UNet.	60
Tabela 6 – Resultados da arquitetura U-Net no conjunto de validação.	61
Tabela 7 – Resultado do modelo U-Net retreinado, obtido no conjunto de teste.	61
Tabela 8 – Resultado da Inferência no Conjunto de Dados AIBL - Arquitetura U-Net retreinada.	61
Tabela 9 – Comparação entre U-Net Proposta e Modelos do Estado da Arte . .	64

LISTA DE ABREVIATURAS E SIGLAS

CNN	Convolutional Neural Network
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MRI	Magnetic Ressonance Imaging
MSE	Mean Squared Error
ReLU	Rectified Linear Unit
RNA	Redes Neurais Artificiais

SUMÁRIO

1	INTRODUÇÃO	16
1.1	OBJETIVOS	17
1.1.1	Objetivo Geral	17
1.1.2	Objetivos Específicos	18
1.2	CONTRIBUIÇÕES	18
1.3	ORGANIZAÇÃO DO TRABALHO	18
1.4	PUBLICAÇÕES	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	APRENDIZADO PROFUNDO	20
2.1.1	Redes Neurais Artificiais	20
2.1.2	Backpropagation	21
2.1.3	Redes Neurais Convolucionais	23
2.1.4	Redes Neurais Residuais	28
2.1.5	InceptionResnetV2	30
2.1.6	EfficientNet	32
2.1.7	U-Net	33
2.2	ENVELHECIMENTO CEREBRAL	36
2.3	IMAGENS DE RESSONÂNCIA MAGNÉTICA	37
2.4	PRÉ-PROCESSAMENTO	39
3	TRABALHOS RELACIONADOS	41
4	MÉTODOS	44
4.1	DATASET	44
4.2	PRÉ-PROCESSAMENTO	45
4.2.1	Co-Registro	45
4.2.2	Skull Stripping	46
4.2.3	Pipeline de Pré-Processamento da nnU-Net	46
4.2.4	Normalização de Modelos Pré-Treinados na ImageNet	47
4.2.5	Seleção de fatias	48
4.3	ABORDAGENS DE REDES NEURAIIS UTILIZADAS	48
4.3.1	Modelos Pré-Treinados em Classificação	48
4.3.2	Modelo Pré-Treinado em Segmentação	49
4.3.3	Processo de Treinamento	51
4.3.4	Otimização de Hiperparâmetros	51
4.4	PÓS-PROCESSAMENTO	52
4.4.1	Fusão das Predições	53
4.4.2	Correção do Viés nas Predições	53
4.4.3	Avaliação de Desempenho em um Conjunto de Dados Externo	54

4.4.4	Recursos Computacionais	54
5	RESULTADOS	56
5.1	ABORDAGEM UTILIZANDO MODELOS PRÉ-TREINADOS DE CLAS- SIFICAÇÃO	56
5.1.1	Comparação de métodos de inicialização dos pesos	56
5.1.2	Ajuste de Hiperparâmetros	56
5.1.3	Correção de Viés	57
5.1.4	Avaliação de Desempenho	58
5.2	ABORDAGEM UTILIZANDO MODELO PRÉ-TREINADO DE SEG- MENTAÇÃO	59
5.2.1	Ajuste de Hiperparâmetros	59
5.2.2	Avaliação de Desempenho	60
5.3	AVALIAÇÃO NO CONJUNTO DE DADOS AIBL	61
5.4	DISCUSSÃO	64
6	CONCLUSÃO	66
6.1	TRABALHOS FUTUROS	66
	REFERÊNCIAS	67

1 INTRODUÇÃO

A velhice causa uma maior probabilidade de problemas cognitivos nas pessoas, uma condição que até agora não pode ser evitada, nem curada ANDERSEN-RANBERG; VASEGAARD; JEUNE (2001). Pesquisas relacionadas a esse tema estão sendo desenvolvidas com o objetivo de entender melhor o comportamento do envelhecimento natural do cérebro.

O envelhecimento cerebral é caracterizado como um processo biológico causado pelo acúmulo de danos moleculares e celulares ao longo da vida (WRIGGLESWORTH *et al.*, 2021). A incapacidade do organismo de reparar os danos leva a uma subsequente perda das funções fisiológicas. Isso causa impacto diretamente em funções sensoriais, motoras e cognitivas que, quando comprometidas, afetam a qualidade de vida da pessoa.

Os cérebros de indivíduos que são cognitivamente normais apresentam alterações relacionadas à idade que incluem uma redução geral no volume e peso cerebral (ANDERTON, 1997). Essas alterações na estrutura estão associadas à atrofia cerebral, alargamento dos sulcos do córtex cerebral e aumento dos ventrículos cerebrais, causando a perda do volume cerebral e possíveis problemas cognitivos. Há evidências de que doenças como Alzheimer ou Esquizofrenia são associadas ao envelhecimento cerebral acelerado (WRIGGLESWORTH *et al.*, 2021).

No contexto médico, o papel da inteligência artificial é transformar dados médicos referentes a um determinado paciente ou procedimento em informações que auxiliam as equipes médicas nas tomadas de decisão. Segundo (TOPOL, 2019), o uso da inteligência artificial na medicina destina-se no auxílio em diagnósticos e interpretação de testes, recomendação de tratamentos, dentre outras aplicações essenciais que envolvem cuidados com a saúde.

O Aprendizado Profundo é um subconjunto do aprendizado de máquina que permite aprender e reconhecer padrões em conjunto de dados por meio de estruturas de Redes Neurais Artificiais (RNAs). As RNAs são modelos conexionistas que consistem em receber N valores de entrada (x_1, x_2, \dots, x_n) e associá-los a uma saída (y) , interligando as saídas em camadas sucessivas de processamento. Esses modelos aprendem com pesos (w_1, w_2, \dots, w_n) e computam a saída $f(x, w) = x_1 w_1, x_2 w_2, \dots, x_n w_n$, segundo GOODFELLOW; BENGIO; COURVILLE (2016).

Em RAJPURKAR *et al.* (2017), foi utilizado um modelo de aprendizado profundo capaz de realizar o diagnóstico de pneumonia por meio de imagens de raio-X do tórax. Além do diagnóstico, o modelo demonstra na imagem, por meio de mapa de calor, as regiões que o modelo definiu com maior ponderação para a classificação da doença.

Outra tarefa em que o uso de aprendizado profundo se destaca é a segmentação de imagens médicas. Um exemplo é a competição BraTS MENZE *et al.* (2015) que

tem o objetivo de reunir pesquisadores para atuar na tarefa de segmentação de tumor cerebral por meio de imagens de ressonância magnética. Nessa competição, diversas pessoas trabalham para alcançar o estado da arte de algoritmos de segmentação de tumor cerebral.

Em COLE *et al.* (2017), BASHYAM *et al.* (2020), POLONI; FERRARI; INITIATIVE *et al.* (2022), foram desenvolvidos estudos que abordam o tema da predição da idade cerebral. Os trabalhos mencionados demonstram a capacidade de um modelo de aprendizado profundo em prever a idade cronológica de pacientes normais (livres de qualquer comprometimento cognitivo) a partir de imagens do cérebro. Imagens de Ressonância Magnética cerebrais foram utilizadas nestes estudos como dados de entrada em modelos de redes neurais que estimam a idade cronológica dos pacientes. Essas imagens possuem diferentes tipos de aquisição, cada uma permite destacar diferentes propriedades físicas do cérebro. Devido as diferentes informações de tecido, estrutura e resolução espacial, o modelo pode detectar detalhes e mudanças sutis nos tecidos cerebrais dos pacientes (POLONI; FERRARI; INITIATIVE *et al.*, 2022).

Os resultados dos trabalhos de COLE *et al.* (2017) e BASHYAM *et al.* (2020) demonstram que modelos de aprendizado profundo apresentam resultados promissores para estimar a idade cronológica a partir de imagens de ressonância magnética de pacientes saudáveis. No entanto, é importante salientar que alguns desses estudos utilizaram conjuntos de dados privados. A utilização de conjuntos de dados privados pode gerar viés nas predições obtidas, uma vez que a falta de transparência e possibilidade de validação ampla por outros pesquisadores podem comprometer a confiabilidade dos resultados.

Com base nisso, o presente trabalho tem como proposta utilizar um conjunto de dados público, o que permite maior transparência e possibilidade de reprodução dos resultados. Além disso, será investigada a robustez do modelo quando aplicado em um conjunto de dados de diferente distribuição. Outra abordagem a ser explorada é a utilização de modelos pré-treinados em uma tarefa de segmentação de tumor cerebral, com o objetivo de aplicá-los posteriormente na tarefa de predição da idade cerebral.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O principal objetivo dessa dissertação é desenvolver um modelo de aprendizado de máquina capaz de estimar a idade cerebral por meio de imagens de ressonância magnética, fazendo uso de um conjunto de dados disponível publicamente.

1.1.2 Objetivos Específicos

- Selecionar base de dados com imagens de Ressonância Magnética Estrutural de pacientes diagnosticados como livres de qualquer comprometimento cognitivo (pacientes controle);
- Realizar preparação de dados com passos específicos da área de análise de imagens de ressonância magnética;
- Treinar e avaliar o desempenho de modelos pré-treinados na ImageNet para a tarefa de predição da idade cerebral;
- Treinar um modelo de segmentação utilizando neuroimagens para posteriormente aplicá-lo na tarefa de predição da idade cerebral;
- Validar os resultados;
- Avaliar o desempenho do modelo em um conjunto de dados externo.

1.2 CONTRIBUIÇÕES

O presente trabalho tem a motivação de desenvolver um método capaz de estimar a idade de sujeitos diagnosticados como saudáveis, ou seja, livres de problemas cognitivos. O método proposto foi o uso de Redes Neurais Convolucionais 2D para predição da idade cerebral, principalmente pelo fato de serem modelos mais simples que as redes 3D. Duas abordagens foram propostas, a primeira é uma abordagem existente na literatura, onde são utilizados modelos classificadores, pré-treinados na ImageNet, para a tarefa de predição da idade cerebral. Como proposta de inovação, uma outra abordagem é realizada: o pré-treino em imagens cerebrais na tarefa de segmentação, visando posteriormente utilizar este modelo na tarefa de predição da idade cerebral. Alguns trabalhos do estado da arte utilizam conjuntos privados; neste trabalho foi utilizado um conjunto de dados público, tornando possível o acesso e a reprodução dos resultados. O repositório do projeto pode ser acessado por meio do endereço <https://github.com/gama-ufsc/brain-age/tree/dissertacao-victor>.

1.3 ORGANIZAÇÃO DO TRABALHO

Essa dissertação está organizada da seguinte forma:

- No capítulo 2, introduzimos conceitos necessários para entendimento do trabalho. São abordados temas relacionados ao aprendizado profundo, arquiteturas de Redes Neurais Convolucionais e conceitos de segmentação de imagens. Também apresenta tópicos para entendimento básico de neuroimagens;

- No capítulo 3, são apresentados alguns trabalhos do estado da arte. Esse capítulo é um compilado das principais técnicas e resultados de cada trabalho presente na literatura;
- O capítulo 4 apresenta os métodos utilizados para desenvolvimento do trabalho. Esse capítulo contém informações do conjunto de dados utilizado, são descritas as etapas de pré e pós-processamento, as abordagens de Redes Neurais utilizadas e as métricas de avaliação;
- O capítulo 5 é constituído pelos resultados obtidos em cada experimento mencionado na metodologia. Além disso, o capítulo apresenta discussões sobre esses resultados, bem como outros experimentos realizados;
- No capítulo 6, são elencadas as conclusões finais e sugestões para trabalhos futuros.

1.4 PUBLICAÇÕES

Durante o decorrer deste trabalho, foram feitas duas contribuições para a comunidade científica relacionada ao tema proposto:

Predição da Idade Cerebral a partir de Imagens de Ressonância Magnética utilizando Redes Neurais Convolucionais: em OLIVEIRA *et al.* (2021) são investigadas técnicas de aprendizado profundo para auxiliar na identificação de biomarcadores do processo natural de envelhecimento. Modelos de Redes Neurais Convolucionais 2D e 3D foram comparados no trabalho proposto.

Redes Neurais Convolucionais para Desenvolvimento de Biomarcador da Idade Cerebral a partir de Imagens de Ressonância Magnética nas Ponderações T1 e FLAIR: em OLIVEIRA *et al.* (2022) dois modelos foram treinados independentemente, um para cada tipo de ponderação (T1-MPRAGE e FLAIR). Após o treinamento, foi criado um novo modelo que consiste da combinação dos outros dois que foram treinados de forma independente. Este método é chamado de ensemble. Essa técnica foi escolhida para garantir que o biomarcador de predição da idade cerebral considere tanto aspectos morfológicos encefálicos quanto aspectos degenerativos crônicos que acometem o sistema nervoso central, notadamente a microangiopatia crônica.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta conceitos importantes para a melhor compreensão do trabalho. Serão abordados temas relacionados a Redes Neurais Artificiais, aprendizado profundo e sua aplicação em imagens de ressonância magnética estrutural.

2.1 APRENDIZADO PROFUNDO

2.1.1 Redes Neurais Artificiais

Um único neurônio artificial (também denominado como unidade) é a forma mais simples de uma rede neural. Baseia-se em um modelo matemático que recebe valores de entrada x_i que são associados a pesos w_i . Os valores são somados em uma função de ativação com um viés aplicado de forma externa. O resultado da função define a saída do modelo (Figura 1), conforme HAYKIN (2009).

Podemos representar matematicamente utilizando as equações (1) e (2):

$$z = \mathbf{w}^T \mathbf{x} + b, \quad (1)$$

$$h = g(\mathbf{w}^T \mathbf{x} + b) = g(z) \quad (2)$$

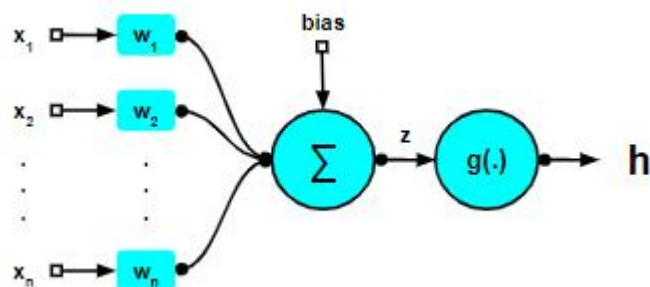
Onde \mathbf{x} corresponde ao vetor de entrada, \mathbf{w} é o vetor de peso e b representa o *bias*. A função $g(\cdot)$ é chamada de função de ativação, de acordo com (BISHOP, 2006) ela permite que a saída da unidade resulte em um valor escalar a partir de ativações decorrentes de operações não-lineares.

Dentre as funções de ativação existentes, exemplos comuns incluem:

Sigmoid: função logística utilizada para classificação binária. Produz valores na faixa de $[0, 1]$.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

Figura 1 – Arquitetura de um Neurônio.



Softmax: classificação multiclasse. Produz valores numa faixa de $[0, 1]$. Cada valor corresponde a probabilidade do dado de entrada pertencer a alguma classe. A soma das probabilidades das N classes equivale a 1.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad \forall i = 1, 2, \dots, K \quad (4)$$

ReLU (Retified Linear Unit): retorna zero para valores negativos e retorna o próprio valor no caso de valores positivos.

$$\text{Relu}(z) = \max(0, z) \quad (5)$$

A arquitetura de uma rede neural contém camadas de neurônios, cada uma conectada à seguinte; dessa forma a informação se propaga por meio das conexões até a camada de saída, \hat{y} , formando uma rede composta por camadas de entrada, camadas ocultas e a camada de saída. A Figura 2 representa um modelo de rede neural contendo duas camadas ocultas.

Esse modelo de redes neurais também pode ser chamado de Propagação Direta (do inglês, *FeedForward*). O objetivo da propagação direta é aproximar a saída estimada $\hat{y} = f(x|\theta)$ de y , sendo $\theta = \{w_{kj}^{[\ell]}, b_k^{[\ell]}\}$ os vetores contendo os parâmetros de pesos e bias do modelo. Esse modelo tem esse nome porque a informação se propaga através das camadas conectadas, sendo processada pela Equação (6) até a saída do modelo (GOODFELLOW; BENGIO; COURVILLE, 2016). Podemos chamar essas camadas de camadas densas (ou camadas totalmente conectadas).

$$\mathbf{h}^{[\ell]} = g_\ell(\mathbf{w}^{[\ell]} \mathbf{h}^{[\ell-1]} + \mathbf{b}^{[\ell]}), \quad (6)$$

onde,

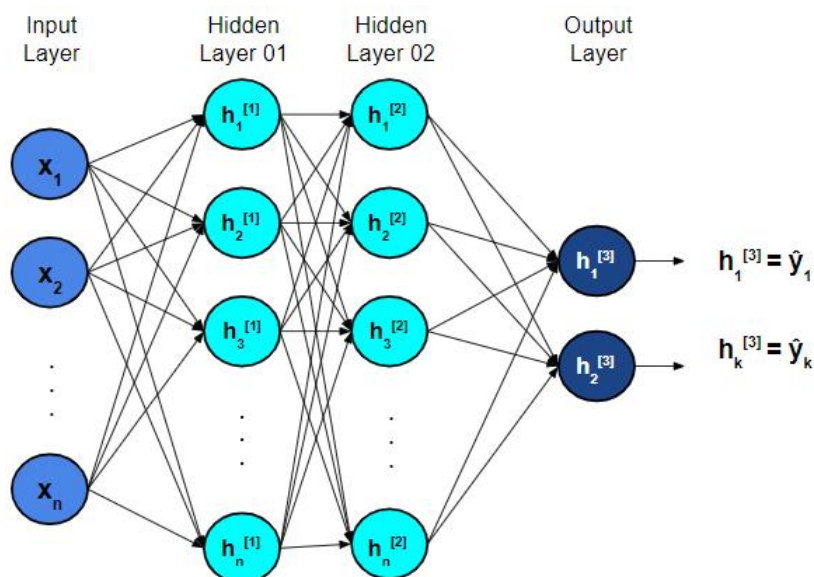
- $\ell \in \mathbb{N}^*$ é a camada correspondente;
- $\mathbf{w}^\ell, \mathbf{b}^\ell \in \mathbb{R}$ e representam os vetores de pesos e bias;
- $g_\ell \in \mathbb{R}$ representa a função de ativação da camada ℓ .
- $\mathbf{h}^\ell \in \mathbb{R}$ representa o vetor de ativações da camada ℓ .

Para ajustar os valores de w que resultam na melhor aproximação da resposta desejada, é comumente utilizado o algoritmo de *backpropagation*.

2.1.2 Backpropagation

Um modelo de aprendizado supervisionado envolve a aplicação de um conjunto de dados de treinamento rotulados e a modificação dos pesos de uma rede neural. Durante o treinamento, os pesos são modificados para minimizar a diferença entre a resposta desejada e a resposta do modelo de acordo com um critério estatístico apropriado. O treinamento da rede é repetido para muitos exemplos no conjunto, até

Figura 2 – Rede Neural FeedForward com duas camadas ocultas.



que a rede atinja um estado estacionário onde não haja mais mudanças significativas nos pesos (HAYKIN, 2009).

A função de custo é um critério importante para descrever de forma quantitativa o quão próximo as previsões chegaram do valor desejado. Ela é representada por L , na equação (7). Essa equação indica o quanto a previsão \hat{y} está incorreta quando o real valor é y :

$$L(\mathbf{y}, \hat{\mathbf{y}}), \quad \hat{\mathbf{y}} = f(\mathbf{x}|\boldsymbol{\theta}) \quad (7)$$

Para medir o desempenho de um modelo no conjunto de treinamento, o custo médio sobre todo o conjunto é calculada, conforme a equação (8).

$$J(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N L(\mathbf{y}^{(i)}, \hat{\mathbf{y}}^{(i)}), \quad \hat{\mathbf{y}} = f(\mathbf{x}^{(i)}|\boldsymbol{\theta}) \quad (8)$$

- $J(\boldsymbol{\theta})$ representa o custo médio;
- N representa o tamanho do conjunto;
- $\hat{\mathbf{y}}$ são as previsões;
- \mathbf{y} são os rótulos.

A maioria dos algoritmos de treinamento envolvem métodos iterativos para minimizar a função de custo, com ajustes nos pesos sendo feitos em uma sequência de etapas (BISHOP, 2006). O método de *backpropagation* é uma solução computacional que permite calcular derivadas e aplicar cálculos iterativos que utilizam otimização numérica para minimizar a função de custo. Esse algoritmo consiste em duas etapas:

1. Propagação: os dados de entrada passam pelas camadas da rede neural e cálculos de custo são realizados;
2. Retro-Propagação: O gradiente da função de custo da última camada é calculado. Os pesos de todos os neurônios da rede são atualizados por meio de cálculo da regra da cadeia.

De acordo com BISHOP (2006), a técnica mais simples aplicada a este método é o gradiente descendente (equação (9)).

$$w^{[\ell+1]} = w^{[\ell]} - \alpha \frac{\partial J}{\partial w_{kj}^{[\ell]}}, \quad \forall k, j, \ell \quad (9)$$

O parâmetro α é uma constante positiva que representa a **taxa de aprendizado**. Seu valor representa o tamanho do passo que controla o quanto se avança na direção do gradiente negativo a partir de um ponto inicial w_0 (WATT; BORHANI; KATSAGGELOS, 2016). O princípio do gradiente descendente é construir um modelo linear da função $g(w)$, determinar a direção descendente neste hiperplano, viajar uma distância ao longo desta direção, de acordo com o tamanho do passo, e repetir o processo até a convergência, conforme WATT; BORHANI; KATSAGGELOS (2016). Pode-se visualizar na (Figura 3) que o algoritmo começa em um ponto inicial (w_0) e o procedimento de descida do gradiente produz uma sequência de pontos (w_1, w_2, \dots, w_N), que reduz o valor de $g(w)$ a cada passo até atingir um ponto estacionário

Quanto maior o valor da taxa de aprendizado, maior a probabilidade do algoritmo ultrapassar o mínimo global e não convergir, todavia apresenta maior velocidade; em contrapartida, a menor taxa de aprendizado tende a convergir para um mínimo global de forma lenta (WATT; BORHANI; KATSAGGELOS, 2016).

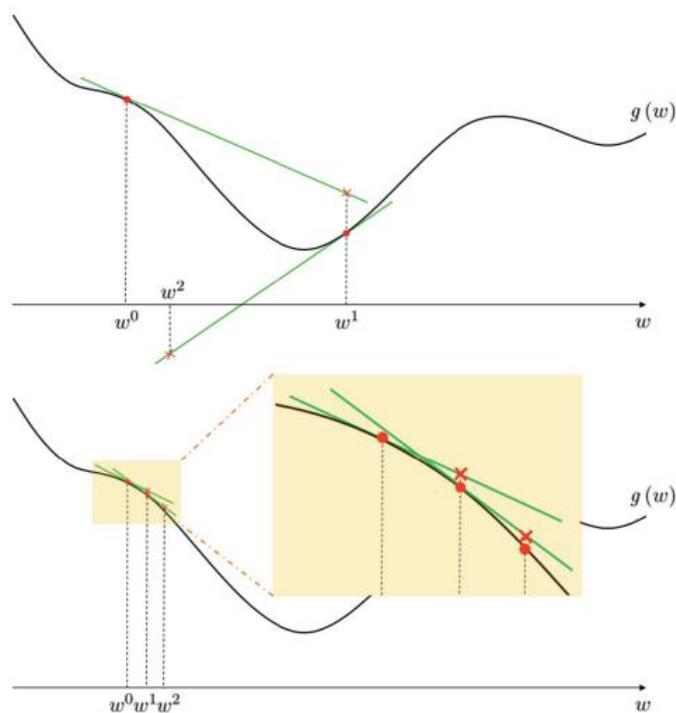
Segundo NIELSEN (2018), essa equação estabelece o quão rápido a função de custo muda ao se alterar os valores de pesos e bias, oferecendo informações detalhadas de como a rede se comporta diante tais modificações.

A partir desse algoritmo computacional, várias arquiteturas de redes neurais foram desenvolvidas. Na próxima seção, será introduzido o conceito das Redes Neurais Convolucionais: redes que são comumente utilizadas para reconhecimento de padrões em imagens, como no trabalho proposto.

2.1.3 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNNs) são análogas às Redes Neurais tradicionais, pois são compostas por neurônios e otimização por meio do aprendizado. Cada neurônio ainda receberá uma entrada e realizará uma operação (como um produto escalar seguido de uma função não linear), como é a base de inúmeras RNAs. Desde os vetores contendo *pixels* sem processamento da imagem até a saída final da pontuação da classe, toda a rede envolve funções de ponderação. As CNNs são utili-

Figura 3 – Efeito da escolha do passo na convergência na descida do gradiente. Em cima - passo muito grande, o que pode ultrapassar o mínimo e, possivelmente, não convergir; em baixo - um passo muito pequeno que faz com que a descida do gradiente venha a convergir para o mínimo lentamente.



zadas principalmente no campo de reconhecimento de padrões em imagens (O'SHEA; NASH, 2015).

Redes Neurais Convolucionais (CNN) são redes muito utilizadas em visão computacional, principalmente quando os dados de entrada são imagens, textos e áudio. Ao contrário das redes neurais de camadas densas, as quais utilizam multiplicações matriciais, as CNNs fazem o uso de operações de convolução em pelo menos uma de suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016).

A arquitetura de um modelo de redes convolucionais possui uma série de estágios (LECUN; BENGIO; HINTON, 2015). A entrada e a saída de cada estágio são conjuntos de *arrays* chamados de mapas de recursos (ou mapa de características). Por exemplo, se a entrada for uma imagem colorida, cada mapa de recursos seria um *array* 2D contendo um canal de cores da imagem de entrada; para uma entrada de áudio, cada mapa de recursos seria um *array* 1D e, para um vídeo ou imagem volumétrica, seria seria uma matriz 3D. Na saída, cada mapa de recursos representa um recurso específico extraído em todos os locais da entrada (LECUN; KAVUKCUOGLU; FARABET, 2010). Em CNNs convencionais, os estágios são compostos por camadas convolucionais, camadas de agrupamento (*pooling*) e camadas densas seguidas de módulos de classificação (ZHANG *et al.*, 2021).

Camada de convolução: Na camada convolucional, um tensor de entrada e um tensor filtro (*kernel*) geram um tensor de saída por meio de uma operação de correlação cruzada (Figura 4). Os filtros operam como uma janela fixa que desliza sobre todas as regiões da entrada, de acordo com um determinado passo (*stride*). A saída da camada convolucional é chamada de mapa de características (Figura 4), pois indica a representação da imagem dada pelo modelo sobre os padrões encontrados em suas dimensões espaciais. Todas as unidades em um mapa de características compartilham o mesmo banco de filtros, gerando assim vários mapas para cada imagem de entrada.

Esses filtros podem ser projetados para diferentes técnicas de processamento de imagens, como: detecção de bordas (Figura 5), desfoque e nitidez. Os diferentes mapas de características formados nesta etapa são empilhados, formando a camada de convolução.

Figura 4 – Operação de correlação cruzada. As partes sombreadas são o primeiro elemento de saída, bem como os elementos tensores de entrada e kernel usados para o cálculo de saída:

Input		Kernel		Output																	
<table border="1" style="border-collapse: collapse;"> <tr><td style="background-color: #e0e0e0;">0</td><td style="background-color: #e0e0e0;">1</td><td style="background-color: #e0e0e0;">2</td></tr> <tr><td style="background-color: #e0e0e0;">3</td><td style="background-color: #e0e0e0;">4</td><td style="background-color: #e0e0e0;">5</td></tr> <tr><td style="background-color: #e0e0e0;">6</td><td style="background-color: #e0e0e0;">7</td><td style="background-color: #e0e0e0;">8</td></tr> </table>	0	1	2	3	4	5	6	7	8	*	<table border="1" style="border-collapse: collapse;"> <tr><td style="background-color: #e0e0e0;">0</td><td style="background-color: #e0e0e0;">1</td></tr> <tr><td style="background-color: #e0e0e0;">2</td><td style="background-color: #e0e0e0;">3</td></tr> </table>	0	1	2	3	=	<table border="1" style="border-collapse: collapse;"> <tr><td style="background-color: #e0e0e0;">19</td><td style="background-color: #e0e0e0;">25</td></tr> <tr><td style="background-color: #e0e0e0;">37</td><td style="background-color: #e0e0e0;">43</td></tr> </table>	19	25	37	43
0	1	2																			
3	4	5																			
6	7	8																			
0	1																				
2	3																				
19	25																				
37	43																				

Figura 5 – Efeito do filtro detector de bordas e ativação linear



Pooling: outra técnica que utiliza do recurso de janela deslizante é o agrupamento (em inglês: *pooling*). A aplicação do agrupamento não leva em conta o uso de filtros convolucionais nem operações de correlação cruzada. Segundo (ZHANG *et al.*, 2021), os operadores de agrupamento são determinísticos, normalmente calculando o valor máximo (Figura 6) ou médio dos elementos na janela de agrupamento. Segundo (KHAN *et al.*, 2020), essa técnica permite redimensionar os mapas de características gerados pelas camadas convolucionais, reduzindo as dimensões, resumindo as informações semelhantes nas vizinhanças dos pixels e produzindo a resposta dominante

nessa região local (Figura 7). A redução no tamanho do mapa de recursos para um conjunto de recursos invariáveis não apenas regula a complexidade da rede, mas também ajuda a aumentar a generalização, reduzindo o overfitting.

Figura 6 – Operação de max pooling

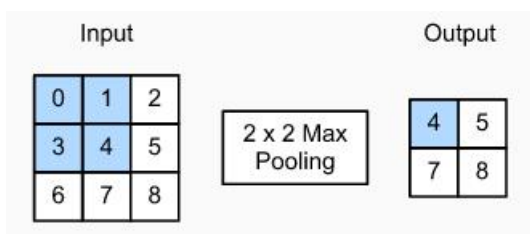
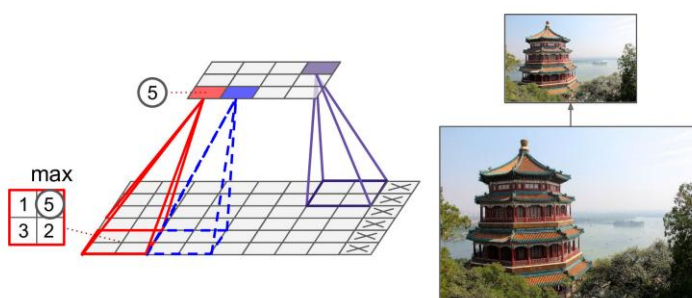


Figura 7 – Aplicando max pooling em uma imagem



Camadas Totalmente Conectadas: após a fase de extração de características é comum aplicar camadas densas (*fully connected*) na saída da rede convolucional (Figura 8) para executar tarefas de aprendizado de máquina, como classificação (figura 9). A Figura 9 representa uma típica arquitetura CNN contendo uma sequência de blocos convolucionais e pooling. Em uma CNN, as camadas convolucionais são normalmente organizadas de forma que diminuam gradualmente a resolução espacial das representações, enquanto aumentam o número de canais (ZHANG *et al.*, 2021).

Figura 8 – Típica Arquitetura CNN contendo uma sequência de blocos convolucionais, compostas por operações de convolução e pooling; e camadas totalmente conectadas

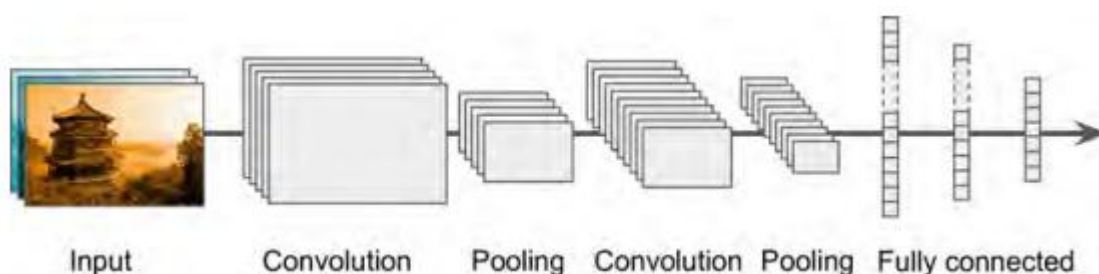
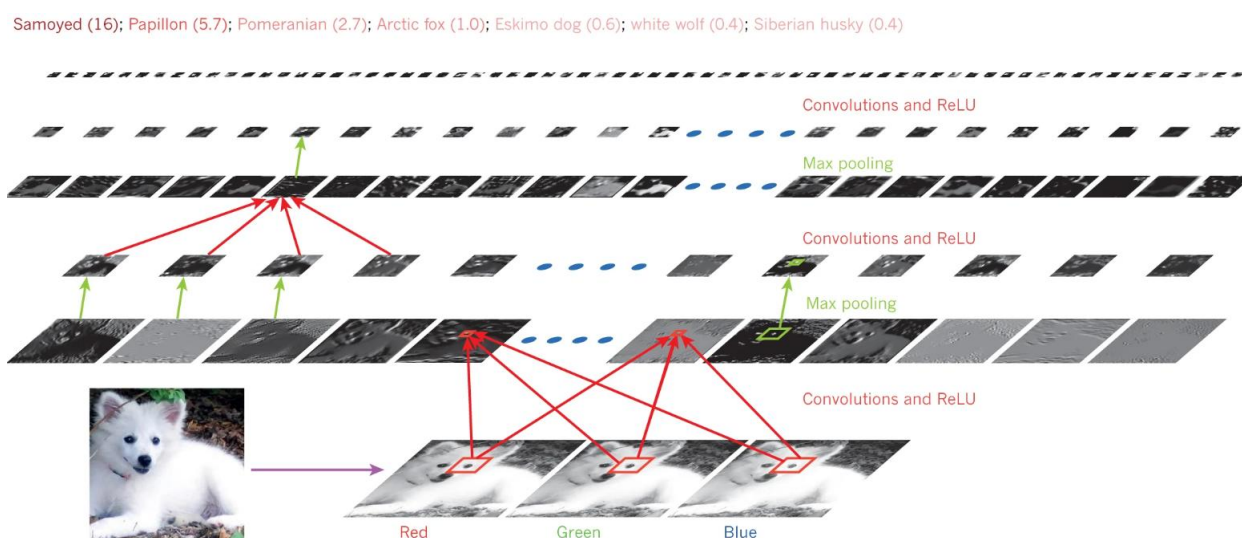


Figura 9 – As saídas de cada camada de uma arquitetura de rede convolucional aplicada à imagem de um cachorro. A imagem demonstra a extração de recursos realizada pelos blocos convolucionais e a probabilidade do cachorro em destaque pertencer a cada classe.



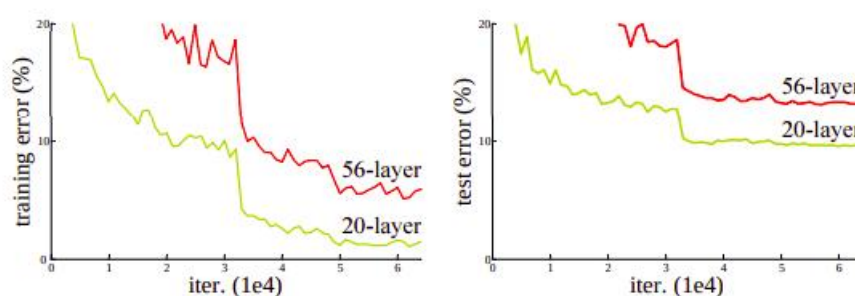
A rede resultante garante informações espaciais dos dados e invariância a certas transformações, o que significa que pode classificar padrões não importa onde eles ocorram dentro da imagem de entrada, (MURPHY, 2012) e (BISHOP, 2006).

As convencionais arquiteturas de Redes Neurais Convolucionais apresentam um formato de conceito simples e inteligente que possibilita aplicá-las em diversas tarefas. Segundo LECUN; KAVUKCUOGLU; FARABET (2010), embora esses modelos sejam muito eficazes, seu treinamento requer um grande número de amostras rotuladas.

2.1.4 Redes Neurais Residuais

Entendendo melhor o conceito das redes neurais profundas, podemos imaginar que, quanto mais profunda for a rede, mais propensa é a um melhor desempenho na tarefa almejada. Conforme descrito em HE *et al.* (2016) e SRIVASTAVA; GREFF; SCHMIDHUBER (2015), há obstáculos no modelo causados pela profundidade da rede. Com o aumento da profundidade, o modelo pode chegar a saturação e até mesmo se degradar rapidamente. Tal degradação não é causada por *overfitting*, e adicionar mais camadas a um modelo profundo pode levar a um maior erro de treinamento (Figura 10).

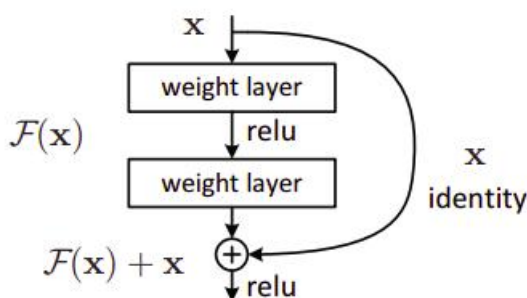
Figura 10 – Erro de treinamento (esquerda) e erro de teste (direita) no dataset CIFAR-10 com redes planas de 20 e 56 camadas



Uma solução para construir modelos profundos é adicionar blocos residuais em sua arquitetura (Figura 11). Em vez de esperar que cada camada empilhada se encaixe diretamente em outras camadas subjacentes, como ocorre em arquiteturas planas, os blocos residuais conduzem essas camadas para que se encaixem em um mapeamento residual $F(x) = H(x) - x$ (Figura 11), modelando a arquitetura para que a entrada da rede seja adicionada em sua saída. Segundo (HE *et al.*, 2016), a formulação de $F(x) + x$ pode ser realizada por redes neurais feedforward com conexões que pulam uma ou mais camadas, chamadas de atalho (do inglês, *skip connections*). Essas conexões realizam o mapeamento de identidade e suas saídas são adicionadas às saídas das camadas empilhadas, que são compostas por blocos convolucionais e função de ativação ReLU. As *skip connections* não adicionam parâmetros extras nem complexidade computacional.

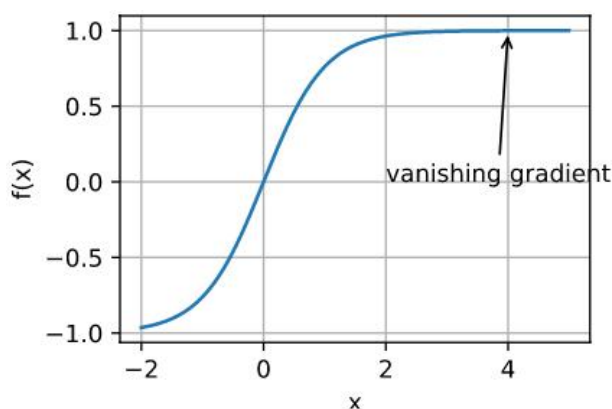
A existência desta solução indica que um modelo mais profundo não deve produzir maior erro de treinamento do que sua contraparte menos profunda. Além disso,

Figura 11 – Bloco residual contendo o somatório de $f(x)$ com camadas de pesos unitários.



é possível eliminar o problema de *vanish gradient*, que ocorre na etapa de *backpropagation*, onde os valores de gradiente podem ser tão pequenos que impedem que os pesos mudem seus valores. Por exemplo, à partir da Figura 12, ao minimizar a função $f(x) = \tanh(x)$ para $x = (4)$ podemos ver que o gradiente de f é quase nulo. Mais especificadamente, $f'(x) = 1 - \tanh^2(x)$, portanto $f'(4) = 0,0013$; conseqüentemente, a otimização ficará estática por um longo período (ZHANG *et al.*, 2021).

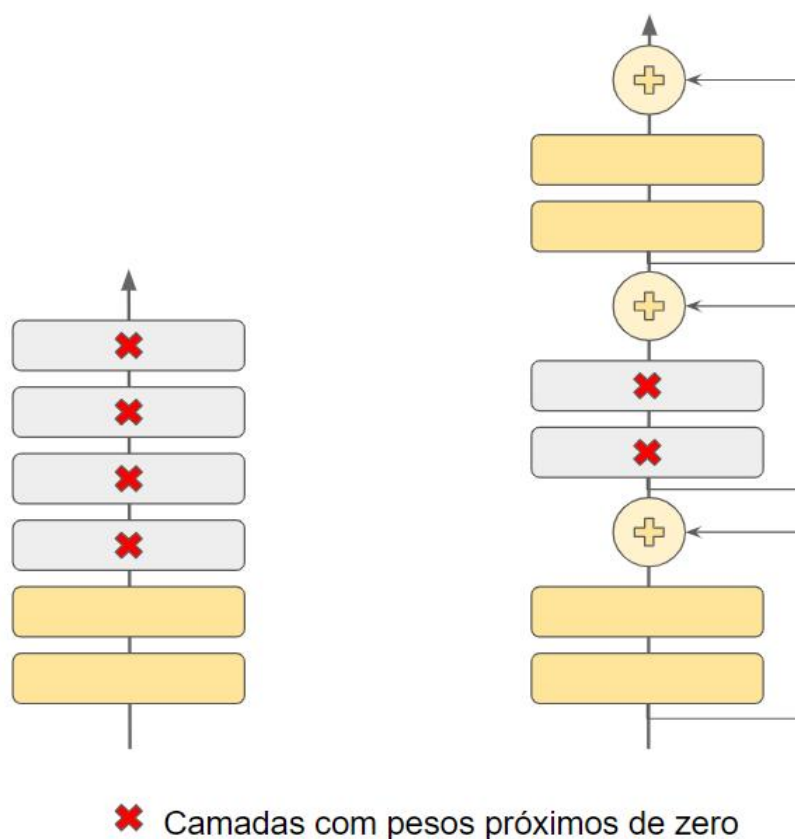
Figura 12 – Representação do Problema de Vanish Gradient, destacando a região em que o gradiente de f é quase nulo



Com a presença das *skip connections* os valores de gradiente da função de custo podem percorrer a arquitetura até as camadas iniciais do modelo. Segundo GÉRON (2017), quando uma rede neural é inicializada, seus pesos são próximos de zero, então a rede apenas gera valores próximos a zero. Ao adicionar uma *skip connection*, a rede resultante apenas exibirá uma cópia de suas entradas; em outras palavras, inicialmente modela a função identidade. Se a função de destino estiver bastante próxima da função de identidade (o que geralmente é o caso), isso acelerará consideravelmente o treinamento. Além disso, ao adicionar muitas *skip connection*, a rede pode começar a progredir mesmo que várias camadas ainda não tenham

começado a aprender. Graças a pular conexões, o sinal pode facilmente passar por toda a rede. A rede residual profunda pode ser vista como uma pilha de unidades residuais, onde cada unidade residual é uma pequena rede neural com uma conexão de salto. A Figura 13 demonstra o funcionamento desse processo.

Figura 13 – Representação de uma rede neural comum (esquerda) e uma rede neural com blocos residuais (direita). Representa dados que passam por toda a rede, eliminando o problema de *vanish gradient*, devido as *skip connections*.



Modelos de arquitetura como esse, chamado ResNet, são disponibilizados com diferentes profundidades, atualmente entre 18 a 152 camadas (HE *et al.*, 2016).

2.1.5 InceptionResnetV2

A Inception Resnet V2 é originada através da combinação de duas arquiteturas: ResNet (HE *et al.*, 2016) e Inception (SZEGEDY *et al.*, 2016). As arquiteturas inception foram criadas para atribuir esparsidade nos dados e reduzir a complexidade computacional causada por várias camadas sucessivas de filtros convolucionais. A combinação de Kernels de diferentes dimensões permite uma ampla distribuição dos dados de entrada. Se duas camadas convolucionais estiverem encadeadas, qualquer aumento uniforme no número de seus filtros resultará em um aumento quadrático da computa-

ção. Conforme SZEGEDY *et al.* (2015), passar de arquiteturas totalmente conectadas para arquiteturas esparsamente conectadas, mesmo dentro das convoluções, é uma maneira de evitar esse problema.

O modelo constitui-se de camadas sucessivas, resultando em uma cadeia de agrupamentos conectados em série, no final concatenados. As camadas tem filtros com diferentes tamanhos, tornando-se a rede mais ampla ao invés de mais profunda. Segundo SZEGEDY *et al.* (2017), isso significa que muitos clusters são concentrados em uma única região e eles podem ser cobertos por uma camada de convoluções 1×1 na próxima camada, aumentando a dimensão dos filtros. Como as redes convolucionais são custosas computacionalmente, os canais de entrada são limitados adicionando as camadas de convolução 1×1 .

Como as redes Inception tendem a ser muito profundas, é natural substituir o estágio de concatenação de filtros da arquitetura Inception por conexões residuais. Isso permite com que a Inception obtenha todos os benefícios da abordagem residual, mantendo sua eficiência computacional.

A Figura 14 representa o esquema da arquitetura da rede InceptionResnetV2. A arquitetura é extensa e possui vários blocos de camadas combinadas entre a inception e a resnet. Como exemplo, o bloco Inception-resnet-A é ilustrado na Figura 15.

Figura 14 – Representação da Arquitetura InceptionResnetV2

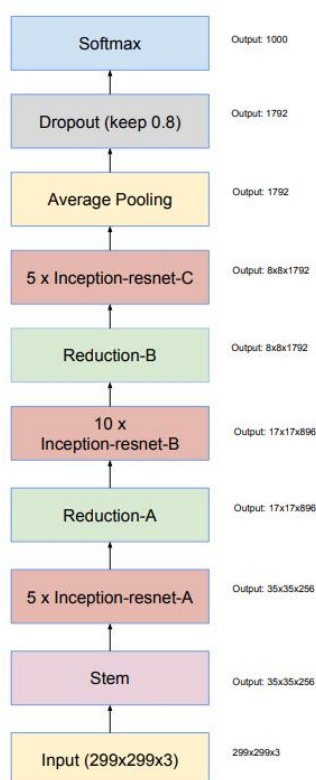
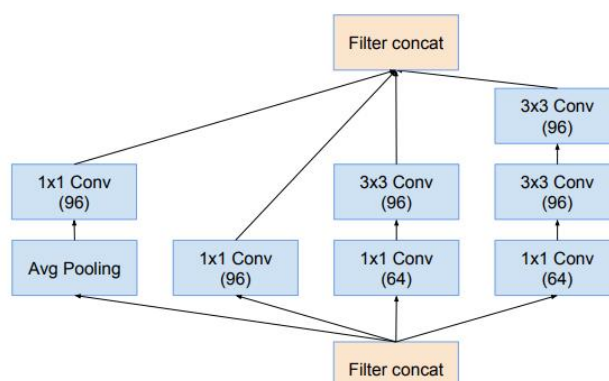


Figura 15 – Esquema para módulos de grade 35×35 da rede Inception-v4 pura.

A premissa ao desenvolver essa arquitetura é introduzir conexões residuais que conectam a entrada do modelo inception com sua saída, tornando modelos inception menos custosos computacionalmente.

2.1.6 EfficientNet

Aumentar a profundidade de arquiteturas de Redes Neurais Convolucionais para alcançar melhores desempenhos são técnicas convencionalmente utilizadas na área de aprendizado profundo. Podemos mencionar as arquiteturas ResNet, que são escaladas aumentando a profundidade por meio da adição de mais camadas no modelo. Ao realizar essas transformações, os dados de entrada podem ser alterados quanto a largura, profundidade e resolução.

Em TAN; LE (2019), é mencionado que embora seja possível dimensionar duas ou três dimensões arbitrariamente, o dimensionamento arbitrário requer um ajuste manual tedioso e ainda produz baixo desempenho. Para solucionar essa questão, novos métodos para escalar modelos CNN foram criados e deu origem à arquitetura EfficientNet. Essa arquitetura usa uma técnica chamada coeficiente composto para escalar modelos de maneira simples, porém de forma eficaz. Em vez de aumentar aleatoriamente a largura, a profundidade ou a resolução, o dimensionamento composto redimensiona uniformemente cada dimensão com um determinado conjunto fixo de coeficientes. Esse método é ilustrado na Figura 16.

A Figura 17 mostra que, em comparação com outras arquiteturas, o desempenho da EfficientNet no conjunto de dados da ImageNet apresenta alto desempenho com o menor número de parâmetros (TAN; LE, 2019).

Figura 16 – (a) modelo sem nenhum método de dimensionamento composto aplicado; (b)-(d) dimensionamentos convencionais contendo técnicas que redimensionam a largura, profundidade ou resolução da rede. (e) é o método de dimensionamento composto que redimensiona uniformemente todas as três dimensões (largura, profundidade e resolução).

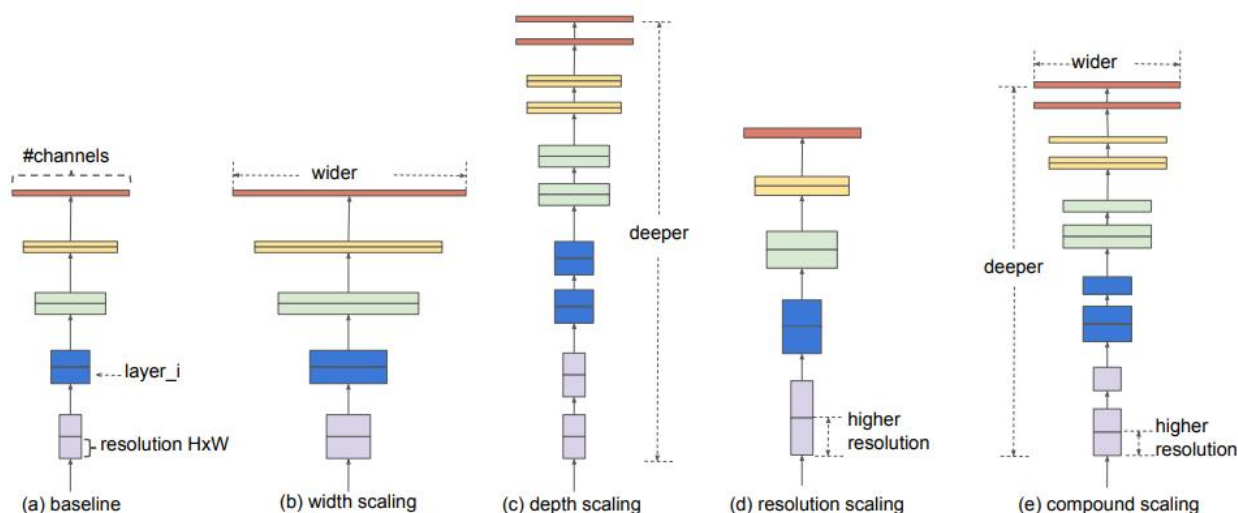
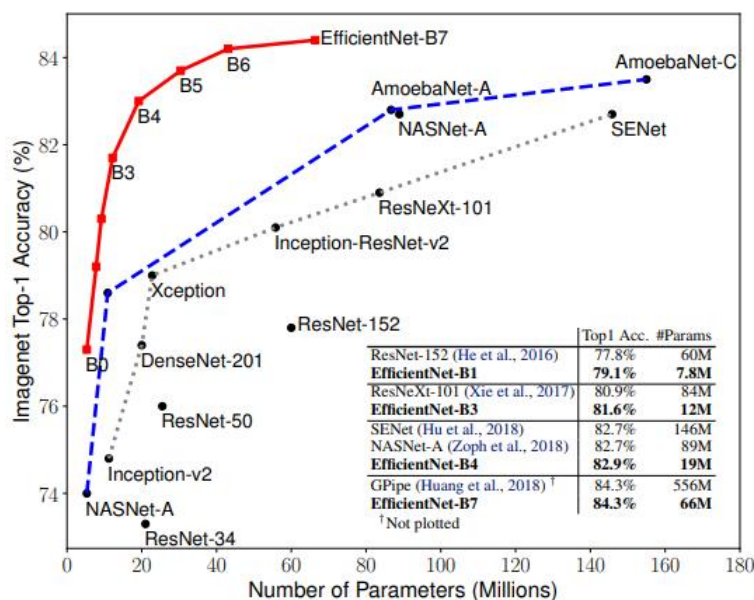


Figura 17 – Performance da Arquitetura EfficientNet com demais Redes na ImageNet



2.1.7 U-Net

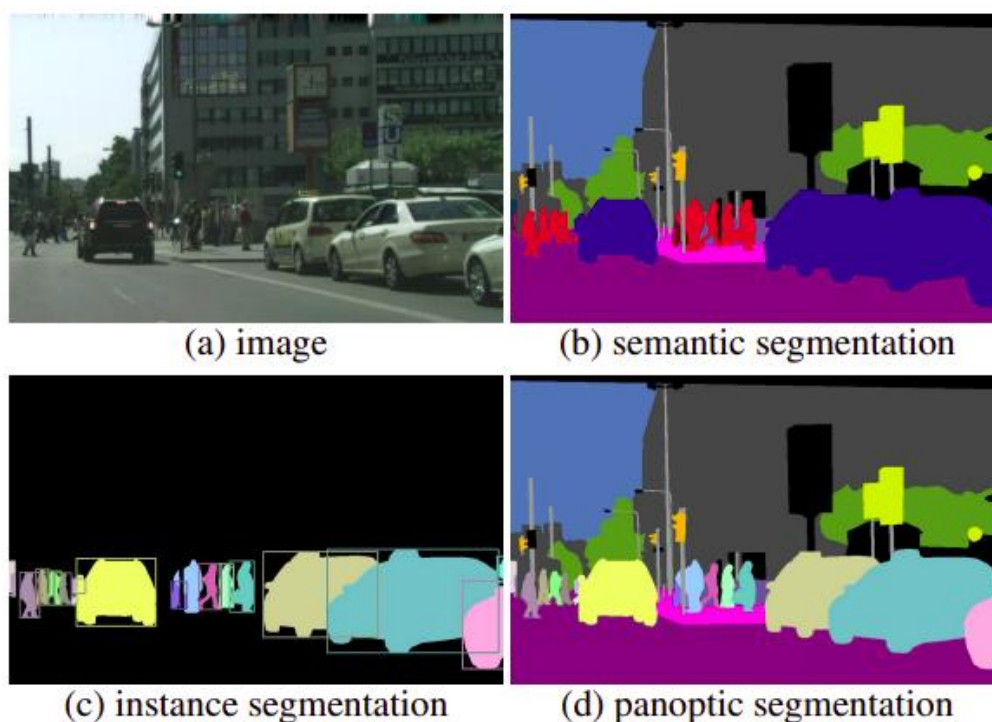
A visão computacional trata de métodos para que máquinas possam ter compreensão de alto nível a partir de imagens ou vídeos digitais. A união com o aprendizado de máquina permite automatizar certas tarefas que estão presentes no cotidiano, como: detecção de incêndios, direção autônoma, detecção intrusos em sistemas de segu-

rança, dentre outros. A utilização de *Deep Learning* em tarefas de visão computacional propõe algoritmos do estado da arte em que o processamento massivo de dados são transformados em informações em tempo real para a máquina, como podemos ver no sistema de navegação autônoma dos atuais veículos da Tesla.

Segundo MINAEE *et al.* (2021), a segmentação é um componente essencial em muitos sistemas de visão computacional. Envolve particionar imagens ou quadros de vídeo em múltiplos segmentos ou objetos, desempenhando um papel central em uma ampla gama de aplicações, incluindo análise de imagens médicas.

A técnica de segmentação pode ser formulada em três aplicações (Figura 18): segmentação semântica, segmentação por instância e segmentação panóptica. A segmentação semântica é vista como um problema de classificação de pixels com rótulos semânticos, onde é previsto um rótulo para cada pixel; já a segmentação de instância é o particionamento de objetos individuais numa imagem, realizando uma rotulação em nível de pixel com um conjunto de categorias de objetos (carros, placas, objetos do ambiente) para todos os pixels da imagem. A segmentação panóptica unifica as tarefas tipicamente distintas de segmentação semântica (atribuir um rótulo de classe a cada pixel) e segmentação de instância (detectar e segmentar cada instância de objeto) (KIRILLOV *et al.*, 2019).

Figura 18 – Exemplos de Técnicas de Segmentação



A arquitetura U-Net, proposta por RONNEBERGER; FISCHER; BROX (2015), consiste em substituir as convencionais Redes Neurais Convolucionais em tarefas que

exigem a localização dos pixels nas imagens. Conforme os autores, em muitas tarefas visuais, especialmente no processamento de imagens biomédicas, a saída desejada deve atribuir um rótulo de classe a cada pixel. Essa rede tem o formato de "U", e não possui camadas totalmente conectadas, apenas é dividida em duas partes: encoder e decoder.

O encoder é responsável pela extração de recursos da imagem, gerando os mapas de características. É composto por blocos convolucionais seguidos por maxpooling, fazendo o *downsampling* na estrutura dos dados. Nessa etapa, o resultado final gera imagens com resolução bem menor que a imagem original.

Após a extração de características, o modelo aprendeu certos padrões nos dados, porém com baixa resolução. O decoder propõe substituir os convencionais operadores de pooling por operadores de *upsampling* seguidos por blocos convolucionais. As dimensões dos mapas de características são expandidas até a resolução original, gerando mapas de características com alta resolução.

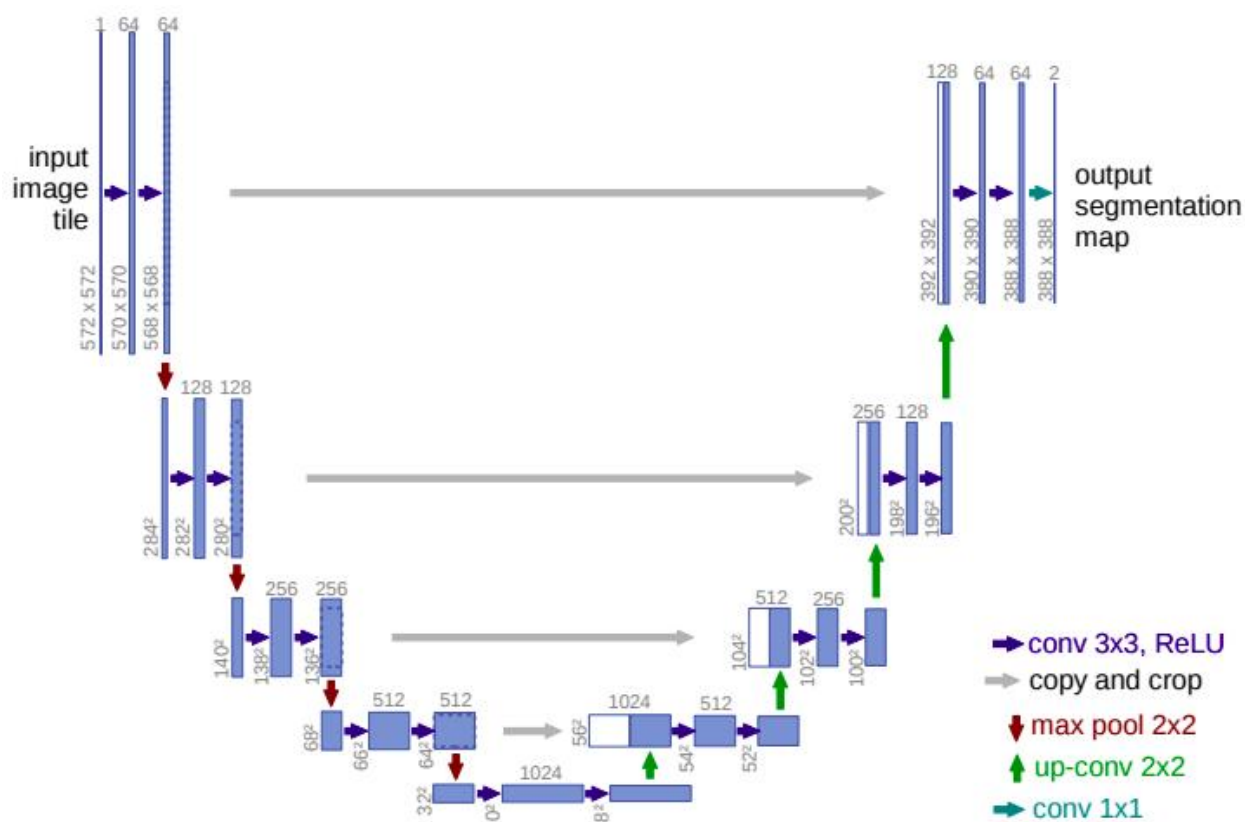
Dessa forma, durante a fase de encoder o modelo consegue aprender características importantes dos dados. Quando chega a fase de decoder, a resolução espacial é aumentada, permitindo a localização das características encontradas nas dimensões originais da imagem.

A arquitetura da U-Net pode ser visualizada por meio da Figura 19. As camadas de descida são representadas pelo encoder e a fase de decoder são as camadas em ascensão.

Especificadamente, para adaptação em qualquer conjunto de dados que contenha imagens médicas, foi desenvolvido um *framework* chamado nnU-Net, proposto por ISENSEE *et al.* (2018). Esse modelo baseado na arquitetura U-Net propõe uma estrutura dinâmica para adaptação dos dados a novas tarefas de segmentação, onde o pré-processamento das imagens é feito de forma automática, sem a necessidade de interação manual. Seu nome vem de "*No New-Net*", pelo fato de não ser uma nova arquitetura de rede, e sim a U-Net com algumas modificações que permitem melhor desempenho da rede.

Para cada tarefa a ser iniciada, o nnU-Net executa automaticamente uma validação cruzada de cinco vezes para três modelos com a arquitetura U-Net configurados automaticamente com diferentes parâmetros. O modelo com a pontuação média mais alta é escolhido.

Figura 19 – Arquitetura U-Net



2.2 ENVELHECIMENTO CEREBRAL

Existem diversas vertentes que levam ao processo de envelhecimento cerebral, desde fenômenos patológicos a ocorrências do cotidiano que levam o indivíduo ao declínio cognitivo. Segundo PETERS (2006), o envelhecimento causa mudanças no tamanho do cérebro, na vasculatura e na cognição. O cérebro encolhe com o aumento da idade e há mudanças em todos os níveis, desde as moléculas até a morfologia. A incidência de acidente vascular cerebral, lesões da substância branca e demência também aumentam com a idade, assim como o nível de comprometimento da memória e há alterações nos níveis de neurotransmissores e hormônios.

A redução de áreas corticais causa deterioração no processo de cognição, assim como a progressão da atrofia na matéria cerebral. Um exemplo citado em KOZIC *et al.* (2022) é a carga de hiperintensidade da substância branca (parcialmente associada à hipertensão arterial) que contribui para a redução do volume da substância cinzenta.

Em KOZIC *et al.* (2022), é explicado que a questão do envelhecimento cerebral saudável e resultados estão ligadas desde a especificidade de gênero ao impacto do estilo de vida de um indivíduo. O impacto da genética, hormônios e experiência têm um

papel a desempenhar no envelhecimento do cérebro. Contudo, conforme (PETERS, 2006), nem tudo é negativo. Níveis mais elevados de educação ou realização profissional podem atuar como fator de proteção, assim como manter uma dieta saudável, ingestão de álcool baixa a moderada e exercícios regulares. O envelhecimento biológico não é completamente determinado pelo envelhecimento cronológico, porém pode ser possível retardar o envelhecimento biológico e até reduzir a possibilidade de sofrer de problemas cognitivos relacionados com a idade.

2.3 IMAGENS DE RESSONÂNCIA MAGNÉTICA

A ressonância magnética utiliza campo magnético de alta intensidade e sinais de radiofrequência para produzir imagens de estruturas anatômicas, permitindo detectar a presença de doenças e analisar várias funções biológicas que ocorrem dentro do corpo humano, (SPRAWLS, 2000). Sua utilização é adequada para gerar imagens de partes não ósseas ou tecidos moles do corpo. O cérebro, a medula espinhal e os nervos, bem como os músculos, ligamentos e tendões são vistos muito mais claramente com a ressonância magnética do que com raios-x e tomografia computadorizada; por esta razão, a ressonância magnética é frequentemente usada para imagens de lesões no joelho e no ombro.

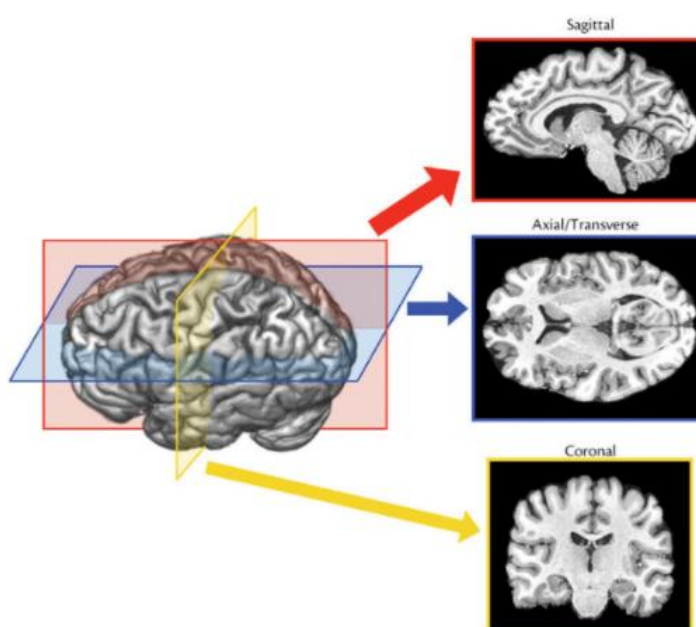
Uma imagem de MRI (*magnetic resonance imaging*) é composta por voxels - representação de pixels no espaço tridimensional. Em neuroimagem é comum utilizar alguns termos padrão para identificar as regiões do cérebro (Figura(20)). O volume pode ser dividido em 3 tipos de planos: coronal (vista superior-inferior e esquerda-direita), sagital (vista superior-inferior e anterior-posterior) e axial (vista anterior-posterior e esquerda-direita).

Há diferentes tipos de modalidades de MRI: estrutural, funcional, difusão, dentre outras. Cada modalidade é composta por diferentes tipos de aquisição (Figura 21) que, individualmente, permite visualizar propriedades físicas de formas específicas; destacam tecidos, estruturas e processos que ocorrem dentro do cérebro, possibilitando analisar a anatomia geral, atividade funcional, microestrutura e composição química (JENKINSON; CHAPPELL, 2018). Patologias que não podem ser identificadas em determinada modalidade podem ser visíveis em outras, tornando-se uma grande vantagem da utilização de MRI (SPRAWLS, 2000).

Imagens estruturais, como as representadas na Figura 21, contém informações relacionadas às estruturas anatômicas do cérebro, possibilitando visualizar os tecidos de matéria cinza, matéria branca e fluido cerebrospinal. É utilizada em ambientes clínicos para inspeção visual de fenômenos que causam mudanças na anatomia do cérebro, como lesões patológicas, deformações anatômicas e neurodegeneração, (JENKINSON; CHAPPELL, 2018).

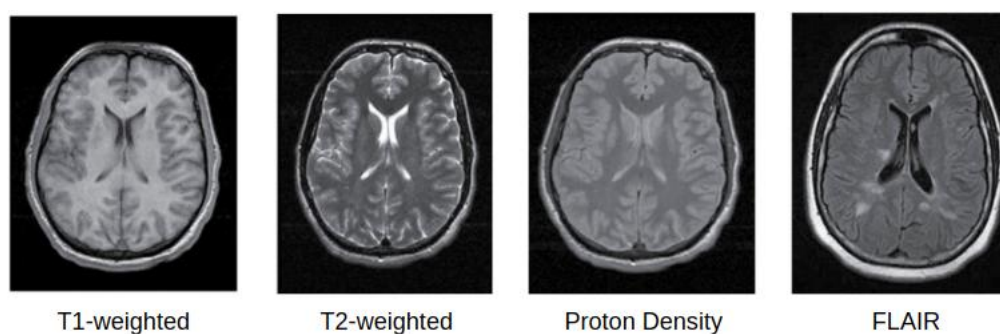
Essa modalidade é considerada um componente essencial para estudos e pes-

Figura 20 – Ilustração dos 3 planos padrão de uma imagem de Ressonância Magnética Estrutural. Plano sagittal (vista superior-inferior e anterior-posterior), axial (vista anterior-posterior e esquerda-direita), coronal (vista superior-inferior e esquerda-direita).



Fonte: (JENKINSON; CHAPPELL, 2018)

Figura 21 – Imagens de ressonância magnética estrutural em diferentes modalidades, demonstrando a diferença de contraste entre as regiões da estrutura do cérebro.



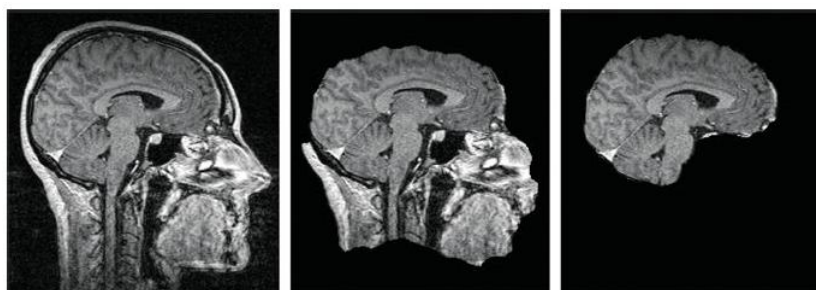
quisas, pois permite obter um alinhamento preciso dos cérebros de diferentes sujeitos, incluindo estudos que utilizam outros tipos de aquisição. Sem uma imagem estrutural, é difícil fazer um estudo em grupo, pois o mal alinhamento das imagens normalmente arruinará a capacidade de localizar mudanças de interesse nas imagens (JENKINSON; CHAPPELL, 2018).

2.4 PRÉ-PROCESSAMENTO

No contexto do aprendizado de máquina, é importante aplicar técnicas de pré-processamento antes de entrar no modelo; isso facilita a comparação entre diferentes fontes de dados, além de reduzir a complexidade computacional. Algumas dentre as técnicas de pré-processamento são: extração do cérebro e co-registro (do inglês: *skull stripping* e *registration*). A técnica de extração do cérebro consiste na remoção do crânio e de estruturas não cerebrais da imagem. O algoritmo retorna um volume ou uma máscara binária, ambos contendo apenas voxels relacionados ao cérebro. Isso evita enviar para um modelo de aprendizado de máquina partes da imagem que não são de interesse; o que pode interferir nos resultados, pois de acordo com (DESPOTOVIĆ; GOOSSENS; PHILIPS, 2015) tecidos não cerebrais, como gordura, crânio ou pescoço, têm intensidades que se sobrepõem às intensidades dos tecidos cerebrais.

Segundo (JENKINSON; CHAPPELL, 2018), a extração do cérebro é um processo útil desde que seja robusto e não falhe: este é um princípio geral em todos os pipelines de análise. Cada estágio do *pipeline* traz consigo uma chance de falha e métodos não robustos devem ser evitados, ou pelo menos monitorado com muito cuidado. A inspeção visual é extremamente importante para assegurar a qualidade da extração cerebral, já que existe a probabilidade de falhas algorítmicas (alguns dos tecidos não cerebrais podem não ser removidos), erros do usuário (imagens erradas utilizadas como entrada) e falha do computador (saída em branco ou corrompida). Por meio da Figura 22 podemos analisar um modelo de extração do cérebro falho e um modelo bem feito.

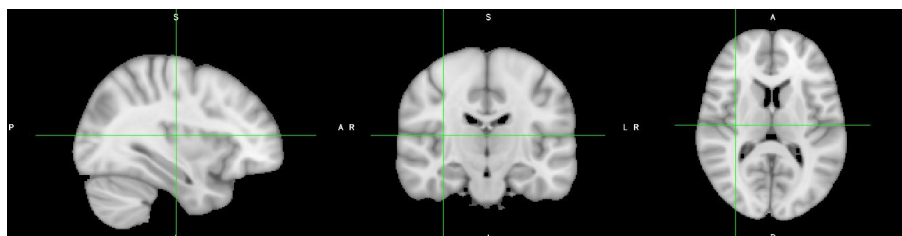
Figura 22 – Ilustração do processo de extração do cérebro (*skull stripping*). Imagem original (esquerda); extração do cérebro mal feita (meio); boa extração do cérebro (direita).



Em casos de análises de neuroimagens contendo diferentes scans e grupos de indivíduos, é comum padronizá-las para um template de referência, fazendo com que cada volume tenha as mesmas dimensões de voxels e sejam orientadas para as mesmas coordenadas. A técnica de co-registro é responsável por essas transformações. O co-registro é o processo de alinhamento espacial de duas ou mais imagens para

um atlas padrão (Figura 23), tendo por finalidade realizar transformações necessárias para que ambas fiquem alinhadas e na mesma orientação.

Figura 23 – Representação do atlas MNI152 - apenas cérebro



3 TRABALHOS RELACIONADOS

No trabalho de COLE *et al.* (2017), foram investigados diferentes modelos e técnicas de pré-processamento visando garantir biomarcadores confiáveis do envelhecimento natural do cérebro. Para comparação, foram utilizados dois modelos: um modelo 3D de Redes Neurais Convolucionais e um modelo de GPR (Gaussian Processes Regression). Para demonstrar que é possível obter bons resultados referentes à predição da idade cerebral utilizando apenas dados brutos, os autores recorreram a um conjunto de neuroimagens estruturais com 2001 indivíduos saudáveis e testaram o desempenho de seus modelos em quatro formas diferentes de dados de entrada, sendo treinados de forma individual: somente matéria branca (WM), apenas matéria cinza (GM), matéria branca e matéria cinza combinadas e por fim, dados T1 brutos. Cada modelo foi avaliado com cada um dos quatro tipos de entrada. Os resultados demonstram a capacidade de obter bom desempenho na tarefa de predição da idade cerebral utilizando imagens de entrada com o mínimo de pré-processamento (skull stripping e co-registro). O modelo CNN atingiu melhores resultados, com MAE de 4,65 anos, enquanto o modelo GPR atingiu o valor de 11,81 anos para os dados brutos. Para as demais entradas, os valores de ambos os modelos mantiveram abaixo de 5,88 anos.

Em BASHYAM *et al.* (2020), foi desenvolvido um modelo de *Deep Learning* que utiliza uma arquitetura de Redes Neurais Convolucionais 2D: inception-resnet-V2 (SZEGEDY; IOFFE; VANHOUCHE, 2016). A estratégia de utilizar essa arquitetura é motivada pela possibilidade de executar a inicialização da rede utilizando um modelo pré-treinado na ImageNet - conjunto de dados com milhões de imagens rotuladas e 1.000 classes diferentes, (DENG *et al.*, 2009). Os autores utilizaram um grande conjunto de dados, com cerca de 11729 imagens de MRI sequenciadas em T1-MPRAGE. O *dataset* compreende uma alta diversidade de estudos, *scanners*, idades e localizações geográficas. A faixa etária escolhida no estudo é de 3 a 95 anos para o conjunto de treinamento. Para investigar deltas de idade cerebral do modelo, utilizaram três diferentes coortes específicas de doenças que, posteriormente, serviram para criar modelos de classificação de doenças por meio de *transfer learning*. O modelo obteve um erro absoluto médio de 3,702 anos.

O satisfatório resultado apresentado em BASHYAM *et al.* (2020) demonstrou a capacidade de utilizar um modelo robusto e pré-treinado para a tarefa de predição da idade cerebral. O modelo que antes era utilizado como classificador, teve a camada do classificador retirada e foi modificado para a tarefa de regressão. Outra possibilidade relevante foi a utilização de dados de entrada 2D, o que garante menos esforços computacionais que volumes de MRI como entrada do modelo.

Em POLONI; FERRARI; INITIATIVE *et al.* (2022), também foram investigadas

técnicas que visam reduzir a diferença entre a idade cronológica e a idade predita pelo modelo. O conjunto de dados utilizado contém aproximadamente 1.550 imagens de ressonância magnética. Os experimentos foram feitos utilizando apenas regiões do hipocampo, considerando que o hipocampo é uma das regiões diretamente relacionadas ao envelhecimento natural do cérebro. Duas redes diferentes foram treinadas: uma para o hipocampo esquerdo e outra para o direito; posteriormente, as saídas foram combinadas, calculando o resultado médio para criar a previsão final e formar um modelo *ensemble*. Utilizando esse método, dois estudos foram feitos. O primeiro estudo investiga a tarefa de predição da idade cerebral em sujeitos com idades entre 20 a 70 anos. O segundo também visa a predição da idade, embora foram utilizados sujeitos com idade superior a 70 anos, com o objetivo de realizar uma comparação pareada por idade entre as predições de idade feitas em pacientes controle (CN), com Alzheimer (AD) e com comprometimento cognitivo leve (MCI). Utilizaram a arquitetura de redes neurais convolucionais 3D, EfficientNet, contendo modificações que transformaram toda a rede 2D em uma arquitetura feita para imagens 3D. O modelo apresenta uma saída linear contendo a idade estimada. Para o primeiro experimento, os resultados das redes treinadas de forma individual são muito semelhantes, enquanto o modelo *ensemble* melhorou os resultados, alcançando o valor de MAE de 3,31 anos. Durante o segundo teste, pacientes CN obtiveram o MAE 3,52 anos, enquanto os resultados das predições feitas em pacientes com AD e MCI aumentaram, obtendo valores de 5,8 para AD e 5,19 para MCI; o que é esperado quando se trata de um modelo treinado em pacientes controle.

Em WOOD *et al.* (2022) foi utilizado um conjunto de dados com 23.302 imagens de ressonância magnética cerebrais de pacientes diagnosticados como saudáveis (faixa de idade = 18–95 years). As imagens são ponderadas em T2 e difusão axial, com diferentes tipos de hospitais e fornecedores de *scanners*. Foi utilizada a arquitetura de Redes Neurais Convolucionais 3D, DenseNet121 (HUANG *et al.*, 2017), para predição da idade cerebral. Os autores demonstram que modelos ponderados em T2 generalizaram bem entre imagens de diversos hospitais e diferentes fornecedores de *scanners*. Avaliaram também a diferença entre modelos treinados utilizando ponderações em T1 e T2 (T1: MAE = 3,86 anos, T2: MAE = 3,83 anos); e a diferença entre ponderações T2 e Difusão (T2: MAE = 3,32 anos, Difusão: MAE = 3,98 anos); os resultados demonstraram pequena melhora estatisticamente significativa em relação às previsões ponderadas em T2 sozinhas. Outro método avaliado foi a diferença da predição da idade do cérebro utilizando imagens com e sem remoção do crânio (*skull stripping*); o tempo de treinamento de imagens que utilizam dados sem *skull stripping* foi significativamente mais rápido (pré-processamento + tempo de previsão = $4,6 \pm 0,8$ s) em comparação com imagens que utilizaram *skull stripping* (pré-processamento + tempo de previsão = $48,9 \pm 3,2$ s). A previsão utilizando apenas imagens ponde-

radas em T2 alcançou um MAE de 2,97 anos. Foi obtida usando varreduras axiais ponderadas em T2 (N treinamento = 15.146, N teste = 4.661).

Em HAN *et al.* (2022) avaliaram 27 algoritmos de aprendizado de máquina para prever a idade cerebral de sujeitos saudáveis. Três conjuntos de dados diferentes foram utilizados: HCP (N=1.113, média de idade = 28 anos), Cam-CAN (N=601, média de idade = 54 anos) e IXI (N=1.113, média de idade = 48 anos). Foi utilizado a ferramenta *FreeSurfer* (DESIKAN *et al.*, 2006), que permite pré-processar e extrair medidas morfométricas de ressonância magnética estrutural. A ferramenta gera uma matriz de características que consiste em medidas morfológicas do cérebro (espessura cortical, área de superfície, volume subcortical e volume intracraniano total); essa matriz foi utilizada como dados de entrada dos modelos. A biblioteca PyCaret foi escolhida para avaliar os diversos modelos de aprendizado de máquina, pois requer menos linhas de código para executar vários modelos de aprendizado de máquina, com o objetivo de simplificar o processo de treinamento. Cada algoritmo foi treinado, separadamente, nos três *datasets*. As amostras foram divididas em um conjunto de treinamento (80%) e um conjunto de teste (20%), de modo que as distribuições de idade e sexo nos dois conjuntos fossem estatisticamente próximas. Para o conjunto de dados HCP, o melhor algoritmo foi o Lasso (MAE = 2,75). Os outros dois conjuntos de dados obtiveram resultados significativamente piores em comparação com o HCP. O Cam-CAN teve o melhor desempenho utilizando o algoritmo Lasso LAR (MAE = 7,08 anos) e o IXI, com o algoritmo ARD (MAE = 8,04 anos).

4 MÉTODOS

Esta seção informa as principais etapas do processo de desenvolvimento do modelo.

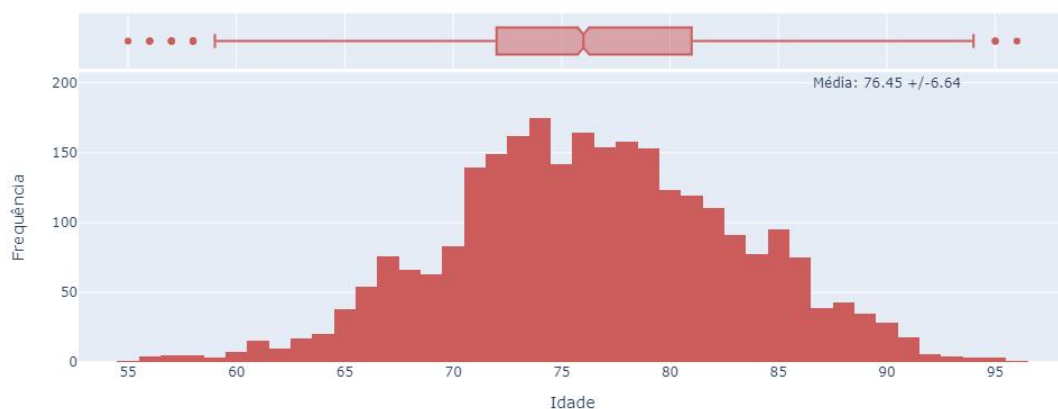
4.1 DATASET

Os estudos do presente trabalho tem por objetivo avaliar a diferença entre a idade cronológica e idade estimada pelo modelo, utilizando um conjunto de dados com uma distribuição específica de idade. Foram utilizados indivíduos com a faixa etária entre 55 e 96 anos, esses que problemas cognitivos podem aparecer com maior incidência quando levamos em conta o quesito do envelhecimento natural do cérebro.

Foram escolhidas imagens de ressonância magnética estrutural contidas no conjunto de dados ADNI (*Alzheimer's Disease Neuroimaging Initiative*). A ADNI tem como objetivo principal testar se a ressonância magnética (MRI), a tomografia por emissão de pósitrons (PET), outros marcadores biológicos e a avaliação clínica e neuropsicológica podem ser combinadas para estimar a progressão do comprometimento cognitivo leve (MCI) e possibilitar o diagnóstico precoce da Doença de Alzheimer (AD).

As imagens escolhidas compreendem o total de 2751 imagens de 666 sujeitos (idade média = $76,45 \pm 6,63$, faixa etária entre 55 e 96 anos). Os dados foram adquiridos em aparelhos de ressonância magnética de alto campo (1,5T e 3,0T), utilizando a sequência T1-MPRAGE. A distribuição dos dados pode ser visualizada por meio da figura 24.

Figura 24 – Distribuição de idades (N=2751)



Apenas os indivíduos diagnosticados como saudáveis e livres de qualquer início de problemas cognitivos foram utilizados, considerando sua idade cronológica próxima a idade cerebral (COLE; FRANKE, 2017).

O dataset foi dividido em treinamento, validação e teste. A maior parte dos dados foi alocada para o conjunto de treinamento, sendo o total de 60% da base, enquanto validação e teste tiveram 20% em cada conjunto. A divisão foi feita a nível de indivíduos, logo, cada conjunto possui indivíduos diferentes; evitando assim utilizar imagens de um indivíduo em conjuntos diferentes, mesmo que as *scans* sejam realizadas em datas distintas. Devido ao custo computacional nas fases de treinamento e ajuste de hiperparâmetros dos modelos, não foi utilizado o método de validação cruzada.

Posteriormente, para avaliar o desempenho do modelo treinado, foi utilizado o conjunto de dados AIBL (*The Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing*). O AIBL tem como objetivo descobrir quais biomarcadores, características cognitivas e fatores de saúde e estilo de vida que determinam o desenvolvimento da doença de Alzheimer. O conjunto de dados contém o total de 630 imagens de 440 sujeitos diagnosticados como livres de comprometimentos cognitivos (idade média = $73,45 \pm 6,38$, faixa etária entre 60 e 92 anos). Os dados foram adquiridos em aparelhos de ressonância magnética de alto campo (1,5T e 3,0T), utilizando a sequência T1-MPRAGE.

4.2 PRÉ-PROCESSAMENTO

Para servir de entrada a modelos de aprendizado de máquina, as imagens de ressonância magnética devem passar por um *pipeline* de pré-processamento, visando reduzir informações que dispersam o objetivo da tarefa a ser realizada.

4.2.1 Co-Registro

A primeira técnica a ser aplicada foi o co-registro rígido, visando utilizar transformações que garantem o alinhamento no mesmo espaço para que as imagens estruturais sejam visualizadas de maneira igual. Nessa etapa, utilizou-se ferramentas do software ANTs (AVANTS; TUSTISON; SONG *et al.*, 2009).

Para alinhamento das imagens no mesmo espaço, é utilizado um atlas padrão, onde as imagens sofrem transformações espaciais com base nas orientações espaciais do atlas. O SRI24 é um atlas baseado em imagens de ressonância magnética anatômica do cérebro humano, gerado por registro não rígido a partir de imagens de 24 indivíduos controle. O SRI24 é comparado com outros atlas frequentemente utilizados (MNI152, Colin27, ICBM452/air12 e ICBM452/warp5) em uma tarefa de normalização de imagens. Os resultados demonstram que seu funcionamento é tão bom quanto os outros quatro atlas avaliados para normalização espacial e, na maioria dos casos, melhor, (ROHLFINGA *et al.*, s.d.).

Todas as imagens passaram pela normalização no espaço SRI24, garantindo o alinhamento e a padronização quanto a orientação espacial. O algoritmo proposto

envolve o mínimo de alterações possíveis na imagem, afim de tornarem mais próximas dos dados reais. Foi feito o co-registro rígido, onde a imagem passou por transformações em 6 graus de liberdade (6 DoF). Tais transformações são compostas por operações de rotação e translação para os três cortes da imagem (axial, coronal e sagital).

4.2.2 Skull Stripping

Posteriormente, foi feito o skull-stripping, que consiste no processo de segmentação do tecido cerebral, removendo tecidos de áreas não cerebrais. Com o objetivo de utilizar um método robusto de extração do cérebro que não seja influenciado por alterações de hardware das máquinas de ressonância magnética, foi utilizado o algoritmo HD-BET (ISENSEE *et al.*, 2019).

O algoritmo HD-BET é disponibilizado publicamente e utiliza um modelo de redes neurais artificiais baseado na arquitetura U-Net para execução da extração do cérebro em diversas sequências de ressonância magnética estrutural. Além de permitir o uso de processamento paralelo, o algoritmo apresentou um bom desempenho em testes feitos por ISENSEE *et al.* (2019), onde foi comparado com cinco algoritmos de extração de cérebro disponíveis publicamente (FSL BET, AFNI 3DSkullStrip, Brainsuite BSE, ROBEX e BEaST), chegando a atingir desempenhos superiores a esses algoritmos quando submetido a testes em datasets de imagens de ressonância magnética.

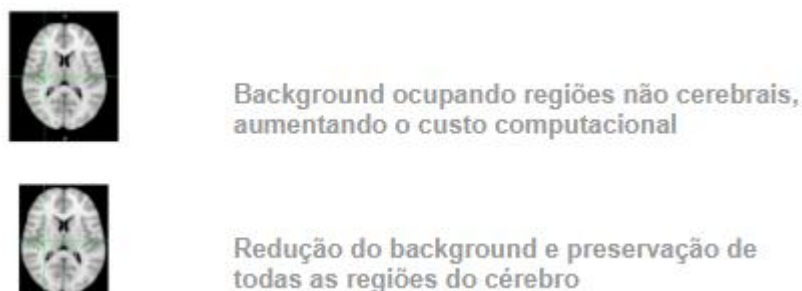
4.2.3 Pipeline de Pré-Processamento da nnU-Net

Após o skull stripping, as imagens passam por outro tipo de *pipeline*, dessa vez os dados de entrada são pré-processados com métodos baseados no trabalho de ISENSEE *et al.* (2018). Esse processo é composto pelos seguintes passos:

Cropping: a primeira etapa do pipeline é o corte (do inglês, cropping). As imagens apresentam grande proporção de *background* (voxels com valor zero). Com o objetivo de eliminar esse volume espacial ocupado pelo *background* e reduzir a complexidade computacional, cada fatia da imagem foi recortada para eliminar regiões com voxel igual a zero. O corte foi feito para abranger toda a região do cérebro dentre as imagens presentes no dataset, resultando nas dimensões de (96 x 96).

Normalização: as imagens de cada paciente foram normalizadas utilizando a normalização z-score, representada pela equação (10). Essa normalização é realizada apenas dentro da máscara de voxels diferentes de zero e todos os valores fora da máscara (*background*) são definidos como 0. Em vez de normalizar toda a imagem, incluindo o *background*, a estratégia proposta produz valores comparativos de intensidade dentro da região do cérebro, independentemente do tamanho da região de fundo ao seu redor. Essa técnica é utilizada em aplicações que envolvem o

Figura 25 – Ilustração do Processo de Corte



pré-processamento de imagens biomédicas, conforme descrito em (ISENSEE *et al.*, 2018a).

$$X[X > 0] = \frac{(X[X > 0] - \mu)}{\sigma} \quad (10)$$

onde,

- X representa o dado de entrada;
- $X[X > 0]$ representa a máscara do cérebro;
- μ é a média dos voxels da máscara do cérebro;
- σ é o desvio padrão dos voxels da máscara do cérebro.

4.2.4 Normalização de Modelos Pré-Treinados na ImageNet

Evitar ordens de grandeza diferentes nos dados permite melhor desempenho em modelos de aprendizado de máquina. A normalização *z-score* foi aplicada em todas as imagens do conjunto (Seção 4.2.3); essas imagens normalizadas serão utilizadas em todos os modelos deste trabalho.

Na Seção 4.3.1, será apresentada uma abordagem que utiliza modelos pré-treinados na ImageNet. Esses modelos tem como requisito aplicar a normalização Min-Max nos dados e transformá-los para o intervalo entre 0 e 1, conforme Equação (11); permitindo maior eficiência no treinamento utilizando as arquiteturas propostas. Portanto, exclusivamente para essa abordagem, após as imagens serem pré-processadas e ter sido aplicado a normalização *z-score*, posteriormente a normalização Min-Max é aplicada.

$$X_{normalizado} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (11)$$

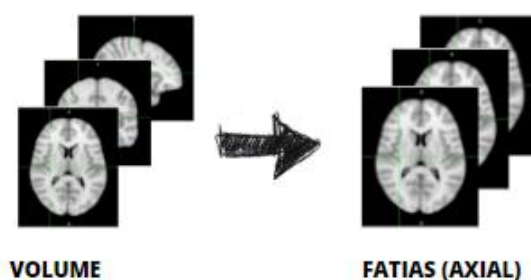
onde,

- X é o dado de entrada;
- X_{min} e X_{max} são os pixels de menor e maior valor, respectivamente.

4.2.5 Seleção de fatias

Partes do volume com maior proporção de *background* e pouca informação do cérebro foram eliminadas estrategicamente. Conforme BASHYAM *et al.* (2020), 40 fatias (imagens 2D) centrais do eixo axial foram selecionadas para servir de entrada no modelo. Deste modo o modelo não necessita de processar todo o volume, inclusive áreas de predominância de regiões não cerebrais.

Figura 26 – Ilustração do Processo de Transformação em Fatias 2D



4.3 ABORDAGENS DE REDES NEURAIIS UTILIZADAS

Esta seção apresenta duas abordagens diferentes para a tarefa de predição da idade cerebral. A primeira consiste em utilizar um modelo de classificação, retirar as camadas de classificação e substituí-las pela tarefa de regressão, possibilitando que o modelo trabalhe como estimador da idade cerebral. Na segunda abordagem, foi utilizado um modelo de segmentação pré-treinado em imagens de ressonância magnética cerebral que, posteriormente, também é modificado para a tarefa de regressão.

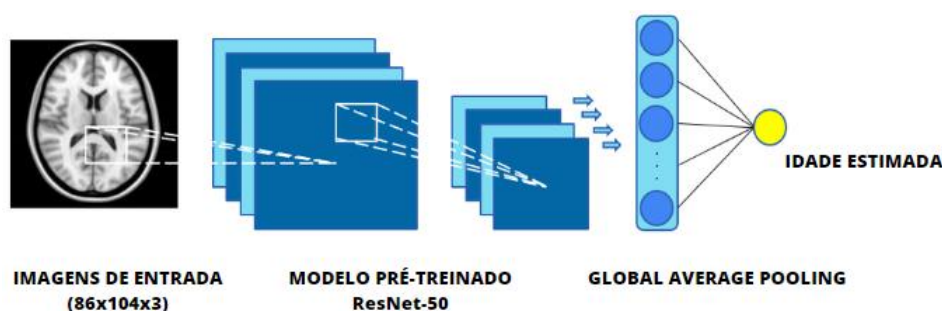
Antes de fazer treinamentos e ajustes em ambas as abordagens, com propósito de comparar tipos de inicialização de redes neurais convolucionais (pesos aleatórios e pesos pré-treinados na ImageNet), dois modelos de arquitetura ResNet50 foram treinados. Por ser um modelo leve e de alto desempenho, a ResNet50 foi escolhida para esse teste. As mesmas configurações foram mantidas em ambos os testes, alterando apenas o método de inicialização dos pesos.

4.3.1 Modelos Pré-Treinados em Classificação

Na primeira abordagem, foram utilizados modelos de Redes Neurais Convolucionais 2D. A preferência do uso de uma arquitetura 2D ao invés de uma 3D é dada pela possibilidade de usar pesos pré-treinados da ImageNet para inicialização do modelo. Para validar o critério de escolha, na Seção 5.1.1 constam testes feitos com a finalidade de comparar um modelo com inicialização de pesos aleatória e inicialização com pesos pré-treinados na ImageNet.

Essa tarefa foi feita utilizando o método de aprendizado por transferência. Foram escolhidos modelos pré-treinados em tarefas de classificação. Cada modelo utilizado teve a camada de classificação substituída por uma camada densa com apenas um neurônio e ativação linear, que representa idade estimada. Portanto, os modelos que antes eram classificadores foram modificados para a tarefa de regressão. Nas camadas finais, foi aplicado um Global Average Pooling e a saída é apenas uma camada densa de ativação linear, contendo uma unidade.

Figura 27 – Representação da Arquitetura do Modelo



Para essa abordagem, quatro arquiteturas de Redes Neurais Convolucionais foram escolhidas. Dentre elas está a InceptionResnetV2 (SZEGEDY *et al.*, 2017), utilizada no trabalho de BASHYAM *et al.* (2020), a qual obteve bom desempenho na tarefa de predição da idade cerebral, alcançando o MAE de 3,7 anos no conjunto de dados utilizado. Outras arquiteturas escolhidas foram a ResNet50 e a ResNet101. Modelos ResNet apresentam bom desempenho utilizando menos parâmetros que a InceptionResnetV2. Utilizou-se também EfficientNet-B4, uma das atuais arquiteturas do estado da arte. Foi escolhido o modelo B4, que apresenta um bom tradeoff entre desempenho e complexidade. Em comparação com modelos ResNet e a InceptionResnetV2, a EfficientNet-B4 apresenta melhor desempenho utilizando menor número de parâmetros quando submetida a testes feitos na ImageNet (TAN; LE, 2019).

4.3.2 Modelo Pré-Treinado em Segmentação

Utilizar modelos pré-treinados em conjuntos de dados de neuroimagem pode apresentar um desempenho superior para extrair características desse tipo de dado do que modelos pré-treinados com imagens diferentes da entrada desejada. Uma vez que a rede reconhece padrões nas imagens, utilizar o método de aprendizado por transferência permite aproveitar as informações aprendidas pelo modelo, como detecção de bordas e texturas nas imagens cerebrais e aplicá-las em outras tarefas de aprendizado de máquina.

Métodos robustos de redes neurais convolucionais para a tarefa de predição da idade cerebral tem se mostrado altamente eficientes,(POLONI; FERRARI; INITIATIVE

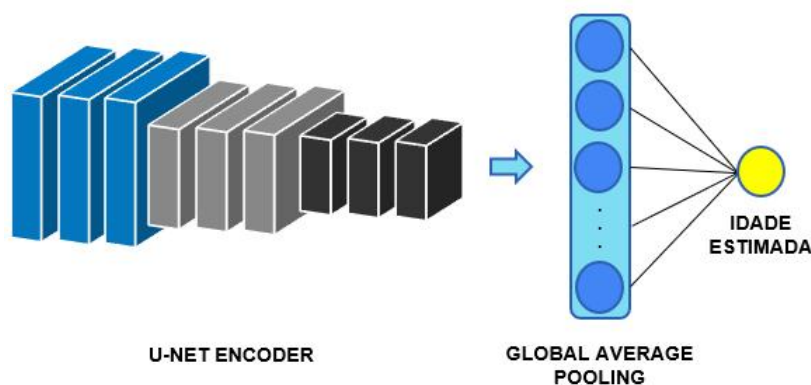
et al., 2022), (COLE *et al.*, 2017), (BASHYAM *et al.*, 2020). Essa sessão propõe a aplicação de um outro método para extração de características das imagens: utilizar uma rede de segmentação pré-treinada em uma tarefa de segmentação de tumores cerebrais para segmentar os dados de entrada; posteriormente, utilizá-los na tarefa de predição da idade cerebral.

Segmentar as regiões cerebrais utilizando um modelo robusto e pré-treinado com dados de ressonância magnética cerebral pode fazer com que a rede extraia bons recursos das amostras para, posteriormente, serem utilizados na tarefa de estimação da idade cerebral.

Para a tarefa de segmentação das imagens de ressonância magnética, utilizou-se uma rede de arquitetura U-Net com auxílio do framework nnU-Net. O modelo não é pré-treinado na ImageNet, portanto não é necessário aplicar a normalização Min-Max, apenas o *z-score* do framework nnU-Net. A U-Net foi escolhida pelo fato de ser convencionalmente utilizada em aplicações de segmentação relacionadas à imagens médicas, gerando bons resultados até mesmo em conjunto de dados pequenos. A rede utilizada foi pré-treinada para tarefas de segmentação de tumor cerebral em imagens de MRI ponderadas em T1, pertencentes ao dataset da competição BraTS (MENZE *et al.*, 2015) na edição de 2020.

A arquitetura da rede foi modificada para a tarefa de predição da idade cerebral, retirando as camadas referentes ao decoder e deixando apenas o encoder. Uma camada de Global Average Pooling e uma camada densa com uma única unidade são acopladas à rede. Dessa forma, as imagens pertencentes ao dataset utilizado para a tarefa de predição da idade cerebral são utilizadas como entrada de um modelo pré-treinado com imagens cerebrais. A figura 28 demonstra a arquitetura proposta.

Figura 28 – Representação da Arquitetura proposta



4.3.3 Processo de Treinamento

O Processo de Treinamento é utilizado em ambas as abordagens. Durante o treinamento, cada fatia é considerada uma amostra independente, entrando de forma aleatória na rede, como proposto em BASHYAM *et al.* (2020). No final, há uma predição para cada fatia dos indivíduos. O processo é ilustrado pela figura 29.

Os modelos são inicializados com pesos pré-treinados. Na abordagem que utiliza modelos de classificação (seção 4.3.1), o modelo é inicializado com pesos pré-treinados na ImageNet. Na segunda abordagem (seção 4.3.2), a inicialização é feita com base em pesos de um modelo pré-treinado na tarefa de segmentação semântica em imagens de ressonância magnética cerebral - mais especificamente na tarefa de segmentação de tumor cerebral da competição (MENZE *et al.*, 2015) de 2020.

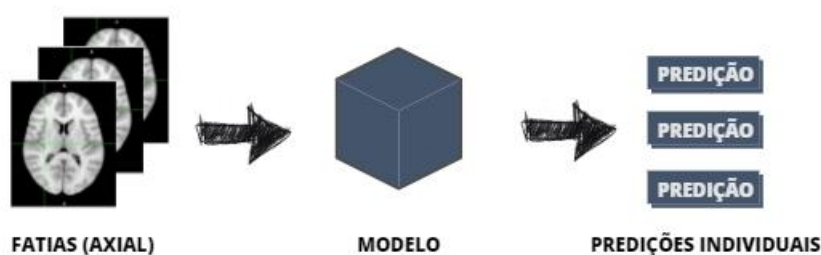
Como função custo, utilizou-se o Erro Médio Quadrático (MSE), permitindo a penalização de elevados erros entre a idade cerebral e idade prevista (equação (12)).

$$MSE = \frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2 \quad (12)$$

onde,

- N é o número total de observações;
- y é o rótulo;
- \hat{y} é o valor estimado.

Figura 29 – Processo de Treinamento



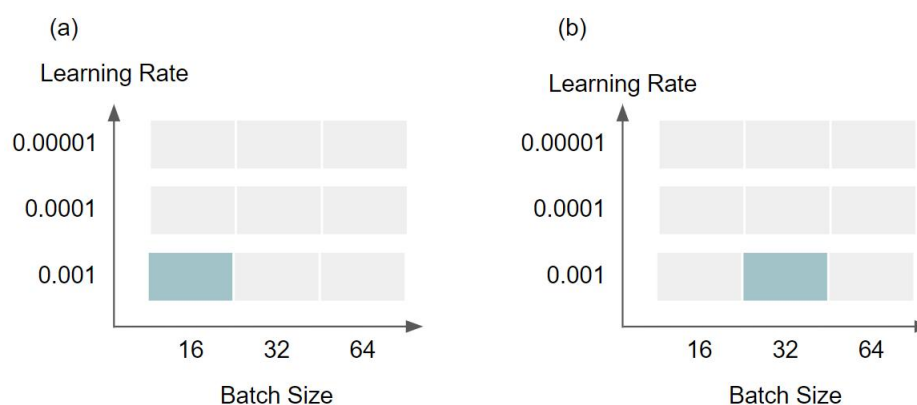
4.3.4 Otimização de Hiperparâmetros

A maioria dos algoritmos de aprendizado de máquina requer várias configurações para controlar seu comportamento na etapa de aprendizado. Essas configurações são chamadas de hiperparâmetros. Os valores dos hiperparâmetros não são adaptados pelo próprio algoritmo, embora se possa projetar um procedimento onde o algoritmo aprende os melhores hiperparâmetros, (GOODFELLOW; BENGIO; COURVILLE, 2016).

Para avaliar a importância da alteração de hiperparâmetros sobre os resultados do modelo, testes de definição da taxa de aprendizado e tamanho do batch foram feitos. Esses testes foram feitos através do método *grid search*, utilizando três valores para a taxa de aprendizado e tamanho do lote. Um hiperparâmetro por vez foi alterado a cada iteração, resultando em 9 diferentes combinações (Figura 30). Outros hiperparâmetros foram avaliados de forma empírica e arbitrária, como: escolha do otimizador, *early stopping* e outras formas de decaimento da *learning rate*. O modelo proposto apresentou ser pouco sensível a alteração desses hiperparâmetros, entretanto, a definição do otimizador Adam (KINGMA; BA, 2014), *learning rate* e *batch size* foram os que mais geraram impacto. Por esse motivo, foram os escolhidos para a fase de ajuste dos hiperparâmetros.

Segundo GOODFELLOW; BENGIO; COURVILLE (2016), o treinamento com tamanhos de lote muito pequenos podem exigir uma pequena taxa de aprendizado para manter a estabilidade devido à alta variação na estimativa do gradiente, além do tempo total de execução que pode ser muito alto devido à necessidade de realizar mais etapas para observar todo o conjunto de treinamento. Valores comuns escolhidos para tamanho de lote variam de 32 a 256; também é comum utilizar o tamanho 16 para modelos grandes, (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 30 – Ajuste dos Hiperparâmetros em Grade. Os blocos destacados representam uma das 9 diferentes combinações. a) Representação da *learning rate* de 10^{-3} e *batch size* de tamanho 16; b) Learning rate de 10^{-3} e batch size de tamanho 32



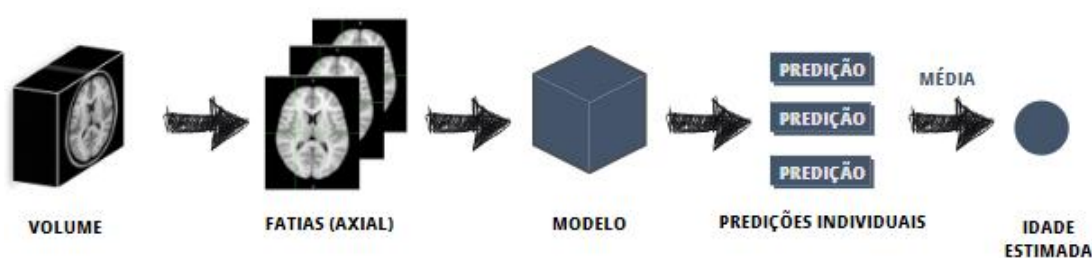
4.4 PÓS-PROCESSAMENTO

Tendo sido o modelo treinado, a próxima etapa é o pós-processamento. Nessa etapa, ajustes serão feitos para fusão das predições feitas em cada fatia e ajuste do viés nas predições.

4.4.1 Fusão das Predições

Nas etapas de validação e inferência, deve-se evitar que fatias aleatórias de indivíduos diferentes sejam enviadas ao modelo. Dessa forma, as 40 fatias referentes ao paciente são inseridas no modelo treinado e a média das predições individuais é utilizada como predição da idade cerebral (Figura 31). Da mesma forma que o processo de treinamento, a etapa de validação e inferência é feita da mesma forma para ambas as abordagens.

Figura 31 – Processo de Inferência



Para avaliar a diferença absoluta entre a idade prevista e a idade cronológica, a métrica de avaliação escolhida foi o Erro Absoluto Médio (Equação (13)).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y - \hat{y}| \quad (13)$$

O coeficiente de determinação r^2 e o MAPE (do inglês, *Mean Absolute Percentage Error*) também são métricas que foram utilizadas para avaliar o desempenho do modelo.

4.4.2 Correção do Viés nas Predições

Estudos relacionados a estimativa de idade utilizando neuroimagens demonstram que, quanto mais longe uma amostra estiver do conjunto de treino, maior ou menor será a idade média das predições. Este efeito resulta na subestimação da idade prevista do cérebro para participantes mais velhos e na sobrestimação para participantes mais jovens, o que aumenta à medida que o desempenho do modelo diminui (POPESCU *et al.*, 2021). Esse viés presente nas predições pode ser corrigido por meio de métodos matemáticos. Nessa seção, iremos abordar uma solução para esse problema.

O conjunto de dados de validação foi utilizado para ajustar o viés das predições. Tendo os registros dos valores de estimativas de idade e os rótulos, um modelo de

regressão linear pode ser treinado com o objetivo de obter o melhor ajuste da reta sob o conjunto. A equação (14) representa o método de ajuste do viés.

$$y' = \alpha \hat{y} + \beta \quad (14)$$

A representação das predições corrigidas sob o conjunto de dados de validação é dada por y' . As predições do modelo são dadas por \hat{y} . Os valores de α e β são parâmetros de uma regressão linear, representando o coeficiente angular e a interceptação da reta, utilizados para corrigir as predições.

Para avaliar a eficiência desse método, tendo em base um modelo treinado, o conjunto de validação foi dividido em 5 *folds*. As predições de cada fold passaram pelo método de ajuste de viés. Ao final de cada fold, é obtido os valores de MAE e o correspondente MAE após a correção do viés. A média dos resultados é feita no final do processo para comparar se houveram mudanças significantes na métrica de avaliação.

A representação do processo de correção do viés pode ser visualizada na figura 32.

4.4.3 Avaliação de Desempenho em um Conjunto de Dados Externo

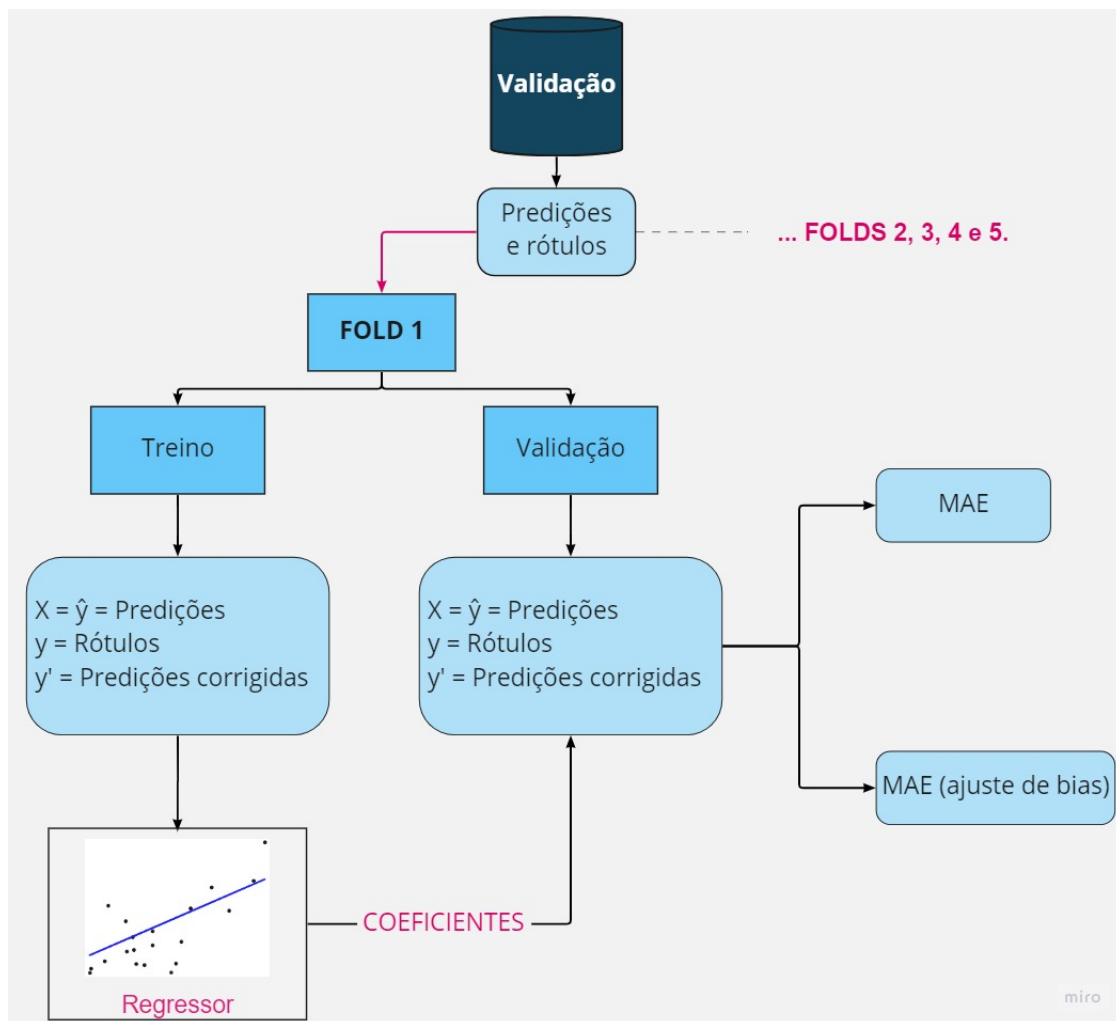
Após a conclusão do processo de criação e treinamento do modelo, a etapa seguinte é realizar uma avaliação do seu desempenho em um conjunto de dados diferente. Essa avaliação é útil para avaliar a capacidade de generalização do modelo para distribuições não vistas no treinamento.

Para fazer essa avaliação, é necessário utilizar um conjunto de dados externo, ou seja, um conjunto de dados que não foi utilizado no treinamento do modelo. Isso é importante para garantir que o modelo não esteja simplesmente memorizando o conjunto de dados de treinamento, mas sim aprendendo padrões gerais que possam ser aplicados em diferentes situações. A inferência será realizada no conjunto de dados AIBL, mencionado na Seção 4.1. O *dataset* foi escolhido como referência por ser um conjunto de dados que já foi utilizado em trabalhos da mesma linha de pesquisa, como em BASHYAM *et al.* (2020).

4.4.4 Recursos Computacionais

O código foi desenvolvido em Python, utilizando o *framework* de *deep learning* PyTorch. A infraestrutura de hardware utilizada durante os experimentos foi fornecida pela plataforma de nuvem Google Cloud, com recursos de 30 GB de memória RAM e uma GPU Tesla T4 com 16 GB de memória de vídeo dedicada.

Figura 32 – Processo de Correção do Viés das Predições, representando o Fold 1.



5 RESULTADOS

Essa sessão apresenta os resultados dos processos de desenvolvimento do modelo.

5.1 ABORDAGEM UTILIZANDO MODELOS PRÉ-TREINADOS DE CLASSIFICAÇÃO

5.1.1 Comparação de métodos de inicialização dos pesos

A Figura 33 mostra um modelo ResNet-50 inicializado com os pesos da ImageNet e inicializado com pesos aleatórios. O modelo pré-treinado na ImageNet obteve melhor performance nas curvas de perda de validação e erro absoluto médio de validação. O treinamento durou aproximadamente 90 minutos para 50 épocas. Foi observado que 50 épocas é suficiente para avaliar o comportamento do modelo, pois as curvas de treinamento convergem e não apresentam melhores resultados quando o modelo é treinado por mais épocas.

Figura 33 – Comparação entre um modelo ResNet50 com inicialização de pesos pré-treinados na ImageNet e um modelo com inicialização de pesos aleatórios



5.1.2 Ajuste de Hiperparâmetros

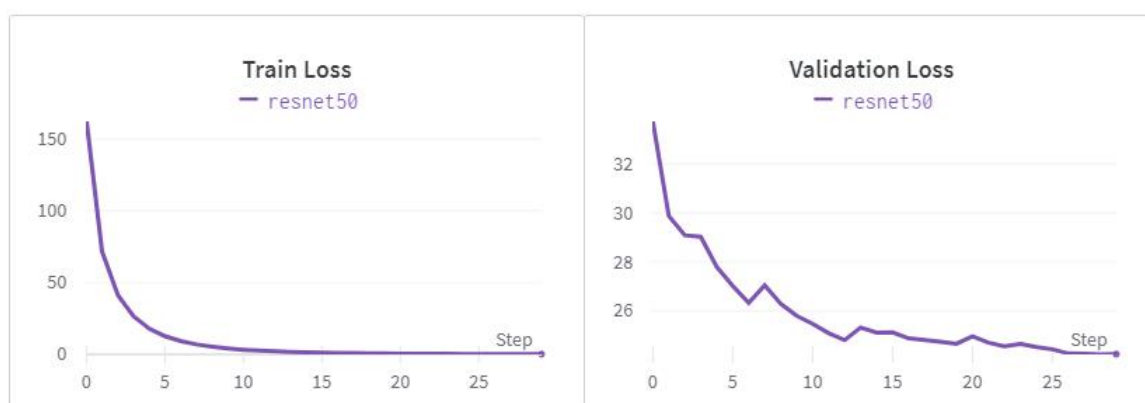
Três tipos de tamanhos de batch foram escolhidos para essa etapa: 16, 32 e 64. Para a taxa de aprendizado, foram escolhidos os valores de 10^{-3} , 10^{-4} e 10^{-5} . A performance de cada modelo foi avaliada em 10 épocas. Outros testes contendo diferentes hiperparâmetros foram feitos, sendo estes o que causaram maior efeito no desempenho do modelo. Tendo em vista a otimização quanto ao tempo de treinamento, optou-se pela escolha do batch de tamanho 64 e definiu-se a taxa de aprendizado com o valor de 10^{-3} . Dessa forma, podemos utilizar um modelo que performa mais rápido devido a escolha do batch. Os resultados de cada treinamento podem ser visualizados na tabela 1.

Tabela 1 – Tuning da Taxa de aprendizado x Tamanho do Lote - ResNet-50

Learning Rate	Batch Size	Tempo (CPU)	Val MAE
10^{-3}	64	27m58s	3,38
10^{-3}	32	32m12s	3,46
10^{-3}	16	44m11s	3,48
10^{-4}	64	28m32s	3,46
10^{-4}	32	33m20s	3,36
10^{-4}	16	45m10s	3,34
10^{-5}	64	28m23s	3,57
10^{-5}	32	31m46s	3,59
10^{-5}	16	52m16s	3,39

O modelo foi inicializado com pesos pré-treinados na ImageNet. Como função de custo, utilizou-se o Erro Médio Quadrático (MSE), permitindo a penalização de elevados erros entre a idade cerebral e idade prevista (Equação (12)). O otimizador escolhido foi o Adam (KINGMA; BA, 2014) com taxa de aprendizado inicial de 10^{-3} e decaimento com um fator multiplicador de 0,1 até a última época. O tamanho do *batch* foi definido como 64 e o modelo foi treinado durante o total de 30 épocas. Após 30 épocas o modelo não apresenta melhoria significativa nos resultados. As curvas de treinamento e validação representadas pela figura 34 demonstram que não ocorreu *overfitting* durante a fase de treinamento, ambas seguem em decaimento até a convergência.

Figura 34 – Loss de treino e de validação - ResNet50



5.1.3 Correção de Viés

A fim de avaliar a eficiência do método de correção do viés nas predições, testes em 5 *folds* foram realizados, de acordo com a seção 4.4.2. Os resultados deste método se encontram na tabela 2. A média dos resultados demonstra pouca significância nas

predições, portanto, esse método não será aplicado nos demais experimentos desta dissertação.

Tabela 2 – Resultados do Teste de Correção de Viés aplicado em 5 Folds.

Fold	Fold MAE	MAE [Correção de Viés]
1	4,00	3,90
2	3,39	3,40
3	3,45	3,35
4	2,75	2,92
5	2,71	2,88
Média	3,26	3,29

5.1.4 Avaliação de Desempenho

Outras arquiteturas foram utilizadas e o mesmo procedimento foi seguido. A tabela 3 reporta os resultados das predições feitas no conjunto de validação.

Tabela 3 – Comparação de resultados obtidos no conjunto de validação: ResNet-50, ResNet-101, InceptionResnetV2 e EfficientNet-B4.

Arquitetura	MAE (anos)	r^2	MAPE(%)	Tempo de Treinamento
EfficientNetB4	3,26	0,58	4,34	2h 17m
InceptionResNetV2	3,48	0,54	4,64	2h 30m
ResNet-50	3,31	0,57	4,41	1h 22m
ResNet-101	3,32	0,57	4,46	1h 51m

Embora a EfficientNetB4 tenha um desempenho marginalmente melhor do que a ResNet-50, o seu tempo de treinamento é consideravelmente maior. Por esse motivo, a ResNet-50 foi escolhida para o restante dos resultados desta seção.

Após a definição de hiperparâmetros, não é necessário utilizar novamente o conjunto de validação. Os dados que antes eram utilizados na validação foram unidos aos dados de treinamento, dessa forma podemos treinar um modelo com um conjunto de dados maior. O novo split fica dividido em 80% para treino e 20% para teste.

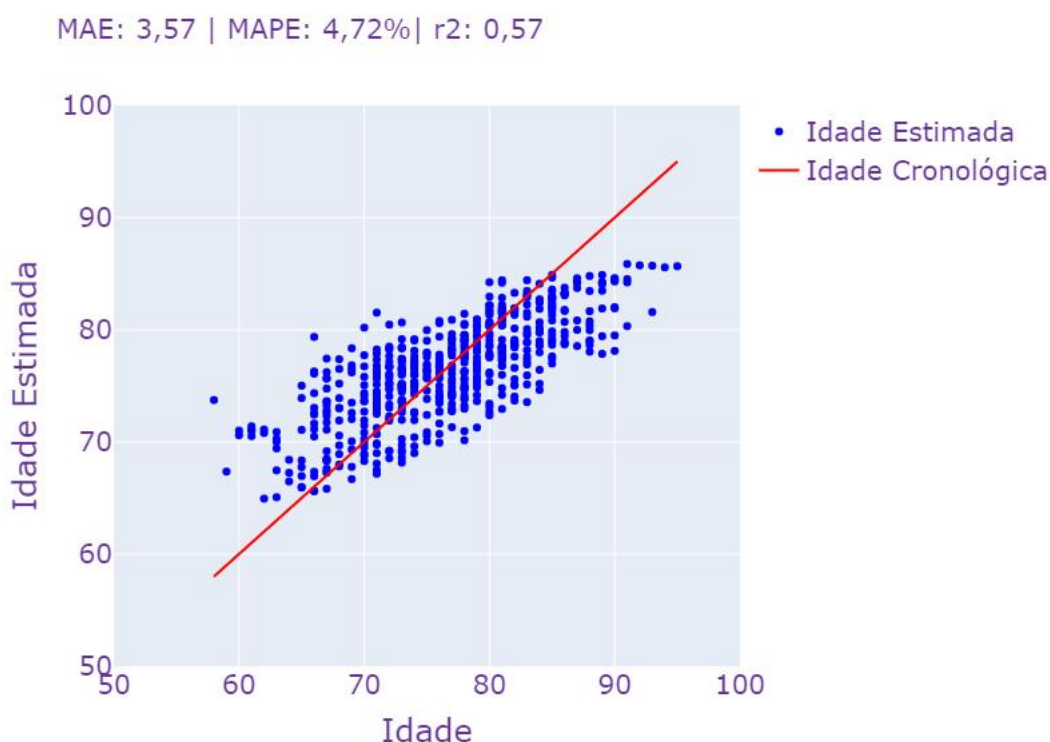
Após o treinamento do modelo, foi realizada a inferência no conjunto de teste. Os resultados estão reportados na tabela 4. Mediante a figura 35 pode-se analisar que o erro absoluto médio alcançado foi de 3,57 anos. A figura 35 demonstra um gráfico de dispersão relacionando a idade predita à idade cronológica do paciente. A linha vermelha refere-se à idade real do paciente e os pontos em azul representam a idade estimada.

A figura 36 demonstra os erros residuais do modelo. A maior parte da distribuição está focada em erros de ± 5 anos. Percebe-se que há poucos casos onde há valores discrepantes, contendo resíduos que chegam a aproximadamente ± 15 anos.

Tabela 4 – Resultado do modelo ResNet-50 retreinado, obtido no conjunto de teste.

Arquitetura	MAE (anos)	r^2	MAPE(%)	Tempo de Treinamento
ResNet-50	3,57	0,57	4,72	2h 1m

Figura 35 – Resultado da Inferência no Conjunto de Teste - Arquitetura ResNet50 retreinada.



5.2 ABORDAGEM UTILIZANDO MODELO PRÉ-TREINADO DE SEGMENTAÇÃO

5.2.1 Ajuste de Hiperparâmetros

Da mesma forma que foi feito na Seção 5.1.2, a fase de ajuste de hiperparâmetros compõe-se por diferentes tipos de taxa de aprendizado e tamanho do lote que foram utilizados em um treinamento em grade. Cada treinamento durou o total de 10 épocas com taxa de aprendizado constante. Os resultados estão apresentados na tabela 5.

Para o treinamento da arquitetura proposta, foi utilizado o otimizador Adam com taxa de aprendizado definida em 10^{-3} . O MSE foi utilizado como função de perda e o MAE como métrica de avaliação. O modelo foi treinado durante 30 épocas, com taxa de aprendizado contendo um fator multiplicador de 0,1 até a última época. O tamanho do lote recebeu o valor de 64.

Figura 36 – Erros residuais no Conjunto de Teste - ResNet50

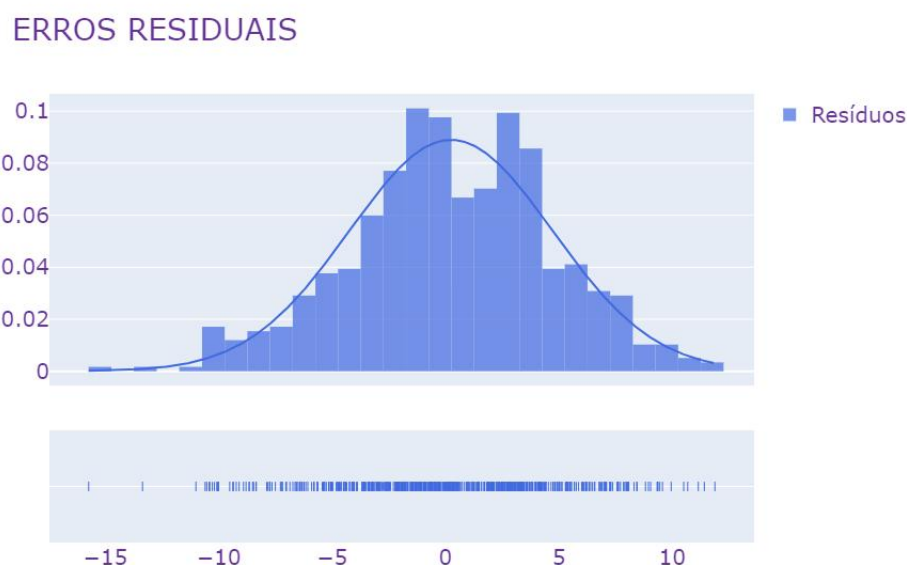


Tabela 5 – Tuning da Taxa de aprendizado x Tamanho do Lote - UNet.

Learning Rate	Batch Size	Tempo	Val MAE
10^{-3}	64	24m27s	3,24
10^{-3}	32	24m27s	3,31
10^{-3}	16	38m56s	3,25
10^{-4}	64	24m26s	3,32
10^{-4}	32	24m14s	3,19
10^{-4}	16	27m59s	3,2
10^{-5}	64	24m10s	52,56
10^{-5}	32	24m36s	20,60
10^{-5}	16	27m54s	3,23

5.2.2 Avaliação de Desempenho

Essa seção refere-se aos resultados do modelo pré-treinado em segmentação. O modelo proposto utiliza a arquitetura U-Net como regressor, eliminando as camadas de classificação.

Inicialmente, após a definição de hiperparâmetros, foi feito um treinamento para avaliar se o modelo está causando *overfitting*. As curvas de perda de treinamento e validação estão representadas na figura 37. Ambas estão em decaimento até a última época, sem apresentar sinais de *overfitting*. A tabela 6 reporta os resultados das predições feitas no conjunto de validação.

Após a verificação do comportamento do modelo quanto ao *overfitting*, o conjunto de validação foi unido com o de treinamento, criando um split de 80% para

Figura 37 – Loss de treino e de validação - UNet

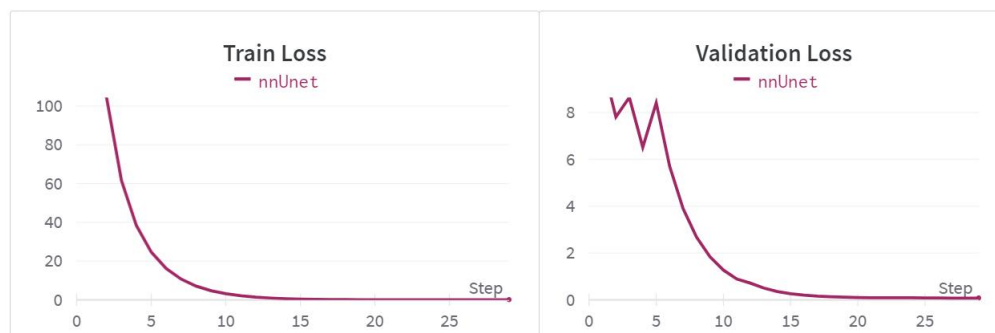


Tabela 6 – Resultados da arquitetura U-Net no conjunto de validação.

Arquitetura	MAE (anos)	r^2	MAPE(%)	Tempo de Treinamento
U-Net	3,14	0,61	4,19	1h 30m

treinamento e 20% teste. Após treinar o modelo, foi realizada a inferência no conjunto de teste e o valor do erro absoluto médio alcançou 3,39 anos. O tempo total de treinamento foi de 1h47min. Os resultados estão reportados na tabela 7.

Tabela 7 – Resultado do modelo U-Net retreinado, obtido no conjunto de teste.

Arquitetura	MAE (anos)	r^2	MAPE(%)	Tempo de Treinamento
U-Net	3,39	0,61	4,46	1h 47m

Por meio das Figuras 38 e 39, pode-se visualizar as previsões no conjunto de teste e os erros residuais. Os resíduos tendem a concentrar no intervalo ± 5 anos, em sua maior proporção. Há casos raros com valores elevados de erros residuais, porém com baixa representatividade da amostra avaliada.

5.3 AVALIAÇÃO NO CONJUNTO DE DADOS AIBL

Devido ao seu melhor desempenho, a arquitetura U-Net foi selecionada como modelo proposto para esta avaliação. Os resultados da inferência no conjunto de dados AIBL estão apresentados na Tabela 8. As Figuras 40 e 41 mostram, respectivamente, o gráfico de dispersão das previsões e os erros residuais.

Tabela 8 – Resultado da Inferência no Conjunto de Dados AIBL - Arquitetura U-Net retreinada.

Arquitetura	MAE (anos)	r^2	MAPE(%)
U-Net	3,77	0,46	5,2

Figura 38 – Resultado do teste utilizando a arquitetura U-Net

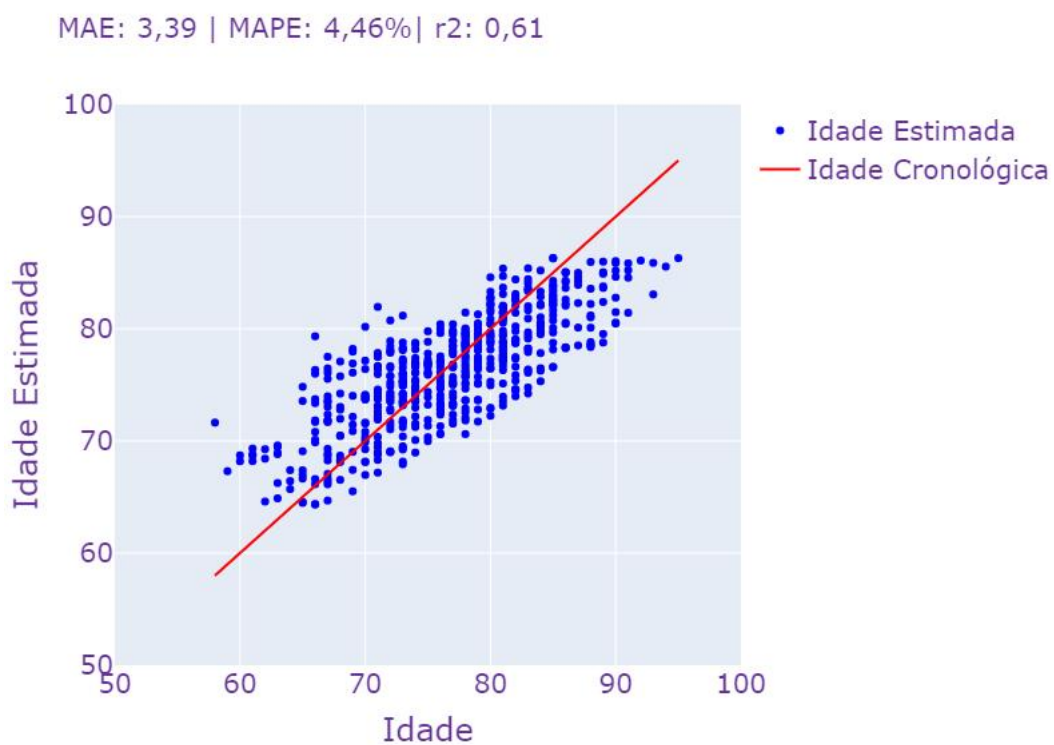


Figura 39 – Erros residuais - U-Net

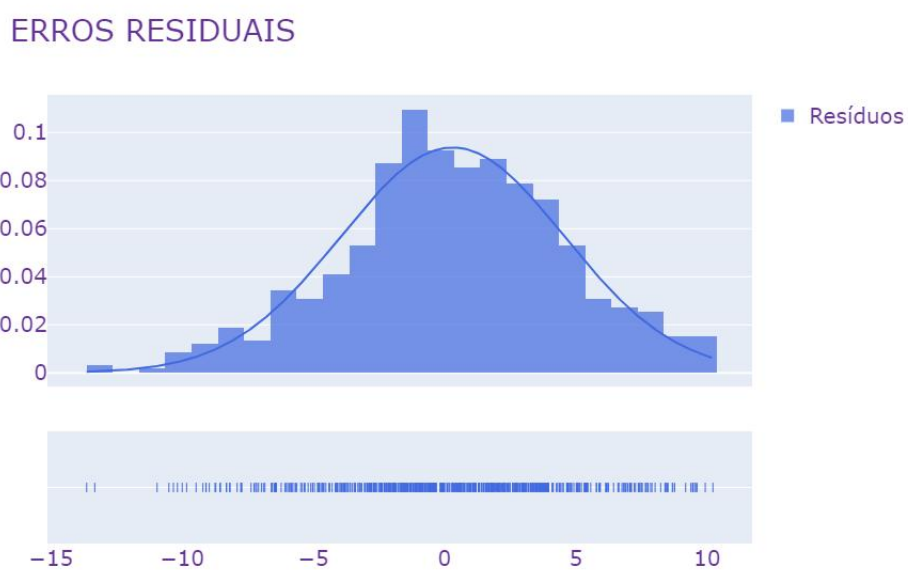


Figura 40 – Resultado da Inferência no Conjunto de Dados AIBL - Arquitetura U-Net retreinada.

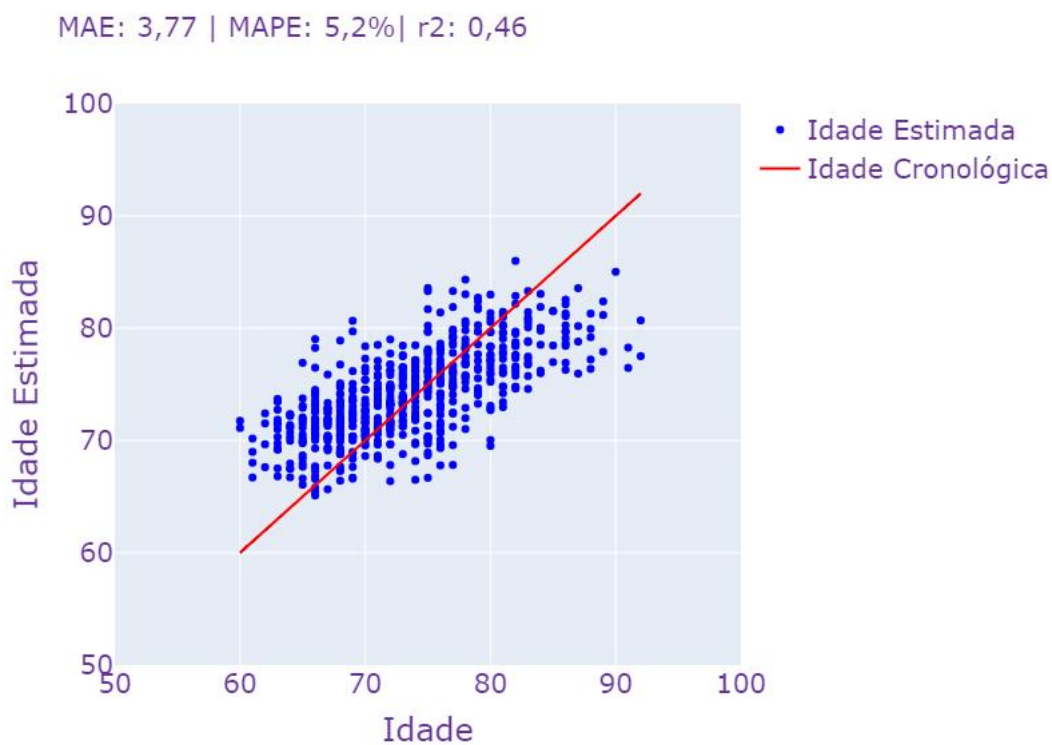
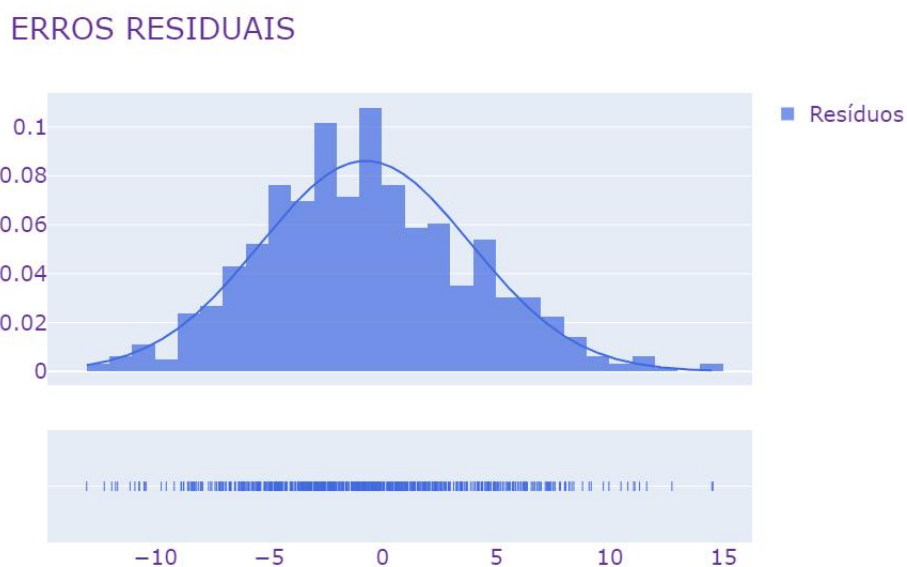


Figura 41 – Erros residuais da Inferência no Conjunto de Dados AIBL - Arquitetura U-Net retreinada.



5.4 DISCUSSÃO

Na primeira abordagem, as arquiteturas EfficientNetB4 e ResNet apresentaram os menores valores de MAE e os maiores coeficientes de determinação r^2 quando avaliados no conjunto de validação. Entretanto, considerando o tempo de treinamento, essas arquiteturas são mais custosas computacionalmente que as outras propostas. Aumentar a profundidade do modelo ResNet-50 para ResNet-101 não apresentou ganhos significativos na métrica avaliada. Entre os modelos propostos, a arquitetura ResNet-50 teve a melhor relação entre desempenho e custo computacional.

Já na segunda abordagem, a arquitetura U-Net proposta apresentou melhor desempenho em ambos os critérios mencionados: MAE e tempo de treinamento. Com um MAE de 3,14 obtido no conjunto de validação e tempo de treinamento de 1h 30min, a U-Net estabeleceu melhores previsões da idade cerebral do que as outras arquiteturas propostas e com menor custo computacional.

Após ter apresentado melhor desempenho em relação às demais arquiteturas, a U-Net foi submetida a um processo de re-treinamento utilizando uma divisão dos dados em 80% para treinamento e 20% para teste. Em seguida, os resultados da inferência no conjunto de teste foram comparados com outros modelos do estado da arte, conforme apresentado na Tabela 9. É importante ressaltar que os resultados da Tabela 9 não são perfeitamente comparáveis, devido ao fato de utilizarem diferentes conjuntos de treinamento e teste.

Tabela 9 – Comparação entre U-Net Proposta e Modelos do Estado da Arte

Arquitetura	MAE (anos)
(HAN <i>et al.</i> , 2022)	2,78
(WOOD <i>et al.</i> , 2022)	2,97
(POLONI; FERRARI; INITIATIVE <i>et al.</i> , 2022)	3,31
(BASHYAM <i>et al.</i> , 2020)	3,72
(COLE <i>et al.</i> , 2017)	4,65
U-Net proposta	3,39

A proposta de utilizar um modelo pré-treinado na tarefa de segmentação de neuroimagens resultou em um desempenho promissor para a solução do problema proposto, quando comparamos com trabalhos mencionados na Seção 3. O modelo proposto utiliza um conjunto de dados público, tornando o acesso menos limitado quando consideramos alguns conjuntos de dados utilizados nos trabalhos mencionados.

Inicialmente, foram desenvolvidos modelos de Redes Neurais Convolucionais 3D antes de optar pela estratégia de utilizar arquiteturas de Redes Neurais Convolucionais 2D. Além da motivação em utilizar modelos robustos pré-treinados em arquiteturas 2D, os modelos 3D apresentavam maior custo computacional, demandando um tempo significativo para treinamento e ajuste de hiperparâmetros.

Outras abordagens para fusão de predições foram testadas ao longo do estudo, incluindo a média ponderada das previsões e o treinamento de um regressor em cima das características extraídas das imagens. Esses resultados foram combinados com os resultados do modelo de redes neurais convolucionais, formando um modelo *ensemble*. No entanto, os resultados não mostraram melhora significativa em comparação com o treinamento isolado do modelo de redes neurais convolucionais. Além disso, essa abordagem aumentou significativamente os custos computacionais devido à fusão de dois modelos de aprendizado de máquina.

Existem algumas limitações que podem ser estudadas para o desenvolvimento de novos modelos com melhor desempenho no futuro. Por exemplo, as predições não se ajustam completamente à linha da idade cronológica, onde pacientes mais jovens podem apresentar idade maior e pacientes mais idosos tendem a apresentar idade menor, como foi observado na avaliação do modelo em um conjunto de dados diferente do que foi treinado, conforme apresentado na Seção 5.3. Outra abordagem que pode ser investigada é a adição de novos conjuntos de dados, a fim de aumentar a população estudada e fornecer ao modelo uma maior quantidade de amostras de treinamento. Fornecer uma quantidade maior e mais diversificada de amostras é de extrema importância na construção de modelos de aprendizado de máquina, especialmente em áreas como a neuroimagem, onde a variabilidade entre indivíduos pode ser muito grande. A adição de novos conjuntos de dados pode permitir que o modelo aprenda padrões mais robustos, melhorando a generalização para casos não vistos durante o treinamento.

6 CONCLUSÃO

O presente trabalho tem o objetivo de construir um modelo capaz de prever a idade cronológica de um paciente utilizando imagens de ressonância magnética cerebrais. Duas abordagens foram utilizadas. A primeira abordagem utiliza modelos de classificação pré-treinados na ImageNet. As arquiteturas desses modelos são alteradas para a tarefa de predição da idade cerebral, substituindo a camada de classificação por uma camada que realiza a operação de regressão. A segunda abordagem utiliza um modelo pré-treinado na tarefa de segmentação de imagens de tumor cerebral. A arquitetura desse modelo também é alterada, tendo seu formato transformado para tarefa de predição da idade cerebral.

Ambas as abordagens demonstraram resultados promissores na tarefa de predição da idade cerebral. O desempenho da U-Net demonstrou ser melhor que os outros modelos propostos, o resultado atingiu o menor MAE com a menor complexidade computacional.

O método de correção de viés demonstra ser pouco significativo para melhorar o desempenho do modelo. O erro absoluto médio altera de forma marginal quando aplicamos a correção nas predições. Para evitar esforços computacionais que não levam a diferenças significativas no desempenho do modelo, optou-se por não utilizar essa técnica.

Para explorar o melhor potencial de ambos os modelos, novos conjuntos podem ser adicionados, contendo diferentes características de pacientes, outras modalidades de ressonância magnética e diferentes configurações de aquisição das imagens.

Tendo em mãos o estimador de idade proposto, pode-se aplicar a inferência em pacientes com declínio cognitivo. Para esses indivíduos, a idade estimada pelo modelo deve ser maior que a idade cronológica. Dessa forma é possível auxiliar na detecção previa de problemas cognitivos, apoiando equipes médicas no diagnóstico prematuro do problema.

6.1 TRABALHOS FUTUROS

- Utilizar o modelo em datasets contendo pacientes com declínio cognitivo e avaliar o comportamento;
- Adicionais outros conjuntos de dados, com diferentes distribuições e tipos de aquisição.

REFERÊNCIAS

- ANDERSEN-RANBERG, Karen; VASEGAARD, Lone; JEUNE, Bernard. Dementia is not inevitable: a population-based study of Danish centenarians. **The Journals of Gerontology Series B: Psychological Sciences and Social Sciences**, Oxford University Press, v. 56, n. 3, p152–p159, 2001.
- ANDERTON, Brian H. Changes in the ageing brain in health and disease. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, The Royal Society, v. 352, n. 1363, p. 1781–1792, 1997.
- AVANTS, Brian B; TUSTISON, Nick; SONG, Gang *et al.* Advanced normalization tools (ANTS). **Insight j**, v. 2, n. 365, p. 1–35, 2009.
- BASHYAM, Vishnu M *et al.* MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. **Brain**, Oxford University Press, v. 143, n. 7, p. 2312–2324, 2020.
- BISHOP, Christopher M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 0387310738.
- COLE, James H; FRANKE, Katja. Predicting age using neuroimaging: innovative brain ageing biomarkers. **Trends in neurosciences**, Elsevier, v. 40, n. 12, p. 681–690, 2017.
- COLE, James H; POUDEL, Rudra PK; TSAGKRASOULIS, Dimosthenis; CAAN, Matthan WA; STEVES, Claire; SPECTOR, Tim D; MONTANA, Giovanni. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. **NeuroImage**, Elsevier, v. 163, p. 115–124, 2017.
- DENG, Jia; DONG, Wei; SOCHER, Richard; LI, Li-Jia; LI, Kai; FEI-FEI, Li. Imagenet: A large-scale hierarchical image database. *In: IEEE. 2009 IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2009. P. 248–255.*
- DESIKAN, Rahul S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. **NeuroImage**, v. 31, n. 3, p. 968–980, 2006. ISSN 1053-8119.

DESPOTOVIĆ, Ivana; GOOSSENS, Bart; PHILIPS, Wilfried. MRI segmentation of the human brain: challenges, methods, and applications. **Computational and mathematical methods in medicine**, Hindawi, v. 2015, 2015.

GÉRON, Aurélien. Hands-on machine learning with scikit-learn and tensorflow: Concepts. **Tools, and Techniques to build intelligent systems**, O'Reilly Media, 2017.

GOODFELLOW, Ian J.; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge, MA, USA: MIT Press, 2016. <http://www.deeplearningbook.org>.

HAN, Juhyuk; KIM, Seo Yeong; LEE, Junhyeok; LEE, Won Hee. Brain Age Prediction: A Comparison between Machine Learning Models Using Brain Morphometric Data. **Sensors**, MDPI AG, v. 22, n. 20, p. 8077, out. 2022. ISSN 1424-8220.

HAYKIN, Simon S. **Neural networks and learning machines**. Third. Upper Saddle River, NJ: Pearson Education, 2009.

HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. Deep residual learning for image recognition. *In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. P. 770–778.

HUANG, Gao; LIU, Zhuang; VAN DER MAATEN, Laurens; WEINBERGER, Kilian Q. Densely connected convolutional networks. *In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. P. 4700–4708.

ISENSEE, Fabian; KICKINGEREDER, Philipp; WICK, Wolfgang; BENDSZUS, Martin; MAIER-HEIN, Klaus H. No new-net. *In: SPRINGER. INTERNATIONAL MICCAI Brainlesion Workshop*. [S.l.: s.n.], 2018. P. 234–244.

ISENSEE, Fabian *et al.* Automated brain extraction of multisequence MRI using artificial neural networks. **Human brain mapping**, Wiley Online Library, v. 40, n. 17, p. 4952–4964, 2019.

ISENSEE, Fabian *et al.* nnu-net: Self-adapting framework for u-net-based medical image segmentation. **arXiv preprint arXiv:1809.10486**, 2018.

JENKINSON, Mark; CHAPPELL, Michael. **Introduction to neuroimaging analysis**. [S.l.]: Oxford University Press, 2018.

- KHAN, Asifullah; SOHAIL, Anabia; ZAHOORA, Umme; QURESHI, Aqsa Saeed. A survey of the recent architectures of deep convolutional neural networks. **Artificial intelligence review**, Springer, v. 53, n. 8, p. 5455–5516, 2020.
- KINGMA, Diederik P.; BA, Jimmy. **Adam: A Method for Stochastic Optimization**. [S.l.]: arXiv, 2014. Disponível em: <https://arxiv.org/abs/1412.6980>.
- KIRILLOV, Alexander; HE, Kaiming; GIRSHICK, Ross; ROTHER, Carsten; DOLLÁR, Piotr. Panoptic segmentation. *In*: PROCEEDINGS of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019. P. 9404–9413.
- KOZIC, Dusko B; THURNHER, Majda M; BOBAN, Jasmina; SUNDGREN, Pia C. Accelerated Brain Aging: Different Diseases—Different Imaging Patterns. **Frontiers in Neurology**, Frontiers Media SA, v. 13, 2022.
- LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.
- LECUN, Yann; KAVUKCUOGLU, Koray; FARABET, Clément. Convolutional networks and applications in vision. *In*: IEEE. PROCEEDINGS of 2010 IEEE international symposium on circuits and systems. [S.l.: s.n.], 2010. P. 253–256.
- MENZE, Bjoern H. *et al.* The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). **IEEE Transactions on Medical Imaging**, v. 34, n. 10, p. 1993–2024, 2015.
- MINAEE, Shervin; BOYKOV, Yuri Y; PORIKLI, Fatih; PLAZA, Antonio J; KEHTARNAVAZ, Nasser; TERZOPOULOS, Demetri. Image segmentation using deep learning: A survey. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, 2021.
- MURPHY, Kevin P. **Machine Learning: A Probabilistic Perspective**. [S.l.]: The MIT Press, 2012. ISBN 0262018020.
- NIELSEN, Michael A. **Neural Networks and Deep Learning**. [S.l.]: Determination Press, 2018. Disponível em: <http://neuralnetworksanddeeplearning.com/>.
- O'SHEA, Keiron; NASH, Ryan. An introduction to convolutional neural networks. **arXiv preprint arXiv:1511.08458**, 2015.

OLIVEIRA, Victor HR; ANTUNES, Augusto; SOARES, Alexandre S; REYS, Arthur D; JÚNIOR, Robson Z; PEDRO, Saulo DS; SILVA, Danilo. Predição da Idade Cerebral a partir de Imagens de Ressonância Magnética utilizando Redes Neurais Convolucionais. **arXiv preprint arXiv:2112.12609**, 2021.

OLIVEIRA, Victor HR; ANTUNES, Augusto; SOARES, Alexandre S; REYS, Arthur D; JÚNIOR, Robson Z; PEDRO, Saulo DS; SILVA, Danilo. Redes Neurais Convolucionais para Desenvolvimento de Biomarcador da Idade Cerebral a partir de Imagens de Ressonância Magnética nas Ponderações T1 e FLAIR. *In: SOCIEDADE PAULISTA DE RADIOLOGIA E DIAGNÓSTICO MÉDICO. 52ª Jornada Paulista de Radiologia (JPR22). [S.l.: s.n.], 2022. P. 83–118.*

PETERS, Ruth. Ageing and the brain. **Postgraduate medical journal**, The Fellowship of Postgraduate Medicine, v. 82, n. 964, p. 84–88, 2006.

POLONI, Katia Maria; FERRARI, Ricardo José; INITIATIVE, Alzheimer's Disease Neuroimaging *et al.* A deep ensemble hippocampal CNN model for brain age estimation applied to Alzheimer's diagnosis. **Expert Systems with Applications**, Elsevier, v. 195, p. 116622, 2022.

POPESCU, Sebastian G; GLOCKER, Ben; SHARP, David J; COLE, James H. A U-net model of local brain-age. **bioRxiv**, Cold Spring Harbor Laboratory, 2021.

RAJPURKAR, Pranav *et al.* Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. **arXiv preprint arXiv:1711.05225**, 2017.

ROHLFINGA, Torsten; ZAHRA, Natalie M; SULLIVANB, Edith V; PFEFFERBAUM, Adolf. The SRI24 Multi-Channel Brain Atlas.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. *In: SPRINGER. INTERNATIONAL Conference on Medical image computing and computer-assisted intervention. [S.l.: s.n.], 2015. P. 234–241.*

SPRAWLS, Perry. **Magnetic resonance imaging: principles, methods, and techniques**. [S.l.]: Medical Physics Publishing Madison, 2000.

SRIVASTAVA, Rupesh Kumar; GREFF, Klaus; SCHMIDHUBER, Jürgen. Highway networks. **arXiv preprint arXiv:1505.00387**, 2015.

SZEGEDY, Christian; IOFFE, Sergey; VANHOUCHE, Vincent. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. **CoRR**, abs/1602.07261, 2016. arXiv: 1602.07261.

SZEGEDY, Christian; IOFFE, Sergey; VANHOUCHE, Vincent; ALEMI, Alexander A. Inception-v4, inception-resnet and the impact of residual connections on learning. *In: THIRTY-FIRST AAAI conference on artificial intelligence.* [S.l.: s.n.], 2017.

SZEGEDY, Christian; LIU, Wei; JIA, Yangqing; SERMANET, Pierre; REED, Scott; ANGUELOV, Dragomir; ERHAN, Dumitru; VANHOUCHE, Vincent; RABINOVICH, Andrew. Going deeper with convolutions. *In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition.* [S.l.: s.n.], 2015. P. 1–9.

SZEGEDY, Christian; VANHOUCHE, Vincent; IOFFE, Sergey; SHLENS, Jon; WOJNA, Zbigniew. Rethinking the inception architecture for computer vision. *In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition.* [S.l.: s.n.], 2016. P. 2818–2826.

TAN, Mingxing; LE, Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. *In: PMLR. INTERNATIONAL conference on machine learning.* [S.l.: s.n.], 2019. P. 6105–6114.

TOPOL, Eric. **Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.** 1st. USA: Basic Books, Inc., 2019. ISBN 1541644638.

WATT, Jeremy; BORHANI, Reza; KATSAGGELOS, Aggelos K. **Machine Learning Refined: Foundations, Algorithms, and Applications.** 1st. USA: Cambridge University Press, 2016. ISBN 1107123526.

WOOD, David A. *et al.* Accurate brain-age models for routine clinical MRI examinations. **NeuroImage**, v. 249, p. 118871, 2022. ISSN 1053-8119.

WRIGGLESWORTH, Jo; WARD, Phillip; HARDING, Ian H; NILAWEERA, Dinuli; WU, Zimu; WOODS, Robyn L; RYAN, Joanne. Factors associated with brain ageing-a systematic review. **BMC neurology**, Springer, v. 21, n. 1, p. 1–23, 2021.

ZHANG, Aston; LIPTON, Zachary C.; LI, Mu; SMOLA, Alexander J. Dive into Deep Learning. **arXiv preprint arXiv:2106.11342**, 2021.