CAIO ATHAYDE NEVES

APLICAÇÃO DE ALGORITMOS DE APRENDIZADO PROFUNDO NA AVALIAÇÃO PRÉ-OPERATÓRIA DE CIRURGIA OTOLÓGICA POR EXAMES DE IMAGEM

Brasília-DF

Junho de 2023

CAIO ATHAYDE NEVES

APLICAÇÃO DE ALGORITMOS DE APRENDIZADO PROFUNDO NA AVALIAÇÃO PRÉ-OPERATÓRIA DE CIRURGIA OTOLÓGICA POR EXAMES DE IMAGEM

Tese apresentada ao Programa de Pós-Graduação em Ciências Médicas da Faculdade de Medicina da Universidade de Brasília, como parte dos requisitos exigidos para a obtenção do título de Doutor.

Orientadora: Profa Dra. Iruena Moraes Kessler

Brasília-DF

Junho de 2023

CAIO ATHAYDE NEVES

APLICAÇÃO DE ALGORITMOS DE APRENDIZADO PROFUNDO NA AVALIAÇÃO PRÉ-OPERATÓRIA DE CIRURGIA OTOLÓGICA POR EXAMES DE IMAGEM

Tese apresentada ao Programa de Pós-Graduação em Ciências Médicas da Faculdade de Medicina da Universidade de Brasília, como parte dos requisitos exigidos para a obtenção do título de Doutor.

BANCA EXAMINADORA

Presidente

Profa. Dra. Iruena M. Kessler

Orientadora – PPGCM/UnB

Prof. Dr. André Luiz Lopes Sampaio

Membro – PPGCM/UnB

Prof. Dr. Henrique Fernandes de Oliveira

Membro – HRT-DF

Profa. Dra. Luciana Miwa Nita Watanabe

Membro – Ebserh

Prof. Dr. Márcio Nakanishi

Suplente – FM/UnB

Aos meus pais Selma e Ronaldo, que me deram a dádiva da vida e o farol para este caminho, e à minha esposa Karine, a maior incentivadora e parceira nesta jornada. Gratidão infinda.

AGRADECIMENTOS

Infindável lista de amigos, parceiros, inspirações e incentivadores. Agradeço imensamente a todos que participaram desta jornada, sintam-se contemplados com parte da conclusão deste trabalho.

Agradeço à minha orientadora Professora Doutora Iruena Kessler, primeiramente pela confiança depositada em mim, e finalmente pela dedicação e incentivo para a elaboração deste trabalho.

Gratidão ao Professor Nikolas Blevins, um mentor de inteligência ímpar e grande incentivador dos meus passos acadêmicos.

Ao Professor Peter Hwang, que abriu a porta para esta trilha acadêmica.

Agradeço profundamente aos ilustres membros da banca examinadora por aceitarem fazer parte deste momento crucial da minha caminhada acadêmica. Agradeço o tempo dedicado à leitura e análise da minha tese, assim como os comentários e sugestões que enriqueceram esta pesquisa.

Ao amigo Professor Doutor Henrique Fernandes de Oliveira, o “prof”, a quem sempre demonstrei minha admiração pela perseverança, dedicação e incentivo à pesquisa científica.

Ao Professor Doutor Márcio Nakanishi, pelo incentivo ao caminho acadêmico e por iluminar meu caminho à minha orientadora e aos mentores.

Ao Professor Doutor André Luiz Lopes Sampaio, exemplo de profissional, com quem tive o prazer de trabalhar e aprender sobre determinação e liderança.

A Professora Doutora Luciana Miwa Nita Watanabe, profissional de conhecida excelência e carisma, que inspira quem passa pelo seu caminho.

Aos colegas parceiros de pesquisa e de laboratório que incentivaram minha busca e contribuíram sem limites para a conclusão dos trabalhos científicos. Obrigado Nour Ibraim, Emma Tran, Trishia El Chemaly, Pooya Roozdar, Christopher Leuze, Yona Vaisbuch, George Liu, e tantos outros. Agradecimento especial a Dra Juliana Gusmão, Dra Rafaela Aquino, Dr Luiz Quaglia e Dr Marcio Lobo, que gentilmente participaram da avaliação por especialistas do Experimento 1 desta tese.

Agradeço aos meus irmãos Michaela e Átila, amigos, parceiros e fiéis companheiros. A todos a quem considero família, como Carlos e Lourdes Moraes, que apoiaram incondicionalmente esta causa.

A Rousseau, Freud e Platão, que me apoiaram, estimularam, me deram conhecimentos e tranquilidade para continuar.

A todos os meus amigos da CEOL Otorrino, pela paciência, incentivo e companheirismo nesta jornada. Todos vocês são muito especiais, e foram fundamentais para a conclusão deste trabalho.

Aos meus amigos Mário Dossi, Letícia Gentil e Rafael Carvalho, muito obrigado pela compreensão e incentivo.

À Fabiane Liano e família, que não mediram esforços para me apoiar nesta jornada.

Por último, mas não menos importante, agradeço a todos que cruzaram o meu caminho, me estimularam e plantaram sementes para o meu crescimento pessoal.

“(...) a questão em tudo e em cada coisa, “Você quer isso mais uma vez e por incontáveis vezes? ", pesaria sobre os seus atos como o maior dos pesos!”

(Nietzsche)

RESUMO

INTRODUÇÃO: A cirurgia otológica desempenha um papel crucial no tratamento de perda auditiva, infecções e tumores da base lateral do crânio. A segmentação precisa das estruturas otológicas a partir de tomografias computadorizadas (TC) pode melhorar significativamente o planejamento cirúrgico e a orientação intraoperatória. OBJETIVO: Desenvolver e validar um sistema de segmentação automatizada de estruturas-chave do osso temporal obtido de TC utilizando algoritmos de aprendizado profundo. MÉTODO: Trata-se de estudo Experimental no qual foram realizados dois ensaios. No primeiro experimento, 150 TC segmentadas manualmente foram utilizadas para construir modelos de segmentação automatizada empregando redes neurais convolucionais (CNN). A análise objetiva dos modelos de segmentação da orelha interna, nervo facial, ossículos e seio sigmoide incluíram o coeficiente de Dice, velocidade de segmentação e distância de Hausdorff média. No segundo protótipo, um moderno algoritmo de aprendizado profundo (SwinUNETR) foi usado para construir um modelo de previsão para segmentação rápida de nove estruturas-chave do osso temporal em 325 TC clínicas. A avaliação objetiva incluiu Dice, precisão balanceada, distâncias de Hausdorff e tempo de processamento. RESULTADOS: No primeiro experimento, os modelos obtiveram coeficientes de Dice de 0,91, 0,85, 0,75 e 0,86 para as respectivas estruturas, e o tempo de segmentação médio foi de 2,7 segundos por estrutura. O segundo modelo alcançou coeficiente de Dice médio de 0,87 para todas as estruturas, precisão balanceada média de 0,94, Distância de Hausdorff média de 0,79mm e tempo médio de processamento de 9,1 segundos por TC. CONCLUSÃO: Neste estudo, a aplicação de algoritmos de aprendizado profundo para a segmentação automatizada de estruturas do osso temporal em TC permitiu a construção de modelos com elevada precisão de acordo com a análise objetiva recomendada atualmente. Os resultados obtidos demonstram o potencial do método para melhorar a avaliação pré-operatória e a orientação intraoperatória em cirurgia otológica.

Palavras-chave: Cirurgia Otológica; Tomografia Computadorizada; Segmentação de Imagem; Redes Neurais Convolucionais; Aprendizado Profundo (Deep Learning)

ABSTRACT

INTRODUCTION: Otological surgery plays a crucial role in the treatment of hearing loss, infections, and tumors of the lateral skull base. Precise segmentation of otological structures from computed tomography (CT) can significantly improve surgical planning and intraoperative guidance. OBJECTIVE: To develop and validate an automated segmentation system of key temporal bone structures obtained from CT using deep learning algorithms. METHOD: This is an experimental study in which two trials were performed. In the first experiment, 150 manually segmented CTs were used to construct automated segmentation models using Convolutional Neural Networks (CNN). Objective analysis of the inner ear, facial nerve, ossicles, and sigmoid sinus segmentation models included the Dice coefficient, segmentation speed, and average Hausdorff's distance. In the second prototype, a modern deep learning algorithm (SwinUNETR) was used to construct a prediction model for quick segmentation of nine key temporal bone structures in 325 clinical CT. The objective evaluation included Dice, balanced accuracy, Hausdorff distances, and processing time. RESULTS: In the first experiment, the models obtained Dice coefficients of 0.91, 0.85, 0.75, and 0.86 for the respective structures, and the average segmentation time was 2.7 seconds per structure. The second model achieved an average Dice coefficient of 0.87 for all structures, an average balanced accuracy of 0.94, an average Hausdorff distance of 0.79mm, and an average processing time of 9.1 seconds per CT. CONCLUSION: In this study, the application of deep learning algorithms for the automated segmentation of temporal bone structures on CT scans allowed the construction of models with high accuracy according to the currently recommended objective analysis. The results obtained demonstrate the potential of the method to improve preoperative evaluation and intraoperative guidance in otologic surgery.

Keywords: Otologic Surgery; Computed Tomography; Image Segmentation; Convolutional Neural Networks; Deep Learning

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

ACI	Artéria carótida interna
AH-Net	*Anisotropic hybrid Network*
AHD	Distância de Hausdorff média
AP	Aprendizado profundo
CAE	Conduto auditivo externo
CAI	Conduto auditivo interno
CNN	*Convolutional neural network*
CO	Cápsula ótica
CSS	Canal semicircular superior
CUDA	*Compute Unified Device Architecture*[1]
DP	Desvio padrão
Exp.	Experimento
FN	Falso-Negativo
FP	Falso-Positivo
HD	Distância de Hausdorff
HD95	Percentil 95º da Distância de Hausdorff
HU	*Hounsfield Units* (Unidades Hounsfield)
LLM	*Large Language Model* (Grande Modelo de Linguagem)
MONAI	*Medical Open Network for Artificial Intelligence* [2]
NF	Nervo facial
NCoT	Nervo da corda do tímpano
OI	Orelha interna

[1] Estrutura de programação para computação paralela desenvolvida pela empresa NVIDIA

[2] Plataforma de desenvolvimento de projetos de inteligência artificial para medicina

Oss	Ossículos
PB	Precisão balanceada
RAM	*Random access memory* [3]
ResNet	*Residual Network*
RM	Realidade mista
RV	Realidade virtual
SS	Seio sigmoide
SV	Similaridade volumétrica
VA	Volume automatizado
VM	Volume manual
VN	Verdadeiro-negativo
VP	Verdadeiro-positivo
VR	Similaridade volumétrica
U-Net	*U-shaped network*

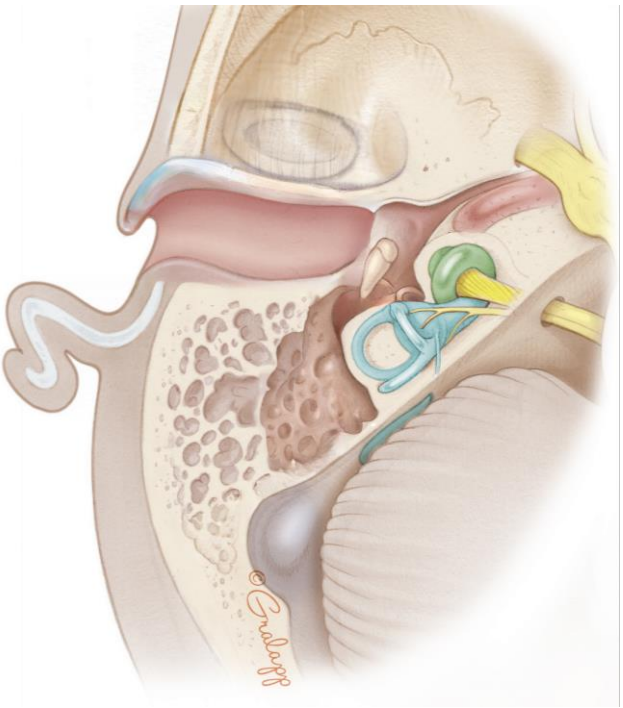
[3] Memória de curto período, componente de computadores para armazenamento de informações de uso imediato.

SUMÁRIO

1 INTRODUÇÃO

A cirurgia otológica é uma disciplina altamente especializada, demandando do cirurgião um conhecimento aprofundado em anatomias radiológica e cirúrgica, bem como uma percepção tridimensional intraoperatória. O pequeno e delicado campo cirúrgico, que inclui procedimentos complexos como implante coclear, timpanomastoidectomia e reparação de deiscência do canal semicircular superior, apresenta desafios significativos devido às delicadas inter-relações entre os ossos e as estruturas neurovasculares vitais no osso temporal. Isso requer um tempo considerável de prática e experiência. Além disso, a proximidade dessas estruturas vitais entre si e com os espaços críticos ao redor da base do crânio lateral torna a cirurgia nessa região particularmente exigente, necessitando de profundo entendimento da anatomia e extenso treinamento cirúrgico. Um esquema da delicada anatomia do osso temporal está apresentado na Figura 1.

Figura 1 - Vista axial da anatomia do osso temporal



Fonte: Atlas de Neurotologia e cirurgia de base do crânio.[4]

[4] Disponível em https://skullbasesurgeryatlas.stanford.edu, acessado em 5 de maio de 2023.

Cirurgiões desta área devem ter um conhecimento abrangente da anatomia do osso temporal e suas estruturas circundantes, bem como treinamento extensivo que requerem anos de dedicação. (ANDERSEN *et al.*, 2015, 2016; ERICSSON, 2008). Devem também ter ampla experiência com técnicas cirúrgicas minimamente invasivas e uma compreensão completa dos riscos e benefícios de diferentes abordagens cirúrgicas. A curva de aprendizado de cirurgias otológicas é longa, e assim como outras cirurgias minimamente invasivas, demandam treinamento extensivo tutorado e em sessões distribuídas ao longo do tempo. (ANDERSEN *et al.*, 2015; CHEN *et al.*, 2023; WIET; SØRENSEN; ANDERSEN, 2017).

O padrão ouro do treinamento de cirurgias da orelha média e interna é a dissecção de osso temporal coletado de cadáveres. (FRITHIOFF; SØRENSEN; ANDERSEN, 2018). A dificuldade de acesso a laboratórios de dissecção do osso temporal e a responsabilidade de não expor os pacientes a riscos se operado por aprendizes não treinados impõe a necessidade da incorporação de técnicas de simulação de cirurgias otológicas. (FRENDØ *et al.*, 2021; FRITHIOFF *et al.*, 2023). Evidências apontam que a carga cognitiva é menor quando as sessões de treinamento são distribuídas em vários dias quando comparado com treinamento intensivo. (ANDERSEN *et al.*, 2016). Neste contexto, o desenvolvimento de um currículo de simulação cirúrgica pode acelerar o aprendizado cirúrgico e potencialmente reduzir o risco para os pacientes quando operados por aprendizes. (DE OLIVEIRA *et al.*, 2022, 2017).

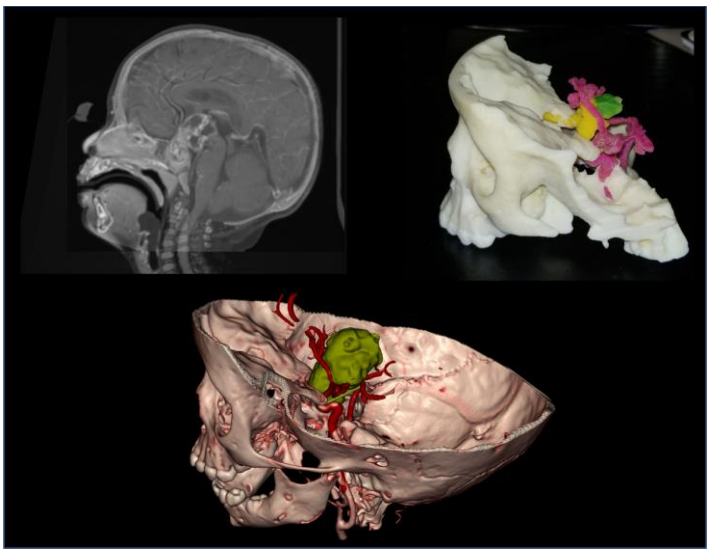
A fim de superar os obstáculos dos riscos de exposição dos pacientes na longa curva de aprendizado, do planejamento cirúrgico individualizado e do treinamento constante, diversas tecnologias foram incorporadas ao ensino e prática médica, como a simulação cirúrgica, uso de sistemas de realidade estendida e de navegação cirúrgica. (NEVES *et al.*, 2020; WON *et al.*, 2018, 2019). Estes sistemas requerem o uso de exames radiológicos, e se beneficiariam de informações adicionais extraídas destes exames, como a identificação e localização prévias de estruturas de interesse e renderização [5] destas estruturas sobrepostas às imagens dos vídeos ou endoscópios.

[5] Renderização é o processo de computação gráfica que constrói a representação visual incluindo texturas, reflexos e sombras a partir de um modelo 3D, que é então projetado em uma tela ou em outro aparelho de visualização.

Neste contexto, a tomografia computadorizada (TC) desempenha um papel central tanto no diagnóstico quanto no planejamento do tratamento de afecções do osso temporal. Embora as TC sejam imagens volumétricas (3D), elas são tradicionalmente apresentadas como uma série de imagens 2D multiplanares, exigindo um processamento mental experiente para transformar essas imagens na representação 3D necessária à tradução da anatomia real.

Um método para a geração rápida e precisa de modelos 3D de alta fidelidade específicos do paciente para planejamento pré-operatório (LOCKETZ *et al.*, 2017) e navegação intraoperatória (BARBER *et al.*, 2018; NEVES *et al.*, 2020) pode oferecer uma variedade de benefícios potenciais tanto para o paciente quanto para o cirurgião, que se beneficiaria da incorporação destes modelos em tempo real no ato cirúrgico, promovendo a ampliação da percepção visual do cirurgião utilizando técnicas de realidade estendida. A figura 2 demonstra o uso de modelagem 3D para a construção de modelos reais e virtuais no planejamento pré-operatório personalizado em um caso de tumor de base do crânio em paciente pediátrico.

Figura 2- Uso de modelagem 3D para o planejamento cirúrgico de craniofaringioma



Fonte: Elaborado pelo autor (2019) [6]

[6] Superior esquerdo: TC e RM T1 fusionadas mostrando grande tumor selar e supraselar. Superior direito: modelo impresso em 3D em tamanho real. Inferior: Simulação de realidade virtual mostrando as relações do tumor com estruturas neurovasculares (vista esquerda).

1.1. REVISÃO BIBLIOGRÁFICA

1.1.1. Realidade estendida

A realidade estendida (RE) é um campo de desenvolvimento tecnológico que incorpora um conjunto de técnicas de renderização de cenários e objetos e a combinação destes objetos virtuais com o mundo real. Neste conjunto estão incluídos a realidade virtual (RV), a realidade aumentada (RA) e a realidade mista (RM). (YUAN *et al.*, 2023). Existe um debate acerca da diferença entre RA e RM, e podem ser consideradas formas da mesma tecnologia. .(SPEICHER; HALL; NEBELING, 2019). Estas técnicas são descritas como tecnologias de imersão, nas quais o usuário estará envolto de informações além do mundo real.

1.1.1.1. Realidade virtual

A realidade virtual é a tecnologia mais conhecida do espectro da RE, na qual o usuário é apresentado a um novo cenário, completamente ausente do mundo real. Na medicina, técnicas de realidade virtual são amplamente utilizadas no desenvolvimento de sistemas de simulação cirúrgica, que apresentam cenários de anatomia normal ou alterada para o treinamento em ambiente seguro e tutorado. (PIROMCHAI *et al.*, 2015; SHAO *et al.*, 2020).

Especialmente concernente a esta tese, técnicas de segmentação anatômica automatizada a partir de exames de imagem podem ampliar o uso de simulação cirúrgica para a simulação paciente-específica. Nesta modalidade de simulação, o modelo virtual é a representação da anatomia de um determinado paciente a ser estudado, por meio do processamento dos exames de imagem daquele indivíduo. A simulação cirúrgica paciente-específica é fundamental no treinamento e planejamento cirúrgico de procedimentos avançados e incomuns. (WON *et al.*, 2019). Idealmente, um paciente a ser operado, teria as estruturas anatômicas relevantes segmentadas de maneira automatizada. Então, com mínimo pré-processamento manual, o cirurgião poderia utilizar o simulador para ensaiar o procedimento em um ambiente seguro. (ANDERSEN *et al.*, 2021; LAI *et al.*, 2022). Como a segmentação anatômica

automatizada das diversas regiões anatômicas ainda é objeto de estudo e desenvolvimento, os estudos que envolvem simulação paciente-específica têm nesta barreira (segmentação automatizada) a sua principal restrição, apesar de já demonstrarem resultados razoáveis de validação de usabilidade. (LAI *et al.*, 2022).

1.1.1.2. Realidade aumentada e realidade mista

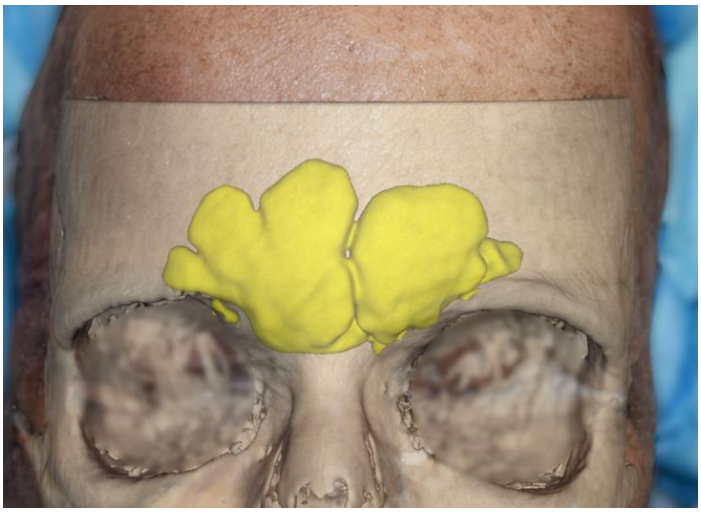
A realidade aumentada é uma técnica que adiciona elementos virtuais em um ambiente real, tornando a experiência imersiva mais rica. Esses elementos virtuais podem ser informações ou objetos que são criados digitalmente e renderizados na visão do mundo real.

Na medicina e na cirurgia, o emprego de dispositivos e soluções de realidade aumentada buscam ampliar a visão do usuário para informações ocultas, como a anatomia subjacente ao plano cirúrgico e facilitar a decisão intraoperatória. (JAMES *et al.*, 2022; LAI *et al.*, 2022; SPEICHER; HALL; NEBELING, 2019; WINNE *et al.*, 2011). Experimentalmente, estudos abordam a projeção de estruturas da face e da base do crânio para auxiliar o cirurgião nas osteotomias (LEUZE *et al.*, 2021) e clinicamente alguns dispositivos já permitem o planejamento intraoperatório da inserção de parafusos em cirurgias ortopédicas. (BUTLER *et al.*, 2023).

Paralelamente, a navegação cirúrgica é uma área de foco na aplicação de técnicas de realidade aumentada. A incorporação de informações anatômicas tridimensionais nos navegadores cirúrgicos pode ampliar a percepção anatômica e aprimorar a segurança dos procedimentos, uma vez que o cirurgião pode ser alertado quando próximo de uma estrutura de interesse. (BROCKMEYER; WIECHENS, 2023; HADDAD; AGHI; BUTOWSKI, 2022; PRISMAN *et al.*, 2011). Particularmente, a segmentação anatômica automatizada é de extrema importância na navegação cirúrgica aprimorada, e sistemas com estas tecnologias podem identificar estruturas de interesse e renderizá-las em aparelhos de navegação cirúrgica que ampliam a visão do cirurgião. (DALY *et al.*, 2010; LEONARD *et al.*, 2016; NEVES *et al.*, 2021; WINNE *et al.*, 2011).

A figura 3 apresenta o uso de técnica de segmentação anatômica para utilização de realidade aumentada para acessos externos ao seio frontal.

Figura 3 - Uso de realidade virtual para acessos externos ao seio frontal



Fonte: (Neves et al. 2020)

1.1.2. Tomografia computadorizada

A tomografia computadorizada é fundamental para fornecer aos cirurgiões de ouvido e base do crânio informações sobre a anatomia exclusiva de um paciente para o planejamento pré-operatório. Porém, a identificação das estruturas-chave e particulares alterações delas por variação interpessoal ou patológica pode ser difícil para radiologistas e cirurgiões, pela complexidade e pela pequena dimensão espacial destas estruturas. Por isto, entender sua orientação e geometria é essencial para procedimentos otológicos bem-sucedidos, como na cirurgia de implante coclear ou ressecção de tumores. (GARE *et al.*, 2020).

Além disso, embora os conjuntos de dados de TC sejam inerentemente volumétricos, os cirurgiões rotineiramente os analisam como representações multiplanares bidimensionais (2D). Durante a avaliação pré e intraoperatória destes dados, uma nova tradução mental das imagens da tomografia para a representação tridimensional é fundamental para o entendimento da anatomia cirúrgica daquele paciente.

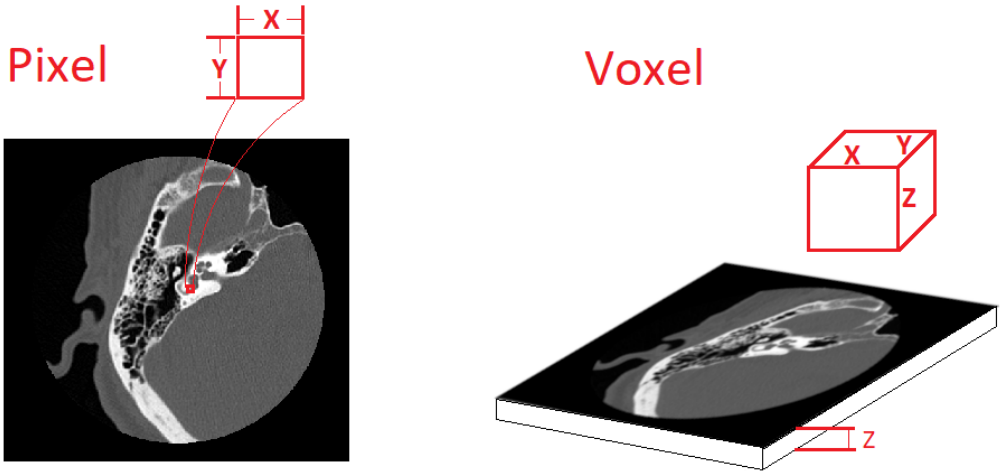
Assim como a aquisição da fluência em linguagens naturais ou de programação, a proficiência na avaliação radiológica pré-operatória requer dedicação intensiva e extensiva no entendimento sobretudo das correlações clínico-radiológicas e anatômico-radiológicas, e em parte nos detalhes técnicos das diferentes modalidades.

Especificamente, alguns detalhes técnicos da TC merecem destaque para a compreensão dos experimentos conduzidos nesta pesquisa, e serão detalhados a seguir.

1.1.2.1. Voxel e espaçamento

Na TC, voxel se refere à unidade tridimensional da imagem gerada. Assim como o pixel é a unidade de uma imagem digital bidimensional, o voxel (contração de “volume” e “pixel”) contém a informação sobre a menor divisão do volume tomográfico (Figura 4). De maneira simplificada, o voxel poderia ser descrito como um cubo no espaço tridimensional da tomografia, e a sua intensidade em tons de cinza representaria a quantidade de radiação absorvida pelo tecido que atravessou.

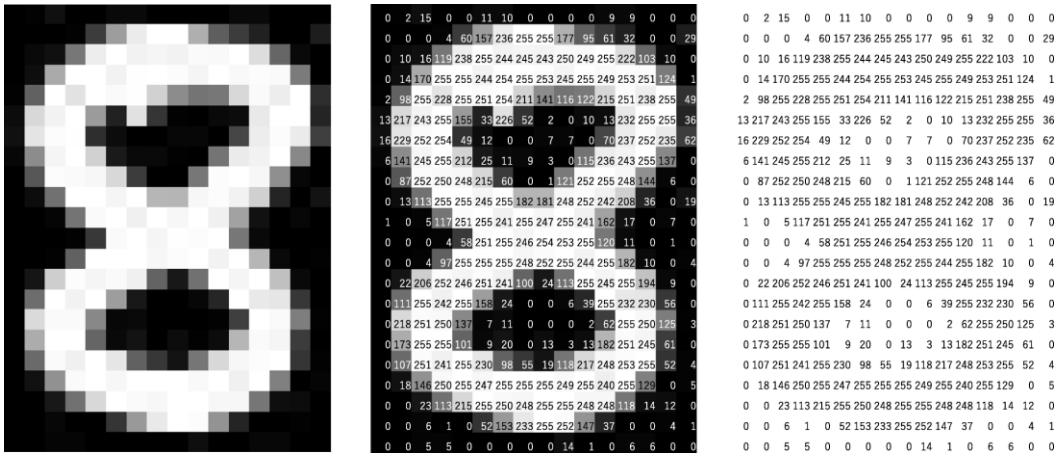
Figura 4 - Representação de pixels e voxels em tomografias



Fonte: Elaborado pelo autor (2023)

Empilhados e enfileirados adequadamente, o conjunto de voxels constitui um volume maior que é a representação digital da região escaneada pelo tomógrafo. Esta organização dos voxels, por sua vez, pode ser representada por uma matriz matemática que descreve a exata posição do voxel no volume total, assim como sua intensidade, que em conjuntos, representam as diferentes estruturas anatômicas. A figura 5 exemplifica este conceito utilizando o número 8, representado visualmente como pixels e como uma matriz matemática das intensidades.

Figura 5 - Representação de imagem por intensidade dos pixels



Fonte: (PATIL; RANE, 2021)

Aprofundando a descrição da representação digital do objeto escaneado na TC, é importante ressalvar que o formato dos voxels não é sempre cúbico. De fato, nos aparelhos de TC clínicos, os protocolos de aquisição e reformatações pré-determinados geram em grande maioria voxels não-cúbicos, mas de formatos paralelepípedos, em que cada dimensão pode representar uma escala de distância diferente no objeto escaneado.

Por padronização, considerando o espaço ortogonal XYZ, o eixo x representa na TC uma posição em relação ao plano sagital original, ou distância latero-lateral em relação à linha média, ou esquerda-direita em termos anatômicos. O eixo y, por sua vez, representa uma posição em relação ao plano coronal original, ou frontal, e representa posições anteriores ou posteriores. Finalmente, o eixo Z representa posições em relação ao plano axial original, e identifica pontos superiores ou inferiores na imagem. Tanto a imagem completa, eventualmente referida como volume, quanto os voxels, possuem as dimensões x, y, e z que podem variar com o aparelho e protocolo de aquisição.

Em um volume considerado isotrópico, os voxels possuem formato cúbico, e consequentemente cada dimensão do voxel representa distâncias similares no objeto escaneado. Alguns algoritmos de processamento de imagem requerem que o volume de entrada seja isotrópico, e muitos outros geram melhores resultados quando são utilizados volumes com esta característica. Contrariamente, volumes anisotrópicos se

referem a imagens em que as dimensões do voxels representam distâncias diferentes. Nos volumes anisotrópicos, normalmente o eixo z é penalizado em comparação com os outros eixos, e a resolução na dimensão supero-inferior é menor quando comparada com as resoluções latero-lateral e antero-posterior. Isto é de grande importância na leitura das imagens tridimensionais, uma vez que nestes volumes anisotrópicos, a incerteza no eixo z pode acarretar redução da precisão na identificação de estruturas anatômicas.

A escala em que cada dimensão do voxel de um volume se correlaciona com as dimensões espaciais do objeto escaneado no mundo real é chamada de espaçamento, que convencionalmente é dada em milímetros(mm). O espaçamento também pode ser interpretado como a resolução do volume escaneado, pois quanto menor o valor do espaçamento, mais detalhada é a imagem. O espaçamento é descrito como um grupo de três valores decimais (em inglês, *tuple*), que se referem às escalas de cada dimensão do voxel ao mundo real nos eixos x, y e z. Neste sentido, um exemplo de TC com espaçamento de (0,3 , 0,3 , 0,625) indica que a medida de cada voxel nos eixos x e y equivale a 0,3 mm no objeto escaneado, e no eixo z equivale a 0,625 mm.

Neste exemplo fica claro como a anisotropia do volume pode reduzir a quantidade de informação vertical em cada voxel, pois neste caso em cada voxel, que só pode ter um valor de intensidade referente à absorção de radiação do tecido, está contida a informação referente a 0,625mm do corpo escaneado, em comparação com as outras dimensões x e y. O exemplo em questão pode não levantar preocupação quanto à perda de informação pois, mesmo a dimensão z tendo aproximadamente a metade da resolução das outras dimensões, a sua escala ainda é submilimétrica. Entretanto, em alguns protocolos clínicos podemos encontrar espaçamentos tão grandes quanto 2mm, o que pode prejudicar sobremaneira a delineação de pequenas estruturas como os canais semicirculares ou o nervo facial.

O processamento de ajuste do tamanho dos voxels e dos espaçamentos para adequar as imagens radiológicas a diferentes aplicações é denominada redimensionamento, que é realizado por meio de técnicas de interpolação da imagem.

1.1.2.2. Redimensionamento e Interpolação

Para a mudança de resolução e consequente de espaçamento de um determinado volume, algoritmos de redimensionamento (*resampling* em inglês) podem ser utilizados com diversas técnicas de interpolação que buscam o incremento ou redução de voxels intermediários, comumente utilizados para gerar imagens isotrópicas a partir de anisotrópicas. A escolha da técnica de interpolação depende do tipo de imagem e do objetivo da análise. Algumas técnicas são mais adequadas para imagens com bordas complexas ou contornos irregulares, enquanto outras são mais eficazes para imagens com texturas e detalhes finos. O tempo e a quantidade de processamento disponível e o tamanho da imagem também influenciam na escolha.

A técnica de interpolação linear é de pequena demanda computacional, em que voxels intermediários são criados por meio da estimativa de um plano que passa pelos pontos discretos em questão e do cálculo de valores intermediários contidos neste plano de acordo com a nova resolução determinada. É simples e rápida, mas pode causar efeitos de *aliasing*[7], já que a imagem resultante pode parecer serrilhada. (GONZALEZ; WOODS, 2018). No entanto, a interpolação linear pode produzir resultados com menor qualidade em relação a outras técnicas de interpolação, especialmente em situações em que a imagem original possui muitos detalhes finos. Nessas situações, técnicas de interpolação mais avançadas podem ser mais adequadas para preservar a qualidade da imagem ao aumentar seu tamanho.

Por outro lado, a interpolação por *splines* utiliza funções polinomiais suaves para criar curvas que se ajustam a grupos de pontos da imagem. Ela é muito precisa e pode produzir imagens suaves e naturais, mas pode ser computacionalmente cara, especialmente em volumes de grandes dimensões. (GONZALEZ; WOODS, 2018). No pré-processamento de TC de ossos temporais, a técnica de interpolação por *splines* pode trazer excelentes resultados, e o tempo de processamento pode ser menor se o volume original for reduzido (cortado) para incluir somente a região de interesse.

Estes processamentos visam a padronização das características dos volumes radiológicos com o objetivo de produzir imagens de resolução ótima que permita a boa identificação das estruturas anatômicas. À anotação para a delineação de objetos em uma imagem denominamos segmentação semântica, muitas vezes chamada

[7] *Aliasing*: Efeito de sobreposição de imagens que causa distorção na imagem final, podendo apresentar aspecto serrilhado.

somente de segmentação, que é uma das tarefas da Visão de Computador (do inglês *Computer vision*), um campo da Inteligência Artificial (IA) que teoriza e desenvolve tecnologia para a obtenção de informações relevantes do processamento de imagens por computadores. (LECUN; BENGIO; HINTON, 2015).

1.1.3. Inteligência artificial

A inteligência artificial (IA) é um vasto campo de estudo que incorpora conhecimentos e técnicas de múltiplas disciplinas, incluindo informática, matemática, estatística, neurociência e psicologia. O objetivo principal da IA é desenvolver sistemas computacionais que possam aprender com dados e tomar decisões com base neste aprendizado, e para isto possui algumas vertentes de pesquisa e desenvolvimento, entre elas o aprendizado de máquina, o processamento de linguagem natural, a representação e raciocínio do conhecimento e a visão de computador. (MOAWAD *et al.*, 2022)

O aprendizado de máquina (*machine learning*) é por muitas vezes confundido com a própria IA, mas na verdade é um ramo de pesquisa da IA. Este ramo se dedica a construir algoritmos e modelos que podem aprender com novos dados para a melhora do seu próprio desempenho ao longo do tempo. O aprendizado de máquina permite que os sistemas reconheçam padrões e façam previsões com base em grande quantidade de dados, de maneira similar, mas em grande escala não possível para os humanos.

O aprendizado de máquina engloba uma grande quantidade de sistemas e algoritmos que podem ser treinados para o reconhecimento de padrões e aprendam com dados novos. As aplicações são inúmeras, e na medicina esta abordagem constitui um enorme campo de pesquisa e de desenvolvimento para a prevenção, diagnóstico e tratamento das doenças. Assim como as aplicações, os próprios algoritmos são objeto de desenvolvimento. Recentemente os grandes modelos de linguagem (LLM do inglês *Large Language Model)* mostraram o avanço deste tipo de sistema.

Os LLM têm em comum o uso dos “transformers”, uma arquitetura de aprendizado profundo que tem como premissa a atenção (VASWANI *et al.*, 2017), e permite que o algoritmo aprenda qual é o próximo resultado de uma sentença por

exemplo. Um algoritmo com esta tecnologia foi utilizado no segundo experimento deste trabalho, em complemento ao experimento inicial que utilizou redes convolucionais tradicionais, que era o estado da arte no momento daquele estudo. Devido ao seu grande sucesso, os *transformers* têm sido utilizados em outras áreas da inteligência artificial com resultados promissores, como no campo da visão de computador. (DOSOVITSKIY *et al.*, 2020; LIU *et al.*, 2021).

A visão de computador (*Computer Vision*, do inglês) é um importante campo de pesquisa e desenvolvimento da IA. (LECUN; BENGIO; HINTON, 2015). É o campo que se destina ao aprimoramento de técnicas e algoritmos para a identificação visual pelas máquinas, e tem grande importância em sistema de segurança, automação de meios de transporte e em especial, na área médica. (ESTEVA *et al.*, 2019, 2021).

O interesse pelas técnicas de visão de computador impulsionou diversas novas linhas de pesquisa nos últimos anos, e a produção científica abordou desde pesquisa em áreas básicas (AHMAD *et al.*, 2023; XIAO *et al.*, 2023) até aplicações clínicas para uso a beira do leito. (GULSHAN *et al.*, 2016).

As implementações de técnicas de visão de computador na medicina são amplas e se estendem nas diferentes tarefas deste campo, listados a seguir:

- Classificação
- Localização
- Detecção de objetos
- Segmentação semântica

Classificação é uma tarefa de processamento de imagens que permite a discriminação de uma imagem analisada em uma determinada classe de acordo com as características desta imagem. (KOTSIANTIS, 2007). É uma tarefa que identifica informação semântica a partir de imagem, porém é limitada a dar uma resposta de acordo com a principal informação da imagem. Um exemplo na área médica é dos estudos pioneiros de classificação de imagens digitais de fundo de olho para avaliação de retinopatia diabética (GULSHAN *et al.*, 2016) e de radiografias de tórax CheXNet. (RAJPURKAR *et al.*, 2017).

Diversos estudos descrevem soluções para a classificação de radiografias em classes incluindo pneumonia (covid em especial) (NISHIO *et al.*, 2020), derrames pleurais, infiltrados, tumores entre outros. Na prática médica, estas técnicas podem

permitir o rastreamento de doenças em larga escala em locais com carência de especialistas e auxiliar os profissionais de saúde tanto nestas áreas, mas também em grandes centros urbanos.

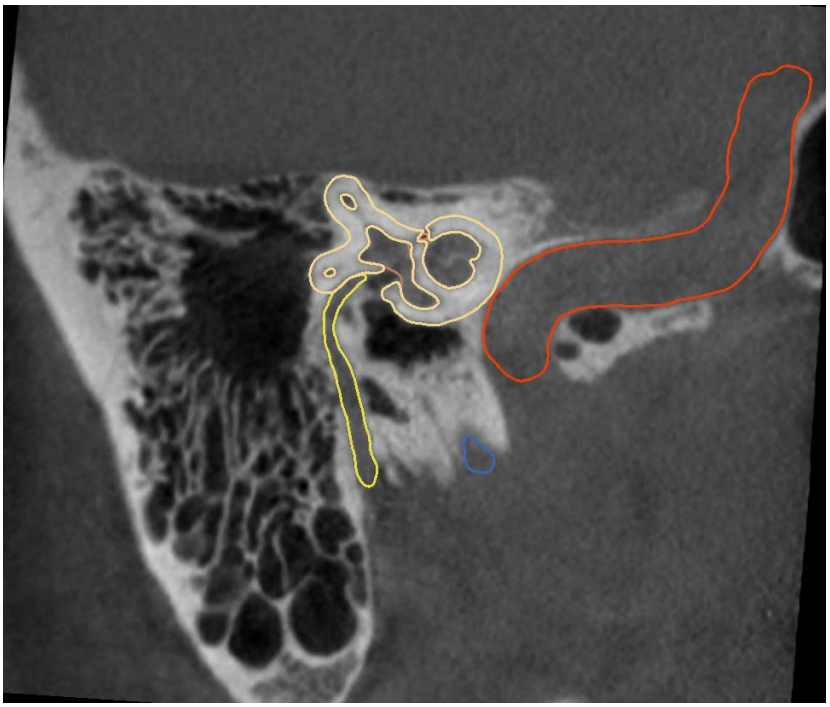
Complementando a tarefa de classificação, e muitas vezes a ela associada, a tarefa de localização reúne técnicas de identificação espacial das características que discriminam a classe da imagem. Utilizando o exemplo de derrame pleural em radiografias de tórax, o sistema, além de classificar a imagem como tal, é também capaz de apontar a região que apresenta características da efusão. A localização é de grande importância para sistemas de radiologia auxiliada por computador, em que sistemas classificam a imagem e sugerem ao radiologista a área de maior probabilidade do achado que sugere aquela classificação. (MOAWAD *et al.*, 2022).

A capacidade de localizar mais de um objeto em uma aplicação de visão de computador é denominada detecção de objetos. A implementação destas técnicas em aplicações diversas promove a identificação e localização de mais de uma classe em uma única imagem. Na radiologia, técnicas de detecção de objetos podem ser utilizadas para a identificação de achados anormais, como consolidações e também sinais de tumores pulmonares. (JAVED *et al.*, 2023; RAJPURKAR *et al.*, 2017).

A segmentação semântica é um processo de análise de imagens que visa identificar e delinear objetos e regiões de similares significados ou interpretação semelhantes. A técnica é particularmente importante no processamento de imagens médicas pois é utilizada para delineação de estruturas anatômicas que podem ter implicância clínica. A segmentação anatômica abrange diversas áreas de interesse, incluindo o diagnóstico e controle de alterações nas estruturas anatômicas (como alterações patológicas causadas às mesmas), simulação de cirurgias, aplicação de realidade estendida (realidade aumentada e realidade virtual) nas etapas pré e intraoperatórias, e navegação cirúrgica, entre outros.

A figura 6 apresenta imagem de tomografia de osso temporal em reformatação oblíqua na qual estruturas anatômicas como o nervo facial, a artéria carótida interna e a cápsula ótica estão delineadas.

Figura 6 - Segmentação anatômica de tomografia de osso temporal



Fonte: Elaborado pelo autor (2023)

A implementação de novas tecnologias para o aprimoramento do pré e intraoperatório buscam aumentar a segurança e a eficiência dos procedimentos. No entanto, esses esforços são limitados pela etapa trabalhosa de segmentação manual de dados de imagem por especialistas altamente treinados. (CHAN *et al.*, 2016; WON *et al.*, 2019).

1.1.3.1. Segmentação anatômica e revisão da literatura pertinente

A segmentação automatizada de estruturas anatômicas do ouvido a partir de exames de imagem compreende um campo de pesquisa de parte do planejamento pré-operatório de cirurgias otológicas. Diversos grupos têm se dedicado ao desenvolvimento de soluções para este problema. (FAUSER *et al.*, 2019; GARE *et al.*, 2020; NAKASHIMA *et al.*, 1993; NEVES, C. A. *et al.*, 2021; NIKAN *et al.*, 2021; NOBLE *et al.*, 2008; WANG *et al.*, 2021). Diante da grande variabilidade anatômica e patológica do osso temporal, um modelo de segmentação automatizada deve ser robusto e generalizável a fim de ser aplicado clinicamente.

Independentemente do método de desenvolvimento de um determinado modelo, estes são baseados em exemplos previamente apresentados. Isto é, os modelos são desenvolvidos por métodos supervisionados, quando os exemplos previamente anotados por humanos são considerados o padrão-ouro da segmentação. Neste contexto, a quantidade de exemplos utilizados no desenvolvimento do modelo tem um papel crucial na sua generalização, pois mais exemplos de variações anatômicas e patológicas podem fazer parte da base do modelo, que pode ter maior capacidade de identificar padrões semelhantes quando da sua implementação.

Tanto a dificuldade da segmentação anatômica de estruturas do osso temporal em exames de imagem, quanto a restrição da capacidade dos métodos empregados até então, limitaram o desenvolvimento de modelos robustos. A figura 7 apresenta um código QR que permite acesso a um vídeo que demonstra a segmentação manual da orelha interna em uma TC.

Figura 7 - Código QR com link para video exemplo de segmentação manual da orelha interna

Fonte: elaborado pelo autor

A partir do avanço da disponibilidade de recursos computacionais, técnicas de desenvolvimento de atlas de segmentação foram empregados para a construção de modelos de segmentação. Na técnica de modelos por atlas, uma ou mais tomografias com estruturas segmentadas manualmente são utilizadas como o padrão (ou “atlas”) para a tarefa. A matriz matemática obtida da distância entre a nova tomografia em análise e o atlas é utilizada para deformar as segmentações do atlas a fim de se adequar ao novo exame. Portanto, neste caso, a segmentação automatizada é fruto

da deformação da segmentação manual pela matriz de deformação advinda da diferença entre a imagem analisada e a imagem considerada atlas.

Os primeiros estudos na área incluem o de Nakashima e colaboradores (NAKASHIMA *et al.*, 1993), que realizaram uma análise computacional de espécimes de histopatologia do osso temporal humano. Em uma série de publicações(NOBLE *et al.*, 2008, 2009, 2011), Noble e sua equipe utilizaram métodos baseados em atlas e outras soluções personalizadas para a detecção automatizada de elementos como o nervo facial, ossículos e a anatomia intracoclear. Mais recentemente, Powell et al. (POWELL *et al.*, 2019) e Gare et al. (GARE *et al.*, 2020) demonstraram boa correlação entre a segmentação automatizada com base em atlas do osso temporal e a referência verdadeira. Hudson et al. (HUDSON *et al.*, 2020) usou modelos baseados em atlas alinhados a exames de micro-CT, requerendo a inserção manual de pontos do trajeto do nervo facial para segmentar esta estrutura em tomografia de espécimes cadavéricos. Ding et al. (DING *et al.*, 2022) publicou um estudo no qual alcançou Dice de 0,77 para a Carótida interna, 0,84 para a cápsula ótica, 0,68 para o conduto auditivo interno, 0,82 para o conduto auditivo externo, 0,62 para o seio sigmoide e 0,57 para o nervo facial.

Entretanto, estes estudos utilizam amostras pequenas, que limitam a generalização dos resultados, e várias destas soluções requerem ação do usuário em forma de marcação de pontos sobre o trajeto do nervo facial ou outros métodos de registro, sendo sujeitos a variação dependente do usuário e limitados na sua escalabilidade.

Técnicas que utilizam algoritmos de aprendizado profundo estão sendo crescentemente utilizadas na última década como abordagem para diversas tarefas de análise de imagens, pois os algoritmos utilizados são capazes de identificar padrões sutis, muitas vezes pouco perceptíveis ao olhar superficial.(MENG; TIAN; BU, 2020)

Fauser et al. (FAUSER *et al.*, 2019) desenvolveram um método híbrido utilizando uma rede de aprendizado profundo associado à técnica de ajustamento de forma. Este estudo é pioneiro na utilização de redes de convoluções para a segmentação anatômica do osso temporal em TC. Os resultados foram encorajadores, pois utilizando apenas 24 tomografias manualmente segmentadas, o sistema produziu respostas razoavelmente boas.

Heutink et al. (HEUTINK *et al.*, 2020) descreveram uma solução automatizada para a segmentação da cóclea com ótimo resultado (Dice 0,9), porém utilizando uma rede neuronal bidimensional, que analisa cada fatia da tomografia separadamente, o que pode limitar a precisão do modelo, além de aumentar o tempo de processamento. Naquele estudo, os autores descrevem que cada tomografia foi processada em 10 minutos. Outra limitação daquele estudo foi a construção do modelo a partir de tomografias de alta resolução, pouco disponíveis clinicamente. Além disso, as tomografias foram cortadas manualmente em um menor volume com centro na orelha interna, o que por um lado facilita o aprendizado pelo algoritmo, mas acrescenta uma etapa manual, o que pode trazer variabilidade.

Ke et al. (KE *et al.*, [*s. d.*]), Lv et al. (LV *et al.*, 2021), Wang et al. (WANG *et al.*, 2021), e Ke et al. (KE *et al.*, 2023), do mesmo grupo, apresentaram em um intervalo de três anos, quatro estudos com 30, 30, 58 e 80 TC clínicas sem alterações patológicas, cortadas com centro na orelha interna e que incluía a orelha interna e o trajeto do nervo facial, porém excluía parte da mastoide, o seio sigmoide e outras regiões periféricas do osso temporal. Nestes estudos, diferentes técnicas utilizando redes convolucionais foram utilizadas para segmentação rápida da orelha interna (Dice 0,71; 0,90; 0,91 0,92), Ossículos (Dice 0,64; 0,85; 0,86; 0,89) e nervo facial (Dice 0,49; 0,77; 0,70; 0,76 em cada estudo respectivamente). No último estudo, o grupo incluiu o Conduto auditivo externo (Dice 0,83), conduto auditivo interno (Dice 0,88), Carótida interna (Dice 0,86). Observamos a evolução dos resultados apresentados pelo grupo medidos pelo coeficiente de Dice ao longo das publicações, e identificamos o uso evolutivo de técnicas de aumento sintético de dados e o ajuste de parâmetros do algoritmo como fatores que podem ter concorrido para a melhora. Entretanto, o tamanho reduzido das tomografias dos conjuntos de treinamento e testes utilizados neste resultante do corte de regiões do osso temporal favorece a precisão e da velocidade de predição apresentados nestes estudos, às custas da generalização destes modelos.

Em consonância com os resultados apresentados nesta tese, Neves et al. (NEVES *et al.*, 2021) publicou o estudo com o maior conjunto de dados até então, validando seus resultados com três diferentes arquiteturas de redes convolucionais

3D (AH-Net[8], ResNet[9] e UNet[10]). Os resultados de Dice foram 0,91, 0,71, 0,86 e 0,85 para a orelha interna, nervo facial, ossículos e seio sigmoide respectivamente. Neste estudo, os autores também apresentaram um método automatizado para a segmentação rápida da cápsula ótica, que se mostrou primordial no estudo de medição automatizada do comprimento do ducto coclear publicado pelo mesmo autor. (NEVES *et al.*, 2022).

Outros estudos foram apresentados nos últimos meses com resultados similares, porém com conjunto de dados menores (DING *et al.*, 2023; HUSSAIN *et al.*, 2021; LI *et al.*, 2020; NIKAN *et al.*, 2021). Ênfase deve ser dada ao estudo de Nikan et al. (NIKAN *et al.*, 2021), que com um conjunto de dados de 39 microTC cadavéricas de altíssima resolução (espaçamento de 50µm), apresentou resultados significativos tanto em microTC quanto em TC clínicas, o que evidencia uma potencial translação clínica daquele modelo. Recentemente, Li Z. et al. (LI *et al.*, 2022) publicou um estudo utilizando 92 TC para treinar um algoritmo de última geração que conseguiu Dice de 0,92 para a segmentação da orelha interna em 10 TC de teste.

Estas publicações demonstram o interesse de diferentes grupos em desenvolver soluções de segmentação automatizada das estruturas do osso temporal utilizando técnicas de aprendizado profundo.

1.1.3.2. Aprendizado profundo

O aprendizado profundo é parte de uma classe de algoritmos de aprendizado de máquina (programas que identificam e retêm informações de padrões e são capazes de reconhecer estes mesmos padrões em novos dados) que utilizam redes neurais de convolução para a identificação de padrões em imagens e outros dados. (ESTEVA *et al.*, 2021; GONZALEZ; WOODS, 2018). Neste processamento, os valores da intensidade luminosa de cada pixel da imagem passam por computações matemáticas para a identificação de padrões, como bordas, linhas e outras

[8] Arquitetura de rede neural híbrida anisotrópica

[9] Arquitetura de rede neural que utiliza funções residuais e conexões de atalho

[10] Arquitetura de rede neural conhecida pela sua forma em U

características que podem em conjunto agrupar informações sobre um determinado objeto.

Estes algoritmos são amplamente utilizados em diversos outros dispositivos do nosso cotidiano como telefones e computadores para a detecção de rostos, pedestres e veículos em imagens de câmeras, e potencialmente podem ter utilidade na identificação de estruturas anatômicas em exames de imagem como a tomografia computadorizada. Todos eles têm em comum serem baseados em redes neurais.

Redes neurais são funções computacionais que possuem inspiração na estrutura e funcionamento do cérebro, em que diversas camadas de unidades de processamento (neurônios) são interconectadas por sinapses que possuem diferentes pesos (facilidade ou dificuldade de ativação). (LECUN; BENGIO; HINTON, 2015). Cada neurônio em uma rede neural recebe entradas de outros neurônios ou de variáveis externas e processa essas informações por meio de uma função de ativação. O resultado da computação de cada neurônio é então transmitida às outras unidades de processamento da rede, que realizam novos cálculos com base nessas informações. (LECUN; BENGIO; HINTON, 2015).

O processo de aprendizado em uma rede neural se baseia em ajustar os pesos das sinapses entre os neurônios para minimizar o erro entre o resultado previsto e o resultado calculado pelo algoritmo. Este processo de treinamento para aprendizado do algoritmo é realizado reiteradas vezes, em um mecanismo de propagação retrógrada dos pesos a partir da comparação da resposta correta (rótulo, ou no contexto deste trabalho, cada estrutura anatômica analisada) e a predição do algoritmo pela da decomposição e análise do objeto de entrada. (LECUN *et al.*, 1998).

Concernente à tese apresentada, as explanações serão realizadas sob o ponto de vista de aplicações de aprendizado profundo no contexto da visão de computador, em que imagens (incluindo exames radiológicos) são os objetos de entrada dos sistemas.

1.1.3.3. Decomposição e análise do objeto de entrada

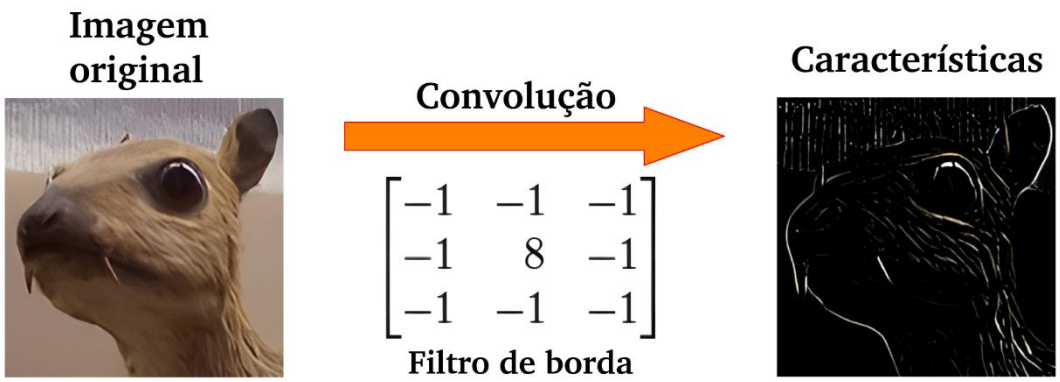
A decomposição e análise da sequência de pixels de uma imagem é uma técnica fundamental em muitas aplicações de visão computacional, incluindo o reconhecimento de padrões e o processamento de imagens. Essa técnica consiste

em converter a imagem em uma representação matemática que possa ser processada por um algoritmo de aprendizado de máquina. O processo é dividido entre propagação anterógrada e retrógrada (do inglês, *forward propagation* e *backward propagation*), que se referem às etapas em que o algoritmo lê e processa as imagens para gerar uma previsão do objeto ou estrutura anatômica (propagação anterógrada), e na etapa de propagação retrógrada, em que o algoritmo se autoajusta de acordo com a diferença entre a sua última predição e a referência, visando reduzir esta diferença e produzir resultados mais precisos. (LECUN; BENGIO; HINTON, 2015).

Nas imagens digitais, cada pixel possui um valor numérico que representa sua cor ou intensidade luminosa. A decomposição da imagem constitui na organização destes valores de pixels em uma matriz matemática (ou vetor) que representa a imagem como um todo. Classicamente, na propagação anterógrada, esta matriz passa por uma série de operações matemáticas (convoluções) que permitem, entre outras coisas, a extração de características (*feature extraction*) da imagem estudada. O uso de técnicas de convolução para a extração de características é uma prática comum em muitas tarefas de visão computacional, como reconhecimento de objetos e classificação de imagens. (GONZALEZ; WOODS, 2018).

As convoluções são operações matemáticas que usam filtros diversos que permitem, em uma imagem, o isolamento de determinada característica. Estas características são informações relevantes da imagem, e se referem às variadas formas, contornos e organização dos elementos presentes nela. (GONZALEZ; WOODS, 2018). Os filtros podem isolar e permitir a identificação de linhas ou curvas específicas que em conjunto, podem identificar um objeto em uma imagem, conforme demonstrado na figura 8.

Figura 8 - Extração de características por meio de filtros de convolução



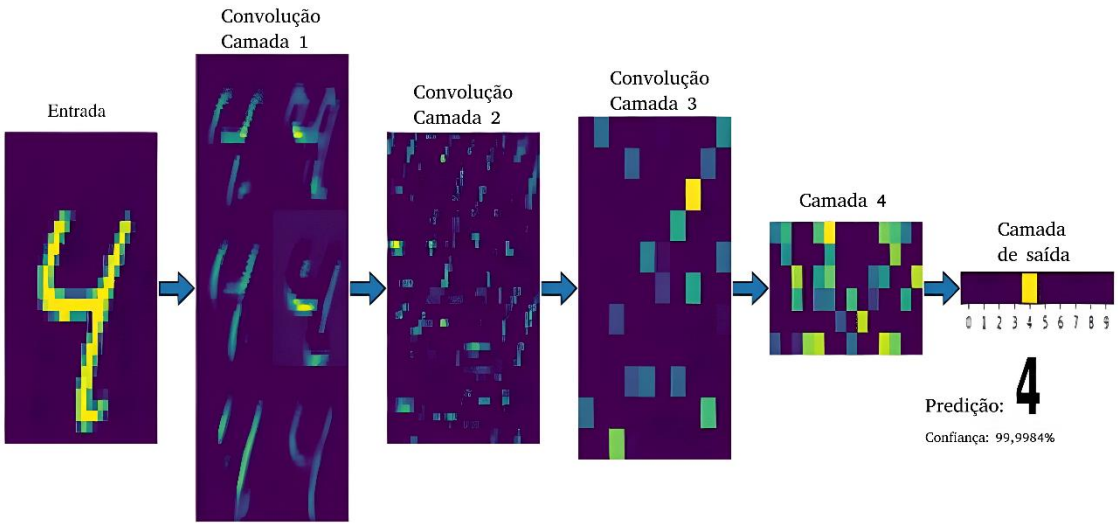
Fonte: il, color. Nvidia, Convolution for developers, 2023[11]

Nesta etapa, cada filtro que varre a imagem produz uma nova dimensão, e esta etapa é caracterizada pela mudança da quantidade de dimensões da imagem. A fim de identificar padrões não visíveis ao olho nu, diversas camadas de filtros (neurônios) são aplicadas a imagem, e isto permite uma análise matemática mais abstrata, em que os padrões analisados são mais sutis que a percepção visual.

Na sequência das operações de convoluções, funções de redução de dimensionalidade são aplicadas para isolar os padrões previamente identificados e alcançar uma predição de resposta. Entre as mais comuns, as operações de *pooling* extraem elementos importantes do espaço hiperdimensional criado pelos filtros para espaços de menores dimensões, compatíveis com a imagem original. Depois, funções de ativação permitem a identificação de associações não lineares das diferentes características dos dados, e ao final da propagação anterógrada, o algoritmo prediz uma resposta para o dado apresentado. Um esquema das camadas de convolução e predição de um algoritmo de classificação de dígitos é apresentado na figura 9.

[11] Disponível em https://developer.nvidia.com/discover/convolution, acessado em 30 de março de 2023

Figura 9 - Representação das camadas da rede neuronal da classificação de dígitos (LECUN et al. 1998)



Fonte: Densenets, Towards AI, 2020[12]

A comparação da predição do algoritmo com a referência verdadeira (*ground truth*) permite o cálculo de um valor de erro, que indica o quão distante do aprendizado eficaz o algoritmo se encontra no estágio de treinamento. Então a propagação retrógrada conduz o mecanismo de aprendizagem propriamente dito.

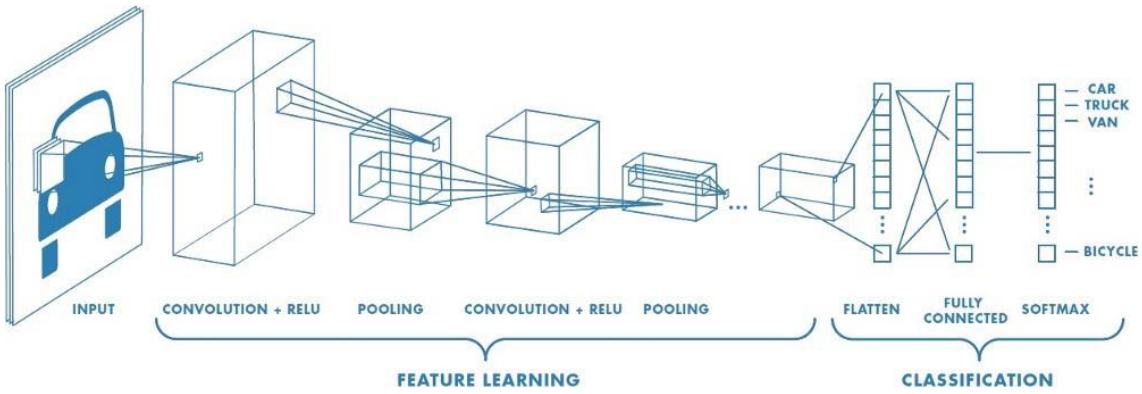
No processo de propagação retrógrada, técnicas de cálculo diferencial (gradiente) são empregadas para o ajuste dos pesos de cada neurônio, considerando o erro calculado entre a predição e o resultado correto. O cálculo do gradiente e de ajuste de pesos é então propagado retrogradamente para todas as camadas do algoritmo, para então um novo ciclo de propagação anterógrada, em busca do menor erro de validação.

A análise da sequência de pixels de uma imagem é uma etapa crítica para o sucesso de muitas tarefas de visão computacional, pois permite que o algoritmo de aprendizado de máquina compreenda as informações visuais contidas na imagem e faça previsões precisas. Essa técnica é usada em diversas aplicações, como

[12] Disponível em https://towardsai.net/p/deep-learning/state-of-the-art-convolutional-neural-networks-cnns-explained%E2%80%8A-%E2%80%8Adensenets, acesso em 05 de maio de 2023

reconhecimento de objetos, detecção de face, classificação de imagem, entre outros, com forme esquematizado na figura 10.

Figura 10 - Arquitetura de algoritmo de aprendizado profundo para classificação



Fonte: Le Reti Neurali Convoluzionali[13]

Utilizando um exemplo concreto do trabalho pioneiro de Lecun para reconhecimento de dígitos no endereço de cartas: ao algoritmo é dado uma imagem digital de um número (exemplo: o número 7), e por meio da decomposição e análise da sequência de pixels desta imagem, o algoritmo responde um conjunto de probabilidade de respostas (exemplo: 1% de ser o número 0, 4% de ser o número 2 e assim em diante, por todos os algarismos de 0 a 9). Neste caso, sendo o treinamento supervisionado, isto é, a imagem digital do número 7 foi rotulada e este rótulo é considerada a resposta correta, o sistema calcula o erro da probabilidade de previsão com a resposta correta (erro de validação), e então o mecanismo de propagação retrógrada ajusta os pesos dos neurônios pelos quais a análise foi realizada para aumentar a acurácia do algoritmo.

Vários parâmetros podem interferir no aprendizado e construção de um modelo de predição, entre eles a taxa de aprendizado, o tipo de função de ativação, a quantidade de ciclos de treinamento (chamados de épocas), e a mais importante, que é a qualidade do conjunto de dados de treinamento.

[13] Disponível em https://www.spindox.it/it/reti-neurali-convoluzionali-il-deep-learning-ispirato-alla-corteccia-visiva, acessado em junho de 2021

1.1.3.4. Conjunto de dados

O sucesso no treinamento de modelos de predição com algoritmos de aprendizado profundo é intimamente dependente da qualidade do conjunto de dados. Aspectos importantes a considerar incluem a curadoria dos dados, a representatividade estatística das diferentes características analisadas, a abordagem de estratégias para dados incompletos, a divisão adequada entre conjunto de treinamento e de teste, e o ajuste do modelo ao conjunto de treinamento. Todos estes fatores podem influenciar na robustez, precisão e generalização dos modelos gerados.

Em projetos de construção de modelos de predição para aplicações médicas algumas questões adicionais devem ser consideradas, como a privacidade, heterogeneidade, a dificuldade da anotação manual, variação entre aparelhos e protocolos de aquisição, e o pré-processamento dos dados. Estes aspectos serão abordados nos próximos parágrafos.

Dados de pacientes são protegidos por regulamentações de privacidade e confidencialidade, e por isto são mais sensíveis. Ao coletar e trabalhar com dados médicos, é fundamental que haja discrição e proteção ativa a fim de evitar vazamento de dados pessoais do paciente. Concernente ao uso de exames radiológicos para a construção de modelos de inferência, a anonimização é uma etapa crucial em que os chamados metadados (informações do paciente e do exame) são apagados de acordo com o protocolo da instituição.

Conjuntos de dados médicos geralmente envolvem dados heterogêneos e desequilibrados. A integração e o processamento desses diferentes tipos de dados podem ser desafiadores, mas são essenciais para criar modelos de aprendizado profundo eficazes e abrangentes. Acontece com alguma frequência um tipo específico de doença ou condição apresentar uma prevalência muito baixa, fazendo com que achados específicos daquela doença fiquem sub-representados considerando todo o conjunto de dados. Técnicas de balanceamento como a geração de dados sintéticos podem ser utilizadas para melhorar a qualidade do conjunto de dados.

A anotação de imagens médicas pode sofrer variação significativa entre diferentes especialistas, o que poderia afetar a qualidade de conjunto de dados e em última análise a performance do modelo de predição. Especificamente na anotação (segmentação) de estruturas anatômicas, o conhecimento detalhado da anatomia

funcional, regional, seccional, radiológica e cirúrgica, associado à sistematização e da segmentação, respeitando rigorosamente os limites de cada estrutura é de grande relevância para a qualidade do conjunto de dados. Neste quesito, é importante ressaltar a grande limitação de pessoal qualificado para segmentação manual de qualidade, que também envolve treinamento e familiaridade com as ferramentas e aplicativos utilizados na tarefa.

Adicionalmente, é esperado que tomógrafos de diferentes modelos e marcas, além de diferentes protocolos de aquisição produzam variações na qualidade, resolução e aspecto das imagens. Isto inclui aspectos como contraste, nitidez e quantidade de ruído, que pode interferir no aprendizado e na predição do modelo construído a partir destes dados. Buscando robustez e generalização do modelo de predição, a inclusão de imagens com variações resultantes de diferentes dispositivos e protocolos de aquisição pode gerar resultados melhores.

Por último, é importante controlar a qualidade dos dados de entrada pelo uso de técnicas de pré-processamento. Estas técnicas incluem ajustes no tamanho, resolução, contraste, distribuição da intensidade e remoção de artefatos das imagens, entre outros. Cada conjunto de dados requer uma abordagem de pré-processamento específica, que é desenvolvida e adaptada para o projeto em questão. Portanto, o conhecimento detalhado do conjunto de dados é essencial para o desenvolvimento dos modelos de predição.

1.1.3.5. Transformações

No contexto de processamento de imagens, transformações são procedimentos que envolvem ajustes e modificações nas imagens para melhorar a qualidade visual e facilitar a análise e processamento das imagens. As transformações aplicadas no domínio espacial da imagem incluem ajustes como a normalização de intensidade, de tamanho, de rotação, de corte, de correção de distorções, e de contraste. A escolha específica do conjunto de transformações para uma determinada tarefa visa otimizar o aprendizado do modelo de predição e melhorar seu desempenho.

A normalização é frequentemente aplicada como um passo de pré-processamento para garantir que as imagens médicas estejam em uma escala consistente e comparável antes do processamento do modelo de predição. A

normalização de escala é uma transformação essencial em processamento de imagens médicas, pois permite a padronização das intensidades das imagens de entrada, e podem permitem a remoção de artefatos metálicos, que nas tomografias computadorizadas apresentam valores de intensidade elevados comparados com os tecidos. Um exemplo é a presença de próteses metálicas na orelha média, comum em casos de estapedotomia.

A presença de objetos metálicos nas tomografias pode gerar sinais muito maiores do que o sinal do osso, dependendo do protocolo de aquisição utilizado. Isso pode resultar em uma comparação inadequada entre tecidos, já que a presença de artefatos metálicos pode levar a uma falsa interpretação da intensidade dos sinais dos tecidos. Este ajuste pode ser feito pela seleção de uma faixa de intensidade permitida e o valor de intensidade de cada pixel é normalizado em um específico, geralmente entre 0 e 1.

Ainda como parte do pré-processamento, transformações de redimensionamento do espaçamento da tomografia são necessárias em grande parte dos algoritmos, e podem ser feitas por técnicas de interpolação conforme descrito na seção Redimensionamento e interpolação.

A anotação de dados médicos é um processo complexo e oneroso pois requer a participação de especialistas para garantir a precisão das anotações. Para enfrentar essas limitações, pesquisadores e desenvolvedores podem utilizar estratégias como a aprendizagem por transferência[14], a criação de dados sintéticos e a aplicação de técnicas de ampliação de dados.

1.1.3.6. Ampliação de dados (*Data augmentation*)

O aumento sintético de dados é uma técnica essencial na construção de modelos de aprendizado profundo para imagens médicas pois ajuda a expandir e diversificar o conjunto de dados disponível. Esta abordagem é particularmente útil quando o conjunto de dados é limitado como nos projetos de segmentação anatômica

[14] Utilização de conhecimentos, técnicas e modelos pré-treinados por outras equipes ou em projetos anteriores como ponto de partida para o desenvolvimento de novos projetos de aprendizado profundo.

em volumes tridimensionais. A técnica consiste em aplicar diferentes transformações nas imagens existentes para criar novas variações, simulando variações naturais nos dados e melhorando a generalização do modelo. (SHORTEN; KHOSHGOFTAAR, 2019). Ao aplicar essas técnicas de aumento de dados, é importante garantir que as transformações utilizadas gerem novas imagens realistas de acordo com o problema em questão. Neste sentido, estas técnicas devem ser utilizadas com cuidado para evitar a introdução de artefatos ou variações irreais que possam prejudicar o desempenho do modelo.

Técnicas comuns de aumento de dados que podem ser utilizados na segmentação de estruturas otológicas a partir de tomografias incluem a rotação do volume, que simula variações no posicionamento da cabeça do paciente no momento da aquisição; escala e zoom, simulando variações na resolução, variação interpessoal do tamanho das estruturas e do tamanho da cabeça como um todo; inversão horizontal, que espelha a imagem no eixo sagital e permite reconhecimento independente de direção e simetria; ajuste de brilho e contraste, adaptando o modelo a variações de iluminação e qualidade das imagens; adição de ruído e artefatos, como ruído gaussiano, ajudando o modelo a lidar com imperfeições e artefatos reais; e deformações elásticas, que simulam variações naturais na forma e estrutura dos órgãos e tecidos, aprimorando a capacidade do modelo de lidar com diferenças anatômicas.

Esta sistematização do uso de técnicas de pré-processamento tem como objetivo preparar o conjunto de dados para o treinamento mais eficaz, e também buscando um modelo de predição robusto. A partir da uniformização do conjunto de dados de entrada e da adição de dados sintéticos pelas transformações para o aumento de dados, segue-se com o treinamento supervisionado do modelo. Neste passo, as imagens tomográficas em conjunto com as respectivas imagens de referência verdadeira (*ground truth*) são processados pela rede para permitir o aprendizado das características de cada estrutura pelo modelo.

1.1.3.7. Treinamento do modelo de aprendizado profundo

Este trabalho discute o treinamento supervisionado de estruturas anatômicas em imagens de TC. Este treinamento se caracteriza pela construção de um modelo

de predição por meio da apresentação ao algoritmo de exemplos, neste caso composto por uma imagem (neste caso de tomografia) junto com a respectiva referência verdadeira para o objeto analisado.

A referência verdadeira é delimitação ou segmentação tridimensionais de determinada estrutura anatômica naquela imagem de TC, normalmente realizada manualmente por um especialista. Em um exemplo, para uma determinada imagem de TC, um especialista delimita (segmenta) tridimensionalmente a orelha interna, produzindo uma máscara daquela estrutura, onde a segmentação, ou todas as áreas determinadas como orelha interna são consideradas primeiro plano (do inglês, *foreground*), e o restante da imagem é considerado plano de fundo (do inglês, *background*).

A máscara é chamada de mapa de rótulos (*labelmap*), um arquivo que armazena as dimensões originais da imagem de tomografia e representa numericamente cada estrutura previamente segmentada em uma ordem pré-determinada. Exemplo, 0 representa todo o plano de fundo, 1 a orelha interna e 2 o seio sigmoide.

O algoritmo aprende a identificar e localizar a partir dos processos de decomposição e análise da imagem de forma reiterada até alcançar resultados estáveis, quando determinamos que o algoritmo convergiu para um modelo estável para aquele treinamento. Vários parâmetros devem ser ajustados para otimizar o treinamento destes algoritmos. Estes incluem a divisão dos conjuntos de treinamento e teste (usados para treinar e validar o modelo), a quantidade total de iterações (quantidade de vezes que o algoritmo avalia todo o conjunto de dados), a taxa de aprendizado (fator que regula a intensidade do ajuste do algoritmo durante a propagação retrógrada), o número de imagens a serem analisadas por vez (tamanho do lote), entre outros.

Uma maneira de entender algoritmos de aprendizado profundo é considerá-los como um modelo de probabilidade de uma distribuição estatística. Considerando que o conjunto de dados apresenta uma distribuição específica, com estruturas maiores, menores, mais ou menos inclinadas e se relacionando de várias formas com os arredores, entre outros, um algoritmo eficiente deve se ajustar a esta distribuição a fim de fazer predições precisas em dados novos.

Neste sentido, é importante que o conjunto de dados seja uma amostra representativa da determinada população de interesse, assim como o conjunto de validação ou de testes. Em projetos de segmentação anatômica, uma amostra de tamanho significativo geralmente é um importante fator limitador, pois a segmentação manual das diversas estruturas em uma quantidade satisfatória de exames requer muita experiência e dedicação.

O processo requer, além de pessoal altamente especializado com conhecimento de anatomia radiológica, dedicação por longos períodos, uma vez que o processo de delineação manual é demorado e exige conhecimento profundo de anatomia radiológica. Isto pode ser um fator limitador da construção de um modelo extremamente eficaz diante da enorme diversidade de variações anatômicas e de alterações relacionadas a alterações patológicas.

Durante o treinamento do algoritmo, os dados do conjunto de treinamento são processados e as características daquelas tomografias são de certa forma incorporadas ao modelo. O conjunto de validação, que é um conjunto distinto, é utilizado para a avaliação de acurácia do modelo durante o treinamento, e o tamanho deste conjunto é importante na construção do modelo. (KOHAVI; EDU, 1993).

A divisão dos conjuntos é um assunto de debate na literatura científica, e não há um consenso estabelecido, e a proporção pode variar muito de acordo com os dados analisados e com o campo de pesquisa em questão. Um aspecto crucial é evitar que os conjuntos se intersectem, pois isto pode gerar um modelo muito preciso durante o treinamento, porém de performance fraca em dados novos, devido a generalização inadequada.

O equilíbrio entre a precisão do modelo no conjunto de treinamento e de testes, e em última análise, a robustez e adequada generalização do modelo, tem íntima relação com os conceitos de sobreajuste (em inglês, *overfitting*) e subajuste (em inglês, *underfitting*). (BISHOP, 2006). O primeiro denota a característica de um modelo se ajustar tão bem a um conjunto de treinamento que se destaca na precisão de treinamento, porém tem performance insatisfatória no conjunto de testes e em dados novos. Isto pode acontecer por desequilíbrio de classes na divisão dos conjuntos ou sobreposição de dados de treinamento e validação. Por outro lado, o subajuste é caracterizado pela incapacidade do modelo de se ajustar ao conjunto de treinamento, e eventualmente falhar na convergência. Isto pode ocorrer por falha no pré-

processamento e consequente ausência de uniformização dos dados, ou por falha do ajuste dos parâmetros de treinamento.

Diante destas considerações, destaca-se um parâmetro de extrema relevância no treinamento de modelos de aprendizado profundo: a taxa de aprendizado. Esta razão controla o grau em que o algoritmo se ajusta durante o processo de propagação retrógrada. O ajuste do algoritmo é caracterizado pela atualização dos pesos são modificados a aproximar a predição do resultado ideal, ou seja, a referência verdadeira. Uma taxa de aprendizado muito alta pode proporcionar um aprendizado inicial mais rápido, porém instabilidades podem ocorrer e dificultar a convergência do modelo. Por outro lado, taxas de aprendizado muito baixas podem proporcionar um treinamento ineficiente e longo.

A faixa ideal da taxa de aprendizado pode requerer testes repetidos em busca de valores ótimos, assim como o uso de mecanismos de ajuste como o algoritmo Adam (KINGMA; BA, 2015), que ajusta a taxa de aprendizado ao longo da rede neuronal, e também proporciona o escalonamento redutivo desta taxa de aprendizado ao longo do treinamento, a fim de favorecer a convergência do algoritmo.

Outros fatores importantes para o treinamento do algoritmo são o tamanho da entrada da rede (*network input*) e o tamanho do lote (*batch size*) de treinamento. O tamanho da entrada da rede se refere às dimensões de cada imagem usada na entrada do algoritmo. (LECUN; BENGIO; HINTON, 2015). Em muitos algoritmos este tamanho é fixo, e todas as tomografias devem ser uniformizadas com pré-transformações para se adequar a esta necessidade. Já em outras arquiteturas de algoritmos, este tamanho pode ser variável e fornecer maior flexibilidade aos dados de entrada. (LIU *et al.*, 2018-). Em todo caso, é necessário ajustar o tamanho da entrada da rede de acordo com a proporção das estruturas analisadas com relação à imagem tomográfica como um todo, e de acordo com os recursos de hardware disponíveis (menor tamanho de entrada são requeridos em sistemas mais modestos).

Por outro lado, o tamanho do lote, que se refere à quantidade de imagens que são analisadas de uma só vez, está mais intimamente ligado aos recursos computacionais disponíveis. Para este hiperparâmetro, quanto mais recursos computacionais disponíveis, maior pode ser o tamanho do lote e menor será o tempo de treinamento.

Apesar de um conjunto de dados bem curado ser o fator mais importante na construção de um modelo, a escolha da arquitetura da rede neural pode contribuir para a precisão e eficiência computacional dele. A literatura dispõe de diversos exemplos de aplicações médicas com arquiteturas de redes neurais convolucionais (CNN) como U-Net (RONNEBERGER; FISCHER; BROX, 2015), ResNet (HE *et al.*, 2016), entre outras.

A arquitetura U-Net, cujo nome é derivado do formato da rede que possui sequencias de blocos de contração seguidos por blocos de expansão, é utilizada em diversos trabalhos segmentação de imagens biomédicas .(CHEN *et al.*, 2017; ÇIÇEK *et al.*, 2016; KAMNITSAS *et al.*, 2017). Neste contexto, estas contrações se referem a etapas de convoluções que permitem capturar características importantes das imagens por um processo de redução dimensional. O conjunto das informações espaciais e visuais relevantes de uma imagem, identificadas na etapa de contração, é denominado mapa de características. Por outro lado, os blocos de expansão consistem em etapas de convoluções e concatenações que reintroduzem informações detalhadas a partir do mapa de características, permitindo assim uma segmentação precisa da imagem. (LECUN; BENGIO; HINTON, 2015).

Por outro lado, a ResNet (HE *et al.*, 2016; SUN *et al.*, 2019) tem sido uma das arquiteturas mais populares em aplicativos de segmentação de imagens. Esta rede introduziu as chamadas conexões de atalho, que permitem a identificação de características em várias escalas de resolução, impedindo a perda de precisão em redes profundas. Implementado sobre a estrutura ResNet (HE *et al.*, 2016), a rede AH-Net (LIU *et al.*, 2018-), também utilizada neste trabalho, utiliza uma rede de entrada dinâmica que permite a utilização de exames com diferentes resoluções espaciais e possui um mecanismo codificador (etapa de contração) derivado de rede 2D que tenta construir o mapa de característica eficientemente em volumes anisotrópicos, característica comum em exames clínicos.

Conforme citado anteriormente, recentemente uma nova abordagem tem sido incorporada aos algoritmos de aprendizado profundo para análise de imagens, denominada *transformers*. Apesar da discussão e detalhamento técnico desta técnica estar fora do escopo deste trabalho, os transformers adicionam um mecanismo de atenção ao aprendizado, e isto significa, simplificadamente, que o algoritmo é capaz

de aprender não só as características da imagem em variadas escalas, mas também a sequência em que as características ocorrem.

Os *transformers* têm apresentado resultados promissores em diversas tarefas (HATAMIZADEH *et al.*, 2022; LIU *et al.*, 2021), e em alguns casos requerem menos recursos computacionais para treinamento (DOSOVITSKIY *et al.*, 2020) quando comparado a outros algoritmos tradicionais como CNN. Por estes motivos, estes algoritmos têm grande potencial na construção de modelos de predição de imagens médicas. (HATAMIZADEH *et al.*, 2022; MYRONENKO; HATAMIZADEH, 2020). Hatamizadeh e equipe descreveram este algoritmo em seu estudo, demonstrando sua aplicação na segmentação semântica de tumores do crânio, com boa performance no desafio BraTs.

A adaptação de algoritmos já desenvolvidos e testados por outras equipes pode incluir a modificações do tamanho da entrada da rede, da quantidade e composição das camadas de redes neurais, entre outros. Com a expansão de projetos de aprendizado profundo (AMODEI *et al.*, 2016; RAJPURKAR *et al.*, 2017; SILVER *et al.*, 2017), é comum o uso de aprendizagem por transferência a fim de reduzir o tempo e recursos necessários para o treinamento de novos algoritmos a partir do zero.

Por fim, nas tarefas de segmentação anatômica em exames de imagem, em que o planejamento e a navegação cirúrgicos são as principais motivações, é mister uma análise objetiva do desempenho do modelo de predição. A precisão da segmentação automatizada pode garantir a implementação do modelo de predição em projetos de aplicação clínica. No entanto, não existe uma métrica padrão-ouro para a medida da acurácia dos modelos. Em vez disso, existe um conjunto de métricas que, quando utilizadas em conjunto, permitem avaliar adequadamente o desempenho do modelo.

As métricas objetivas podem ser divididas em dois grupos, as medidas de precisão, que em sentido amplo medem a taxa de similaridade ou de intersecção do resultado da previsão com a referência verdadeira; e as medidas de erro, que apontam para a diferença entre o previsto e a referência.

Entre as medidas de precisão mais utilizadas para avaliação de segmentação em imagens médicas temos o coeficiente de similaridade de Dice (Dice), a precisão balanceada e a Similaridade de volume.

O coeficiente de Dice é um método de avaliação que considera a razão entre a intersecção dos objetos analisados e a soma dos seus volumes. A equação desta métrica é dada por:

$Dice = \frac{2 \times A \cap B}{A+B}$ ou $Dice = \frac{(2 \times VP)}{2 \times VP+FP+FN}$, onde VP, FP e FN significam verdadeiro positivo, falso positivo e falso negativo respectivamente. O Dice é provavelmente a medida mais utilizada na descrição da precisão de modelos de segmentação em imagens médicas, utilizado durante o treinamento e validação dos modelos. Este índice varia de 0 a 1, e é eventualmente discriminado como um valor percentual.

A precisão balanceada (PB) é outra medida que pode oferecer informações sobre a precisão do modelo de segmentação. Esta técnica leva em consideração tanto a sensibilidade quanto a especificidade, e pode complementar a avaliação da precisão pelo coeficiente de Dice. A PB apresenta em porcentagem qual a proximidade do resultado da previsão com a referência (SHERER *et al.*, 2021), e sua equação é dada por:

$PB = 1/2 \times \left(\left(\frac{VP}{VP+FN}\right) + \left(\frac{VN}{VN+FP}\right)\right)$, onde VN é verdadeiro negativo.

A similaridade de volume (SV) se refere a um cálculo da razão entre o volume da segmentação automatizada pelo volume proveniente da segmentação manual, considerada a referência. Esta avaliação objetiva descrever a diferença média entre as duas técnicas de segmentação. Neste estudo, optamos por utilizar esta razão pois ela pode discriminar possíveis tendências sistemáticas de sub ou superestimação do volume da estrutura pelo modelo de segmentação.

$SV(\%) = \frac{(VA-VM)}{VM} x100$, onde VA [15] é o volume da estrutura segmentada automaticamente, e VM[16] pelo método manual.

Por outro lado, a principal medida de erro na análise de segmentação anatômica é a dada pela Distância de Hausdorff (HD, de *Hausdorff's Distance*). Essa medida, que de maneira simplificada pode ser entendida como a diferença entre as bordas de segmentações manuais e automatizadas, é frequentemente usada no desenvolvimento de sistemas de navegação cirúrgica (NEVES *et al.*, 2021).

[15] VA – Volume automatizado

[16] VM – Volume manual

Apesar de ser um conceito mais abstrato matematicamente, a HD é uma métrica importante que expressa a incerteza da borda da segmentação. No entanto, devido à sensibilidade da HD a pontos de dados aberrantes, os estudos de segmentação biomédica costumam utilizar a distância de Hausdorff média (AHD, *Average Hausdorff Distance*), que é a média das HD por todos os pontos da imagem. Esta variante da HD é menos sensível a dados aberrantes e oferece informação relevante acerca da precisão do modelo de segmentação. Para uma perspectiva mais ampla sobre o modelo, as avaliações de precisão podem também utilizar a HD95, que é a Distância de Hausdorff dentro do percentil 95º. Esta técnica oferece uma ideia da margem de erro máxima da segmentação automatizada, apesar de ser mais sensível a dados aberrantes. (DUBUISSON; JAIN, 1994).

1.1.4. Estruturas anatômicas de interesse

O osso temporal, que constitui a região lateral da base do crânio, abriga importantes estruturas anatômicas, sobretudo vasculares e nervosas, que podem ser tanto alvo de manipulação quanto de preservação em uma abordagem cirúrgica nesta região. Nesta seção, as principais estruturas do osso temporal consideradas relevantes no planejamento cirúrgico de cirurgias otológicas serão detalhadas e sua escolha justificada para seleção, segmentação manual e construção do modelo de segmentação automatizado. Detalhes tomográficos de cada estrutura incluem a disposição, uniformidade de intensidade e interface com os arredores, que são relevantes para o potencial aprendizado do modelo.

1.1.4.1. Carótida interna

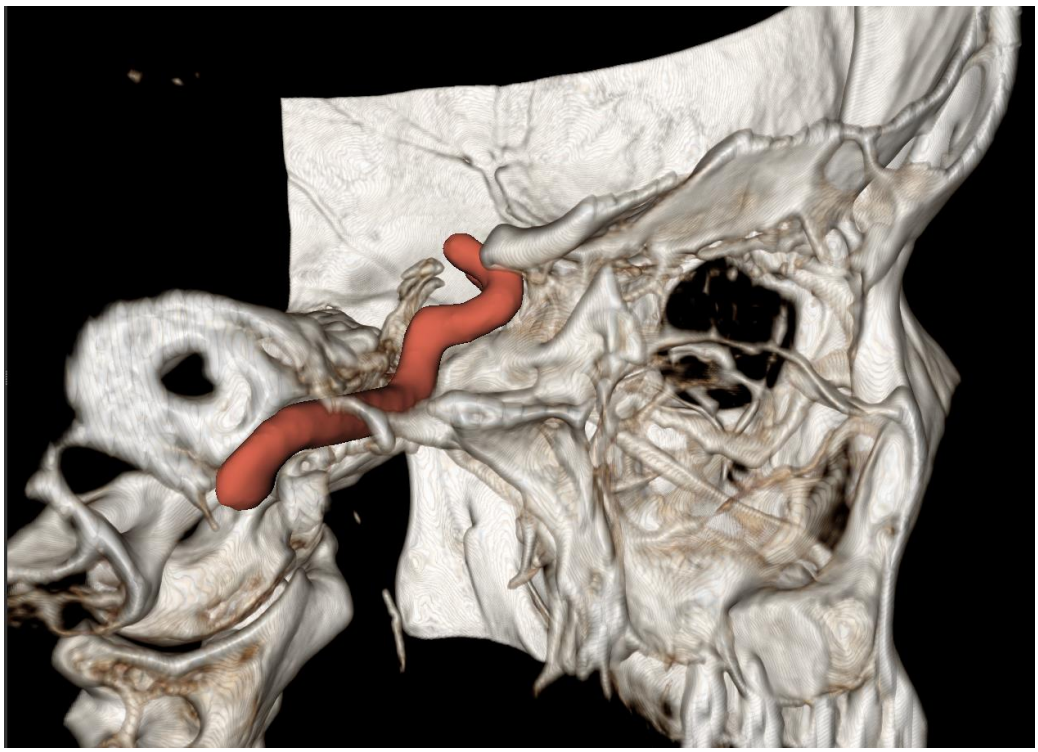
A artéria carótida interna (ACI) é o vaso que conduz o maior fluxo sanguíneo arterial para o crânio, e sua importância é vital para o organismo. Anatomicamente, penetra no osso temporal em trajeto ínfero-superior pelo canal carotídeo, se deflete anterior e medialmente na região petrosa do osso temporal e segue um trajeto aproximadamente retilíneo até o forame lácero, onde se deflete superiormente para a região cavernosa. Na primeira deflexão, a ACI possui relação próxima com a cóclea,

e pode apresentar deiscências. Nesta região a ACI também possui íntima relação com o músculo tensor do tímpano e com a porção óssea da tuba auditiva.

Em imagem de TC sem contraste, técnica exclusiva abordada neste trabalho, a ACI apresenta densidade de partes moles circundada pelo denso osso da região petrosa, e seu trajeto é menos claro a partir do segmento cavernoso. Variações anatômicas incluem graus diversos de estreitamento e curvatura do canal carotídeo, proximidade com a orelha interna, interface maior ou menor com pneumatizações do osso temporal inclusive no ápice petroso, além de deiscências. Calcificações das paredes da artéria, geralmente em regiões de deflexão, podem aparecer na tomografia como áreas de densidade de osso e ser um desafio para a segmentação.

A figura 11 demonstra o trajeto da artéria carótida interna direita desde a entrada do canal carotídeo até o segmento pós-clinóide.

Figura 11 - Renderização da artéria carótida interna (vermelho) em tomografia computadorizada, vista lateral direita.



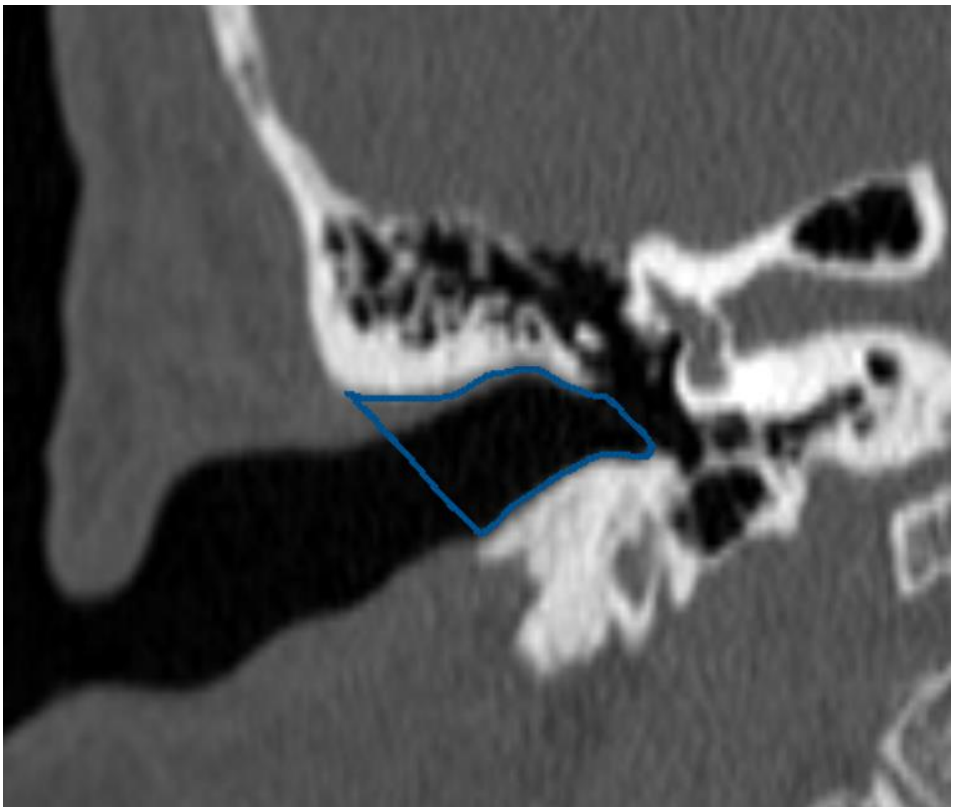
Fonte: Elaborado pelo autor (2023).

1.1.4.2. Conduto auditivo externo

O conduto auditivo externo (CAE) se estende da sua abertura externa na região do pavilhão auditivo até a membrana timpânica. Em indivíduos sem alterações patológicas, o CAE é geralmente um canal ósseo preenchido por ar e recoberto por delgado tecido epidérmico, grosso modo de trajeto em direção lateral-medial. Porém, devido a alterações inflamatórias, tumorais ou congênitas, o CAE pode ser preenchido por partes moles ou áreas de hiperostose, prejudicando a sua identificação.

Cirurgicamente, o CAE apresenta importância crucial no planejamento de procedimentos realizados por via transcanal, cada vez mais explorada pela expansão das técnicas endoscópicas para cirurgia otológica. A análise do grau de deflexão do CAE pode ser importante no planejamento de abordagens infracocleares e para o acesso de diferentes afecções otológicas. A delineação do CAE em TC é demonstrada na figura 12.

Figura 12 - Conduto auditivo externo direito (azul) visto em TC de osso temporal



Fonte: Elaborado pelo autor (2023)

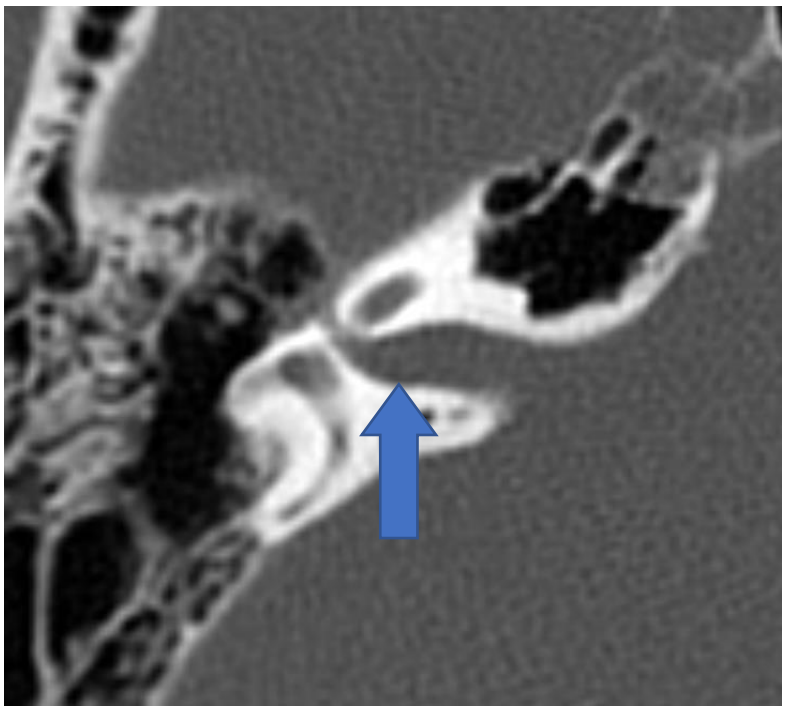
1.1.4.3. Conduto auditivo interno

O conduto auditivo interno (CAI) é o canal ósseo do osso temporal entre o poro e o fundo, pontos de abertura para a fossa posterior medialmente e para o vestíbulo lateralmente, respectivamente. Na tomografia, o CAI aparece como um canal preenchido por material com densidade de partes moles envolto por osso cortical de espessura variável de acordo com a pneumatização do osso temporal. Nele passam os importantes VII e VIII nervos cranianos provenientes do tronco cerebral com destino ao osso temporal.

O CAI pode ter formatos variados, incluindo o formato cônico, de base maior no poro, como também semelhante a um fuso, com a região central mais larga que as extremidades, mesmo em indivíduos normais. O planejamento cirúrgico e a monitorização intraoperatória de ressecção de schwannomas do VIII podem se beneficiar da segmentação do CAI por conter estruturas nobres incluindo neste caso o alvo cirúrgico.

A figura 13 mostra a disposição anatômica do CAI em TC (axial).

Figura 13 - Conduto auditivo interno (seta azul) direito em TC de osso temporal



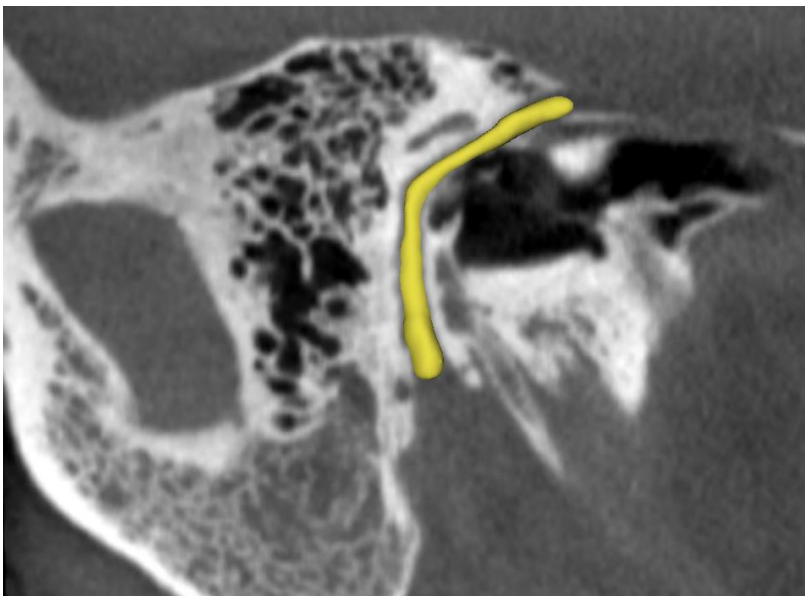
Fonte: Elaborado pelo autor (2023)

1.1.4.4. Nervo facial

Provavelmente a estrutura anatômica mais relevante em cirurgias otológicas em geral, o nervo facial (NF), responsável pela mímica facial, possui um trajeto com diversas flexões e interrelações marcantes com outras estruturas do osso temporal. Após passar pelo CAI, o NF passa através da cápsula ótica superiormente à cóclea, e se deflete anteriormente, onde compreende o gânglio geniculado. O NF, então sobre a região do epitímpano anterior, inicia seu trajeto de direção póstero-lateral denominado segmento timpânico, que tem interface com a orelha média latero-inferiormente e com a cápsula ótica medial-superiormente. Então o NF se deflete inferiormente, no seu segmento mastoideo até o forame estiloideo.

A identificação precisa do trajeto do NF tem grande importância nas cirurgias otológicas por ser primariamente uma estrutura a ser preservada, e devido ao seu trajeto peculiar, pode surpreender cirurgiões durante as abordagens. Neste trajeto, o NF possui ampla variação da sua interface com os arredores, e destaque deve ser dado a possível mudança do conteúdo da orelha média e cavidades mastoides causadas por problemas inflamatórios, o que torna o NF uma estrutura particularmente difícil de segmentar. (Figura 14)

Figura 14 - Nervo facial (em amarelo) em uma TC de osso temporal, reformatação oblíqua



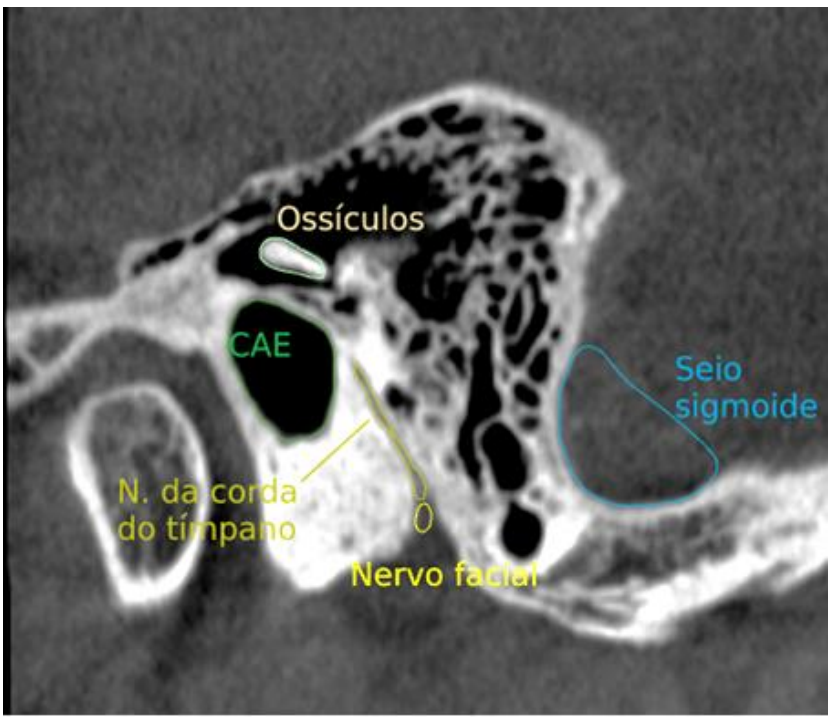
Fonte: elaborado pelo autor (2023)

1.1.4.5. Corda do tímpano

O nervo da corda do tímpano (NCoT) é responsável pela condução de estímulos gustatórios aferentes e de estímulos parassimpáticos para as glândulas salivares. Este nervo emerge do segmento mastoideo do NF e se direciona superior e anteriormente para a cavidade timpânica, para então seguir seu caminho para a fossa infratemporal. A sua principal importância clínica reside no fato de ser a margem lateral do recesso do facial, área de abordagem do acesso transmastoideo à janela redonda para cirurgia de implante coclear.

Na TC, o NCoT é mais bem visualizado nos cortes sagitais e coronais, e apresenta densidade de partes moles permeado por osso cortical, rodeado por variados graus de pneumatização. Devido ao seu pequeno volume e localização em uma região passível de grandes variações morfológicas e patológicas, a segmentação do NCoT é desafiadora. A figura 15 apresenta a delineação do nervo corda do tímpano em TC.

Figura 15 - Segmentação de estruturas do osso temporal incluindo o nervo corda do tímpano. TC em reformatação sagital



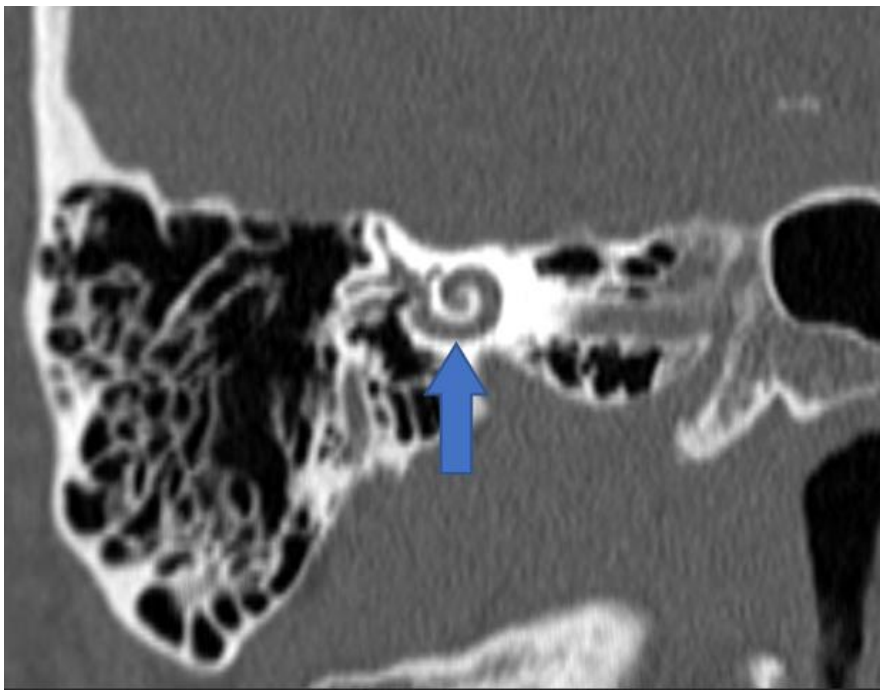
Fonte: elaborado pelo autor (2023)

1.1.4.6. Orelha interna

A orelha interna (OI) constitui o labirinto membranoso, está localizada em região centralizada do osso temporal e é composta pela cóclea anteriormente e pelo vestíbulo posteriormente. Anatomicamente, a OI está envolta pela cápsula ótica e tem relação com o CAI medialmente, NF superiormente e medialmente, orelha média medialmente e ACI anteriormente. A OI, em última análise, é responsável pela sensibilidade a estímulos sonoros e de movimento da cabeça, e nisto reside a sua vital importância na audição e no equilíbrio.

Alguns pontos anatômicos de importância clínica são as janelas oval e redonda, esta última por onde eletrodos são inseridos em cirurgias de implante coclear. Nas imagens de TC, a OI é uma região de densidade de líquido geralmente bem delimitada pela diferença de densidade da cápsula ótica e com pouca variação anatômica interpessoal o que faz desta estrutura uma boa candidata para a tarefa de segmentação automatizada. A figura 16 mostra a identificação da orelha interna em TC.

Figura 16 - Orelha interna direita em TC, reformatação oblíqua



Fonte: Elaborado pelo autor (2023)

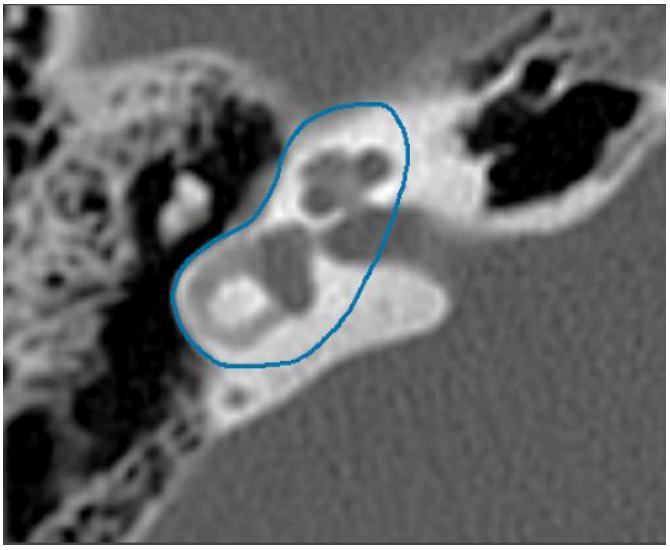
1.1.4.7. Cápsula ótica

A cápsula ótica (CO) é o osso compacto que envolve a orelha interna. Anatomicamente, está relacionado nas suas bordas internas com a OI, e externamente com a fossa craniana média superiormente, a fossa posterior posterior e medialmente, com a ACI antero-lateralmente, e medialmente com a orelha média. Nas imagens de TC, a CO é uma região radiointensa com pouca pneumatização, e tem com a OI uma transição marcante pela interface osso-líquido consistente na maior parte da sua extensão.

A integridade da CO em volta dos canais semicirculares é um aspecto de importante relevância clínica. A ausência desta integridade caracteriza a deiscência dos canais semicirculares, que pode ser causa de sintomas como perda auditiva e tonturas. Conhecer o tamanho e a localização de tais déficits em relação à anatomia circundante pode ajudar a selecionar abordagens cirúrgicas ideais (por exemplo, transmastoide versus fossa média). A partir da segmentação da CO, a possibilidade de identificação do nicho da janela redonda pode ajudar os cirurgiões a preverem e planejar abordagens para o ouvido interno durante o implante coclear.

A figura 17 a seguir demonstra a posição da cápsula ótica em volta da orelha interna.

Figura 17 - Cápsula ótica envolvendo orelha interna em TC, corte axial



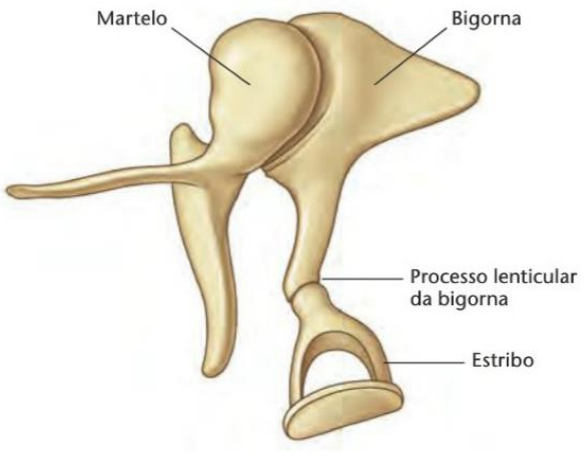
Fonte: Elaborado pelo autor (2023)

1.1.4.8. Ossículos

A cadeia ossicular da orelha média, constituída pelo martelo, bigorna e estribo, é responsável pela transmissão da vibração sonora da membrana timpânica até a orelha interna. Neste trabalho nós nos referiremos ao conjunto dos constituintes da cadeia ossicular como ossículos e a segmentação incluirá os três indiscriminadamente. Os ossículos estão localizados na orelha média, circundados por ar nas cavidades aeradas ou eventualmente por secreção ou mucosa edemaciada em casos de inflamação.

Em imagens tomográficas, os ossículos possuem aspecto radiodenso em sua maior constituição, sobretudo no colo e cabeça do martelo assim como no corpo da bigorna. Porém, pelas pequenas dimensões, as extremidades do ramo longo da bigorna e do cabo do martelo, assim como as cruras do estribo podem ser difíceis de identificação, sobretudo em casos inflamatórios. A segmentação dos ossículos pode fornecer informações relevantes acerca da integridade deles e sua disposição relativa a outras estruturas, possibilitando a avaliação de possíveis estratégias cirúrgicas para cada caso. O desenho dos ossículos está demonstrado na figura 18 a seguir.

Figura 18 - Desenho dos ossículos, visão medial.



Ossículos da audição direitos articulados (vista medial)

Fonte: (DRAKE, 2011)

1.1.4.9. Seio sigmoide

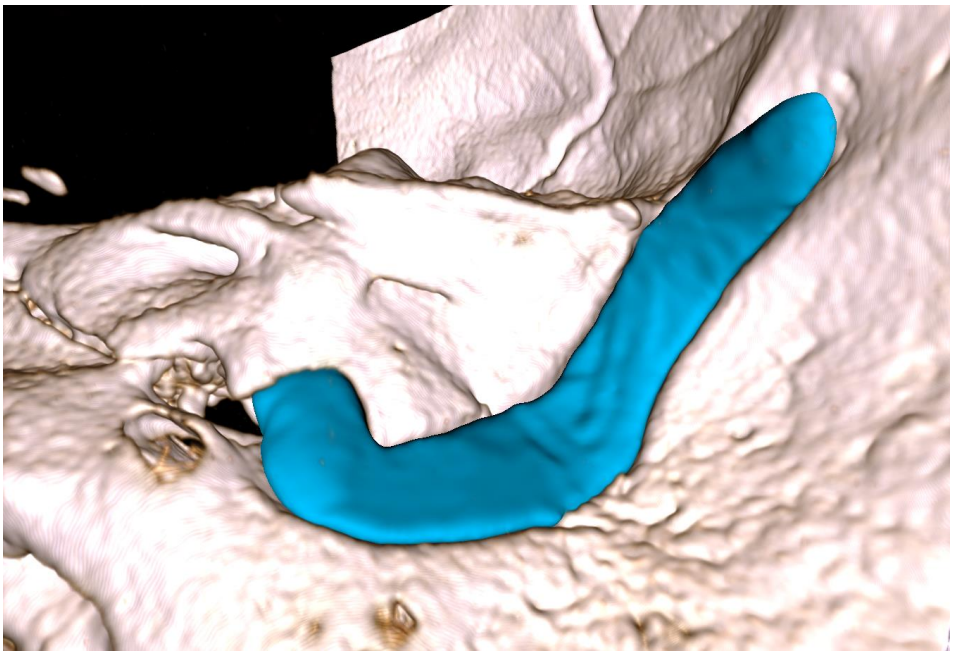
O seio sigmoide (SS) é o principal vaso de drenagem de sangue venoso do crânio. Está localizado na parede lateral-posterior do crânio e escoa em trajeto medialmente à mastoide todo o fluxo venoso da região. O SS se inicia a partir do seio transverso e toma um trajeto descendente e depois medial pela parede lateral da fossa posterior. Então o SS se deflete em graus variados onde recebe o nome de bulbo da jugular, e a partir da sua emergência pelo forame jugular em direção crânio-caudal passa a ser denominada jugular interna. Neste estudo, considerando a importância destes diferentes segmentos na cirurgia otológica, denomina-se SS as porções a partir do seio transverso até o forame jugular.

O SS pode conter deiscências ósseas em seu trajeto, e estas podem ser causa de sintomas como zumbido pulsátil, podendo levar a tratamento cirúrgico. Em cirurgias que incluem mastoidectomia, a segmentação do SS pode oferecer dados importantes para o planejamento cirúrgico e monitorização intraoperatória desta estrutura. Achados normais podem incluir grande variação no calibre do SS (SINGH *et al.*, 2019; VAN OSCH *et al.*, 2019), graus variados de curvatura na região justa-mastoidea, e da curvatura e extensão do bulbo da jugular.

A característica marcante do SS, nas imagens tomográficas, de não ter clara transição para o conteúdo da fossa posterior, torna a segmentação do SS relativamente difícil, pois apenas a impressão do SS no osso pode ser utilizada como parâmetro anatômico.

O entendimento completo do trajeto do seio sigmoide por meio de imagens bidimensionais pode não ser fácil, e a figura 19 apresenta imagem do seio sigmoide em uma modelagem 3D.

Figura 19 - Renderização 3D do seio sigmoide (azul) do lado direito como visto a partir da fossa posterior.



Fonte: elaborado pelo autor (2023)

1.2. JUSTIFICATIVA

A cirurgia otológica representa o tratamento de escolha para diversas doenças inflamatórias e de alteração auditiva. Entretanto, a complexidade da anatomia do osso temporal torna estes procedimentos particularmente desafiadores, exigindo treinamento extenso e cuidadoso.

Atualmente, o treinamento em cadáveres é considerado o padrão ouro, porém apresenta limitações importantes, sobretudo na incapacidade de simular alterações patológicas. Alternativamente, simuladores de realidade virtual emergiram como uma ferramenta de treinamento promissora, mas a sua eficácia depende do delineamento preciso das estruturas anatômicas, uma tarefa exigente em termos de dedicação e especialização.

A navegação cirúrgica intraoperatória avançada depende fortemente da segmentação anatômica, e pode se beneficiar significativamente de métodos de segmentação aprimorados. A segmentação automatizada e precisa de estruturas anatômicas-chave é um passo crucial no desenvolvimento de sistemas inteligentes de treinamento e navegação que podem auxiliar cirurgiões de diversos graus de experiencia, melhorando os resultados para os pacientes.

Estudos preliminares sugerem que técnicas deaprendizado profundo poderiam ser utilizadas para o desenvolvimento destes sistemas de segmentação anatômica, porém apresentam amostras escassas, poucas estruturas e precisão limitada.

Este estudo se justifica pela necessidade do aprimoramento da segmentação anatômica em exames radiológicos para o planejamento pré-operatório e navegação intraoperatória personalizados em procedimentos otológicos por meio do uso de técnicas de aprendizado profundo utilizando uma base de dados robusta.

2 OBJETIVOS

Geral:

Desenvolver e validar um sistema de segmentação automatizada de estruturas-chave do osso temporal em TC utilizando algoritmos de aprendizado profundo, com vistas a otimizar o planejamento pré-operatório e navegação intraoperatória de cirurgia otológica.

Específicos:

1. Construir uma base de dados de TC de osso temporal com estruturas anatomicamente relevantes para cirurgia otológica manualmente segmentadas;
2. Adaptar algoritmos de aprendizado profundo para o treinamento supervisionado de modelos de segmentação anatômica em TC de osso temporal;
3. Construir modelos de segmentação automatizada de estruturas anatômicas em TC;
4. Desenvolver técnicas para a segmentação automatizada da cápsula ótica;
5. Implementar servidor de segmentação anatômica para TC de osso temporal acessível remotamente pela internet;
6. Avaliar a precisão do algoritmo com uso de métricas objetivas validadas na literatura biomédica, incluindo Dice, precisão balanceada, similaridade volumétrica, distância de Hausdorff e tempo de processamento;
7. Comparar a eficiência e a precisão dos modelos desenvolvidos com métodos existentes de segmentação anatômica;
8. Discutir o potencial dos modelos desenvolvidos para o aprimoramento dos simuladores de realidade virtual e de navegação intraoperatória.

3 MÉTODOS

3.1. Tipo de Estudo

Trata-se de um estudo experimental realizado no Laboratório de Realidade Aumentada do Departamento de Otorrinolaringologia da Universidade de Stanford, entre julho de 2019 a janeiro de 2023. No estudo foram desenvolvidos modelos para a segmentação automatizada de estruturas anatômicas do osso temporal, utilizando algoritmos de aprendizado profundo. Estes algoritmos foram adaptados para o treinamento supervisionado de conjuntos de imagem de TC juntamente com suas respectivas segmentações manuais de referência.

3.1.1. Comitê de ética

O estudo foi iniciado após aprovação do comitê de ética da Universidade de Stanford (N.º 38946) que garantiu a dispensa de consentimento informado, uma vez que este estudo foi conduzido com dados anonimizados.

3.2. Amostra

O estudo foi dividido em 2 etapas de experimentos, com um intervalo de dois anos entre os ensaios. No primeiro experimento, o pesquisador analisou cinco estruturas anatômicas do osso temporal em 150 TCs. No segundo experimento, outras 175 TCs foram acrescentadas ao conjunto de dados, totalizando 325 amostras. Nesta segunda etapa, além das cinco estruturas estudadas anteriormente, quatro estruturas anatômicas adicionais foram também analisadas, perfazendo um total de nove estruturas anatômicas examinadas ao longo do estudo.

3.3. EXPERIMENTO 1 – MODELOS DE ESTRUTURAS ÚNICAS

O primeiro experimento se constituiu no desenvolvimento e avaliação de um sistema de segmentação automatizada da orelha interna (OI), nervo facial (NF), Ossículos, Seio sigmoide (SS) e cápsula ótica (CO), que são estruturas anatômicas

relevantes no planejamento pré-operatório de cirurgias otológicas e da base lateral do crânio.

3.3.1. Base de dados

Foram selecionadas 150 TC do osso temporal com resolução variável entre 0,125 e 0,3m. Exames com anatomia normal e aqueles alterados por doença ou intervenções (por exemplo, otomastoidites, pós-operatório etc.) foram incluídos. Foram excluídas imagens com que apresentavam artefato de movimento ou artefatos metálicos que prejudicassem a identificação das estruturas.

3.3.2. Estruturas anatômicas selecionadas para segmentação

3.3.2.1. Nervo facial

O NF foi segmentado do gânglio geniculado até o forame estilomastoideo. A segmentação manual incluiu os voxels contíguos no trajeto anatômico do NF com intensidade -200 a 500 HU, que corresponde a faixa de partes moles até o limite do osso compacto (>500HU).

3.3.2.2. Orelha interna

A OI foi segmentada considerando todo sinal de partes moles dentro da cápsula ótica, até a janela redonda e incluindo os giros da cóclea e os canais semicirculares. A faixa de intensidade para a segmentação da OI variou de -300 a 500 HU. Cuidado foi tomado para não segmentar porções dos feixes nervosos do modíolo, eventuais deiscências do nervo facial e regiões do aqueduto vestibular porventura alargado.

3.3.2.3. Ossículos

A segmentação dos ossículos incluiu o martelo, a bigorna e as partes visíveis do estribo, na faixa de 0 até 2000HU. Parte da faixa que compreende tecidos moles foi

incluída, respeitando as referências anatômicas, para incluir o ramo longo da bigorna e as cruras do estribo, cujos valores HU são menores.

3.3.2.4. Seio sigmoide

A segmentação do SS se estendeu da sua transição do seio transverso, pela região justa-mastoidea, bulbo da jugular, até a sua eminência pelo forame jugular. A principal referência anatômica foi a impressão do SS na região medial do osso temporal.

3.3.3. Segmentação de referência

A segmentação (tridimensional) manual das estruturas-chave foi realizada pelo autor após treinamento em segmentação anatômica em exames de imagem. Um programa análise de imagens médicas de código aberto (3D Slicer (FEDOROV *et al.*, 2012)) foi utilizado para gerar as anotações volumétricas das estruturas de interesse.

O treinamento em segmentação anatômica foi realizado por duas horas diárias durante 6 meses, com recursos online e sob supervisão do diretor do laboratório Dr Nikolas Blevins.

3.3.4. Pré-processamento

Com o objetivo de uniformização do conjunto de dados para serem utilizados no treinamento do algoritmo, as transformações de pré-processamento das imagens incluíram:

3.3.4.1. Adequação da dimensão e espaçamento das tomografias para um espaço isotrópico de 0,25mm.

Após análise da demanda computacional do algoritmo, e considerando a resolução requerida para a identificação adequada das estruturas anatômicas, o

espaçamento de 0,25mm foi considerado ideal e aplicado em todo o conjunto de dados.

Este valor é compatível com aparelhos de tomografia de última geração, permite segmentação de pequenas estruturas como os canais semicirculares, e é computacionalmente eficiente, considerando os recursos disponíveis.

3.3.4.2. Transformação dos volumes de lado esquerdo para direito

Com o objetivo de evitar a introdução de viés de seleção, no conjunto de treinamento, as TC de osso temporal esquerdo foram transformadas para TC de osso temporal direito através da modificação da matriz posicional das imagens. Assim, todas as tomografias do conjunto de treinamento eram natural ou artificialmente do lado direito, e na etapa de aumento de dados, o algoritmo as transformou aleatoriamente em lateralidade esquerda durante o treinamento.

3.3.4.3. Normalização da intensidade:

Foi realizada a padronização da intensidade das imagens entre -500 a 2000 Unidades Housnfield (HU, do inglês *Hounsfield Units*), isto é, intensidades abaixo de -500 foram consideradas -500 e acima de 2000 foram consideradas 2000 HU. Isto evita impedir que objetos metálicos (valores HU elevados, como 10000HU) distorçam a faixa de valores analisados e eventualmente impeçam a convergência do algoritmo.

3.3.4.4. Aumento de dados:

Neste experimento as seguintes transformações foram utilizadas durante o treinamento:

- Variação da intensidade da imagem em ±10%: A cada época do treinamento, a intensidade das TC era aleatoriamente modificada no intervalo de mais ou menos 10%.

- Inversão da lateralidade das imagens: A cada época, as TC foram aleatoriamente invertidas entre lado direito e esquerdo através de técnica de espelhamento.

3.3.5. Divisão do conjunto de dados entre treino e teste

As imagens foram aleatoriamente divididas entre dois grupos: o de treinamento, com 125 imagens, e o de teste, com 25 imagens. Dentro do grupo de treinamento, as imagens foram subdivididas em 4 partes para o treinamento propriamente dito e 1 parte (25 imagens) para a validação cruzada. A validação cruzada é a verificação da performance do modelo durante o treinamento, realizado em um conjunto de validação.

3.3.6. Arquiteturas das redes[17]

Três arquiteturas de redes de convolução (U-Net, ResNet e AH-Net) (HE *et al.*, 2016; LIU *et al.*, 2018-; RONNEBERGER; FISCHER; BROX, 2015) foram adaptadas para a tarefa de treinamento. Todas estas arquiteturas são validadas na literatura para tarefas de identificação de objetos e para segmentação de imagens biomédicas.

3.3.7. Treinamento do algoritmo

Um servidor de treinamento com sistema operacional Linux (Ubuntu 18.04) com 40GB de memória RAM[18] e 24GB de memória de vídeo (2 unidades de NVIDIA TITAN XP) foi utilizado para a tarefa. A plataforma Clara SDK, um conjunto de aplicativos de código aberto desenvolvido para acelerar projetos de aprendizado profundo em imagens médicas (NVIDIA CLARA IMAGING, 2022) foi utilizado para o desenvolvimento da aplicação.

---

[17] Três diferentes arquiteturas de redes neurais previamente validadas em estudos de segmentação anatômica em imagens radiológicas foram adaptadas para este trabalho.

[18] Memória de curto período, componente de computadores para armazenamento de informações de uso imediato

Neste experimento foi implementado o treinamento de um modelo por estrutura (modelo de rótulo único ou estrutura única) em vez de um único modelo para todas as estruturas (de múltiplas estruturas). Esta abordagem é conhecida como treinamento de modelo *single-label*.

Para cada estrutura de interesse, o treinamento foi conduzido por 2000 épocas (ciclos de análise de todas as imagens do conjunto de treinamento), com taxa de aprendizado (*learning rate*) inicial de 10^{-4}, que decaíam a um terço do valor a cada 400 épocas.

3.3.8. Servidor de predições

Um servidor de internet para a predição remota de novos dados foi implementado com os modelos gerados pelo treinamento do algoritmo. Utilizou-se o módulo de anotação assistida por inteligência artificial da plataforma 3D Slicer (SACHIDANAND *et al.*, 2019) para gerar a segmentação automatizada da orelha interna, nervo facial, ossículos e seio sigmoide nas 25 tomografias do conjunto de testes. Neste conjunto de dados a segmentação foi acompanhada de pós-processamento, que incluiu a remoção de artefatos falsos positivos e maximização das margens das segmentações dentro de faixas de intensidade Hounsfield para cada estrutura.

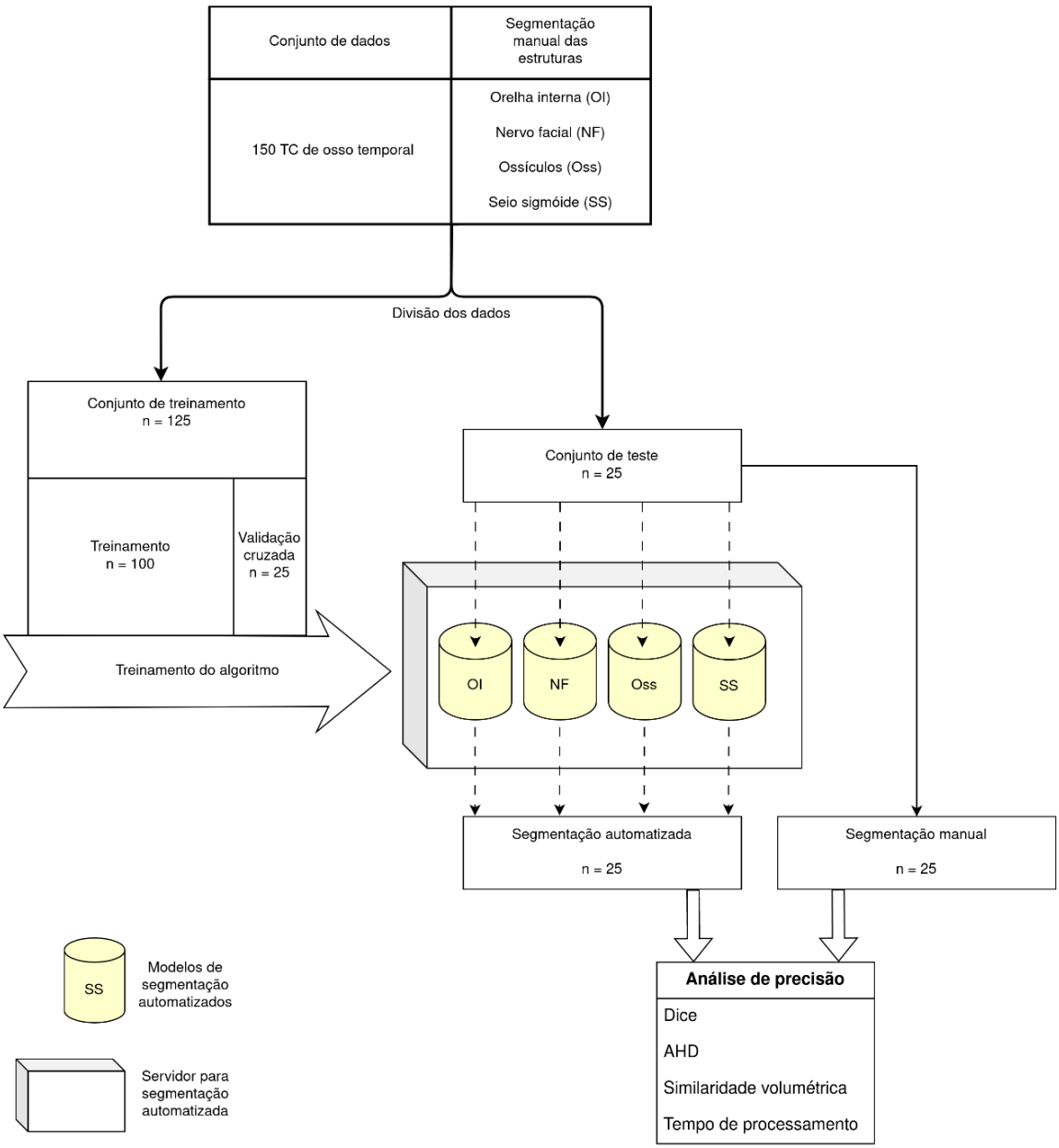
3.3.9. Cápsula ótica

Neste primeiro experimento, uma abordagem computacional para a segmentação automatizada da CO foi desenvolvida a partir da segmentação da OI. Utilizando funções do editor de segmentos da plataforma 3D Slicer, a segmentação da CO se inicia como uma cópia da segmentação da OI, depois o programa desenvolvido expande o volume da OI para a região adjacente em direção centrífuga, mas seleciona apenas voxels com intensidade acima de 500HU, que corresponde a osso cortical.

3.3.10. Avaliação objetiva

A avaliação objetiva foi conduzida utilizando o coeficiente de similaridade de Dice, a distância de Hausdorff média (AHD) e o coeficiente de similaridade de volume. O tempo para segmentação automatizada e manual foram medidos para cada estrutura do conjunto de testes e esta comparação foi utilizada para analisar a eficiência do processo de segmentação. A figura a seguir apresenta o diagrama do estudo.

Figura 20 - Diagrama do experimento 1 (modelos de estruturas únicas)



Fonte: elaborado pelo autor (2023)

3.3.11. Avaliação subjetiva

Uma avaliação subjetiva foi conduzida através da avaliação de 4 tomografias segmentadas (2 delas manualmente segmentadas e outras 2 com segmentação automatizada, sem conhecimento dos participantes). Sete profissionais (cirurgiões otológicos e neuro-radiologistas) responderam a questionários baseados em pontuação em escala de Likert de 5 pontos para avaliar a precisão das segmentações desconhecendo o método de segmentação utilizado. Os resultados das avaliações entre os métodos foram comparados utilizando teste T não-pareado com intervalo de confiança de 95% utilizando o Excel para Microsoft 365. Este método foi empregado para avaliar se havia diferença entre as médias de avaliação dos dois grupos.

3.4. EXPERIMENTO 2 – MODELOS DE MÚLTIPLAS ESTRUTURAS

No segundo experimento, um novo modelo de segmentação simultânea de múltiplas estruturas foi desenvolvido, utilizando um maior conjunto de dados associado à ampliação da quantidade de estruturas anatômicas segmentadas.

3.4.1. Base de dados

Neste experimento, o conjunto de dados foi expandido para incluir um total de 325 TC do osso temporal, o que representa um acréscimo de 175 novas TC em relação ao Experimento 1. O critério de exclusão utilizado anteriormente foi mantido, descartando imagens de TC que apresentassem artefatos metálicos significativos ou movimentos que impedissem a visualização clara das estruturas. Da mesma forma, seguimos com a inclusão de estudos provenientes de diversos aparelhos, desde que tivessem uma resolução adequada (com espaçamento de até 0,3mm). Esse critério é semelhante ao aplicado no primeiro experimento.

Apesar das segmentações já terem sido realizadas nos estudos do primeiro experimento, optamos por realizar novamente as segmentações para todo o conjunto de dados para este experimento.

3.4.2. Estruturas anatômicas selecionadas para segmentação

3.4.2.1. Artéria Carótida interna

A ACI foi segmentada na faixa de intensidade de partes moles, desde a sua entrada no canal carotídeo até o segmento cavernoso. Nesta região, a variação interpessoal de curvatura é grande, e a TC sem contraste não permite a identificação clara das diferentes tortuosidades.

3.4.2.2. Nervo da corda do tímpano (NCoT)

O NCoT foi segmentado a partir da sua emergência do nervo facial até a sua entrada na orelha média, onde não é mais possível sua identificação em TC clínicas.

Neste caso, devido ao pequeno volume do NCoT e seus arredores muitas vezes composto de osso compacto, a faixa de intensidade de segmentação foi de -200 até 800HU.

3.4.2.3. Conduto auditivo externo

O CAE foi delineado desde o poro acústico externo até o anel timpânico. Lateralmente o CAE foi limitado por um plano que tangencia a superfície do osso temporal, e medialmente pelo anel timpânico e o cabo do martelo. A faixa de -1000 a 300 HU foi definida para esta estrutura.

3.4.2.4. Conduto auditivo interno

O limite medial do CAI é definido pelo poro acústico interno, definido por um plano que tangencia o osso temporal na face da fossa posterior. O limite medial é o fundo, anatomicamente o local da emergência dos nervos facial e vestíbulo-cocleares para gânglio geniculado e labirinto, respectivamente.

3.4.2.5. Nervo facial

O NF foi segmentado de maneira similar ao Experimento 1, desde a emergência do NF pelo CAI, incluindo o gânglio geniculado, e os segmentos timpânico e mastoideos.

3.4.2.6. Orelha interna

Conforme o Experimento 1, dentro da faixa de tecido de densidade de partes moles, o conteúdo da orelha interna foi segmentado.

3.4.2.7. Ossículos

O estribo, martelo e bigorna, quando existentes, foram segmentados dentro da faixa de tecido ósseo na TC.

3.4.2.8. Seio sigmoide

O seio sigmoide foi segmentado conforme o experimento 1, desde a transição com o seio transverso, até o forame jugular.

3.4.2.9. Cápsula ótica

A segmentação de referência da CO foi implementada de acordo com a descrição do Experimento 1, e refinamento manual foi realizado em cada exame de TC para remover eventuais falsos positivos.

3.4.3. Segmentação de referência

Todos os volumes de TC e seus respectivos mapas de rótulos das estruturas utilizados no primeiro estudo foram minuciosamente revisados e as novas estruturas segmentadas. Neste experimento, para algumas estruturas como a OI e SS, foram desenvolvidos modelos assistentes de segmentação, o que reduziu o tempo de processamento manual, porém sem prescindir de avaliação manual pormenorizada.

Estes modelos assistentes são modelos de segmentação semi-automatizada, em que pontos nas extremidades das estruturas são marcados pelo usuário, e o modelo conduz a segmentação da estrutura anatômica, que é então refinada pelo usuário. Os modelos assistentes também são construídos reduzem o tempo de segmentação completamente manual,

3.4.4. Pré-processamento

A uniformização dos dados ocorreu de forma semelhante ao primeiro experimento, embora com algumas modificações para a adequação ao novo algoritmo utilizado.

3.4.4.1. Adequação da dimensão e espaçamento das tomografias para um espaço isotrópico de 0,25mm;

3.4.4.2. Normalização da intensidade

A padronização da intensidade das imagens entre -500 a 1800 HU foi conduzida. A modificação da extensão da faixa de intensidade correspondente a osso (de 2000 para 1800HU) foi realizada para aumentar o gradiente entre as partes moles e permitir ao algoritmo uma melhor identificação das diferentes estruturas.

3.4.4.3. Aumento de dados

No experimento 2, técnicas para aumento de dados foram implementadas como no Experimento 1, com a adição de outras descritas a seguir.

- Variação da intensidade da imagem em ±10%

- Inversão da lateralidade das imagens: A inversão da lateralidade da TC (direita ou esquerda) foi realizada aleatoriamente na probabilidade de 20% a cada época.

- Rotação dos eixos do volume tomográfico em ±15°
  Foi introduzido um fator aleatório de rotação das TCs em ±15° para simular variações da posição da cabeça durante o exame

- Ampliação em ±10%
  Neste experimento, os volumes e respectivos mapa de rótulos (segmentações de referência) das estruturas foram ampliados aleatoriamente na faixa de ±10% do tamanho original com o objetivo de aumentar a variação do tamanho das estruturas no treinamento do algoritmo.

- Adição de ruído Gaussiano
  Durante o processo de treinamento, a adição aleatória de ruído gaussiano às TC foi realizada para introdução de variações nas imagens.

3.4.5. Divisão do conjunto de dados

As imagens foram aleatoriamente divididas entre os grupos de treinamento e de teste, sendo 265 imagens no grupo de treinamento e 60 imagens no grupo de teste.

3.4.6. Arquiteturas

Neste experimento, adaptamos o algoritmo SwinUNETRpara a construção do nosso modelo. Trata-se de um algoritmo híbrido, que combina características das CNN com o Swin Transformer, é um modelo baseado em *transformers*...

3.4.7. Treinamento do algoritmo

O servidor de treinamento com 24GB de memória de vídeo (NVIDIA RTX4090) foi adaptado para a tarefa com a incorporação da plataforma MONAI. (CARDOSO *et al.*, 2022). Esta é uma evolução da plataforma Clara SDK, com modificações estruturais que aprimoram o manejo dos dados e permitem maior flexibilidade ao treinamento de algoritmos de aprendizado profundo para medicina.

O treinamento de múltiplas estruturas (um modelo de predição para todas as estruturas simultaneamente em uma única análise) foi conduzido por 1000 épocas (ciclos de análise de todas as imagens do conjunto de treinamento), com taxa de aprendizado (*learning rate*) inicial de 2×10^{-4}, com algoritmo Adam de otimização.

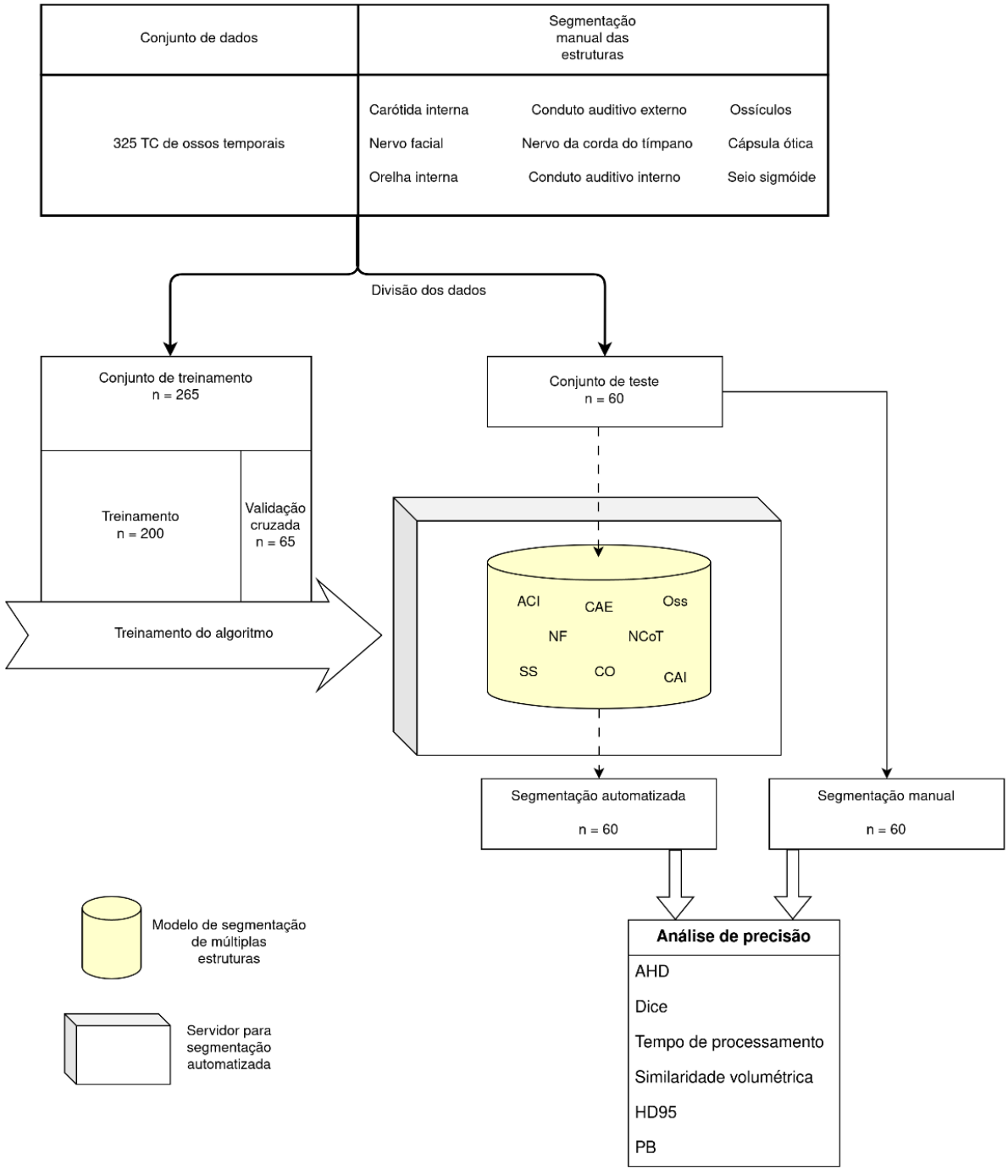
3.4.8. Servidor de predições

Da mesma forma que o Experimento 1, um servidor de internet que permite a segmentação automatizada de maneira remota foi implementado a partir das ferramentas da plataforma MONAI. Este servidor foi utilizado para a predição do conjunto de testes. O pós-processamento das segmentações incluiu apenas a remoção automática de pequenas ilhas de segmentação falsos positivos das estruturas. Este processamento não foi implementado para a segmentação do NCoT, pois se trata de estrutura de pequeno volume e isto poderia excluir verdadeiros positivos.

3.4.9. Avaliação objetiva

A avaliação objetiva foi conduzida utilizando o coeficiente de similaridade de Dice, o cálculo da precisão balanceada, a similaridade volumétrica, a distância de Hausdorff média, e o percentil 95º da distância de Hausdorff do (HD95). O tempo para segmentação automatizada foi medido para cada tomografia, e representa o tempo de processamento total para aquele volume por se tratar de um modelo de segmentação de múltiplas estruturas simultaneamente. O diagrama do estudo é representado na figura 21.

Figura 21 - Diagrama do experimento 2 (modelo de múltiplas estruturas)



Fonte: elaborado pelo autor (2023)

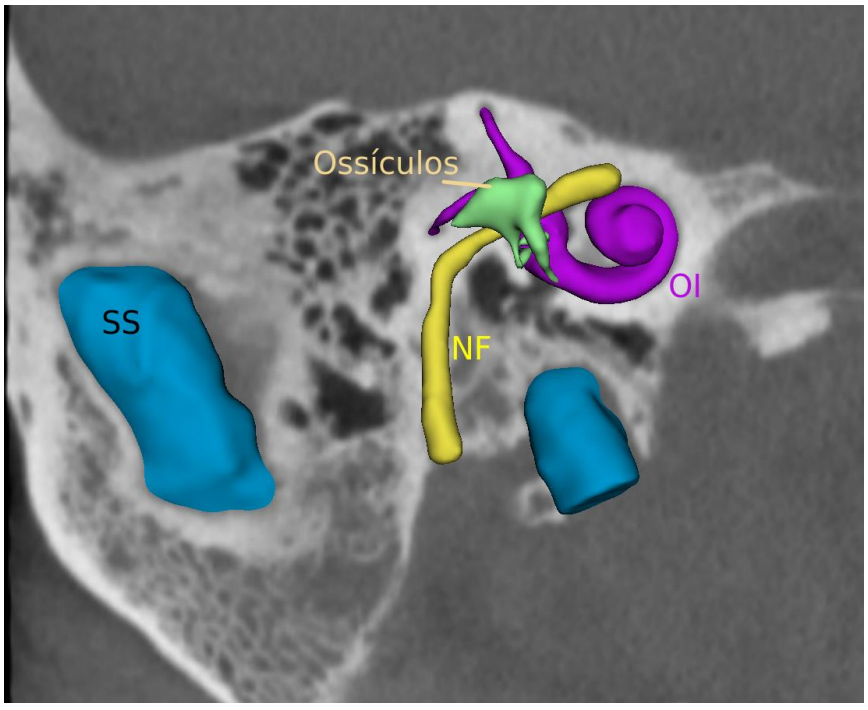
4 RESULTADOS

Entre julho de 2019 e janeiro de 2023, 325 TCs de osso temporal foram selecionadas para análise e processamento. Dois experimentos foram realizados neste período. O primeiro foi realizado com o total de 150 TCs entre janeiro e maio de 2020, para a segmentação de 5 estruturas anatômicas. Já o segundo experimento contou com o total de 325 TC (foram acrescidas 175 TCs às 150 TCs do experimento 1) , para a segmentação de 9 estruturas anatômicas do osso temporal.

4.1. EXPERIMENTO 1 – MODELOS DE ESTRUTURAS ÚNICAS

No conjunto de 150 TCs de ossos temporais, a OI, o NF, os Ossículos e o SS foram manualmente segmentados pelo pesquisador, como no exemplo da figura 22.

Figura 22 – Segmentação manual de estruturas do osso temporal. Renderização 3D sobre TC



Fonte: elaborado pelo autor (2023)

A avaliação clínica do conjunto de dados mostrou exames normais em 69% (104/150), alterações pós-operatórias em 9% (14/150), achados inflamatórios ou

relacionados à ventilação inadequada da orelha média em 11% (17/150) e deiscência do canal semicircular superior foi identificada em 3% (4/150). (Quadro 1)

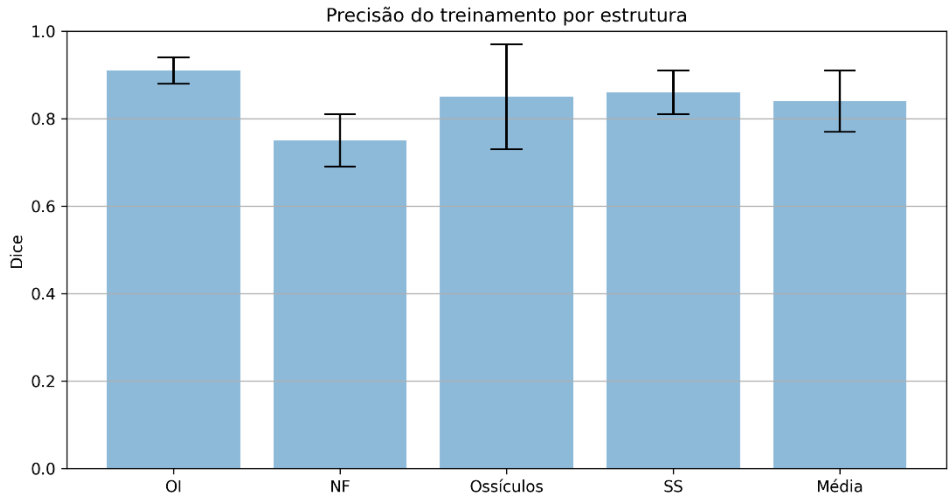
Quadro 1 - Achados radiológicos do conjunto de dados do Exp. 1

Normal		**104**	**69%**
Focos de otosclerose		15	**5%**
Prótese estapedotomia		4	**3%**
Ossiculoplastia		4	**3%**
Mastoidectomia		5	**3%**
Implante coclear		1	**1%**
Total pós-operatório		14	9%
Opacificação da orelha média		9	**6%**
Opacificação de células da mastoide		5	**3%**
Esclerose da mastoide		7	**5%**
Total de achados inflamatórios *		17	11%
Deiscência de CSS		4	**3%**
Outros		4	**3%**

Fonte: elaborado pelo autor (2020)
Legenda: CSS – Canal semicircular superior

A precisão da predição do modelo, medida pelo coeficiente de similaridade de Dice no conjunto de validação cruzada, foi de 0,86 ± 0,08 (Dice ± desvio padrão, DP) para a OI; 0,77 ± 0,11 para o NF; 0,84 ± 0,07 para os Ossículos e 0,86 ± 0,09 para o SS.(Figura 23)

Figura 23 – Gráfico do coeficiente de Dice por estrutura ao final do treinamento dos modelos do experimento 1, medidos em validação cruzada.



Fonte: elaborado pelo autor (2023).

Usando o processo de predição automatizada no conjunto de testes (n = 25), os resultados foram de Dice de 0,91 ± 0,03 (Dice ± desvio padrão, DP) para OI; 0,75 ± 0,06 para o nervo facial; 0,85 ± 0,12 para os ossículos e 0,86 ± 0,05 para o seio sigmoide. (Tabela1)

Neste experimento, foi calculada a distância de Hausdorff média (AHD), cujo resultado foi de 0,15 mm para a OI, 0,21mm para o nervo facial, 0,24mm para a os ossículos e de 0,45 mm para o seio sigmoide. A similaridade volumétrica entre as segmentações manuais e automatizadas foi de 108,3% ± 13,3% para OI; 101,2 ± 24,3 % para o NF; 99,4 ± 0,31 para os ossículos e 96,3 ± 18,9 para o SS. (Tabela 1).

O processo de segmentação automatizada proposto no Exp. 1 durou em média 2,7 segundos por estrutura analisada, enquanto a CO foi processada e segmentada em 2,2 segundos. Em comparação, a segmentação manual de cada estrutura foi realizada em 211,1 segundos, em média, conforme tabela 1.

Tabela 1 - Resultados da análise objetiva do Exp. 1. do conjunto de testes (n=25)

		Orelha interna	*Ossículos*	*Nervo facial*	*Seio sigmoide*
Dice	ResNet (DP[19])	0.91 (0.03)	0.87 (0.04)	0.69 (0.11)	0.85 (0.04)
	U-Net (DP)	0.91 (0.04)	0.86 (0.06)	0.73 (0.07)	0.81 (0.05)
	AH-Net (DP)	0.91 (0.03)	0.85 (0.12)	0.75 (0.06)	0.86 (0.05)
Tempo para segmentação (s)	Manual (DP)	224.2 (54.6)	110.3 (19.2)	221.8 (59.1)	323.3 (100.0)
	ResNet (DP)	4.58 (0.52)	4.75 (0.62)	4.83 (0.56)	4.64 (0.55)
	U-Net (DP)	6.82 (0.74)	6.72 (0.69)	6.84 (0.77)	6.71 (0.71)
	AH-Net (DP)	2.61 (0.82)	2.70 (0.61)	2.73 (0.66)	2.65 (0.73)
Distância média de Hausdorff (mm)	ResNet (DP)	0.23 (0.18)	0.23 (0.18)	0.46 (0.42)	0.45 (0.15)
	U-Net (DP)	0.25 (0.21)	0.22 (0.16)	0.38 (0.20)	0.62 (0.21)
	AH-Net (DP)	0.25 (0.24)	0.23 (0.14)	0.24 (0.19)	0.45 (0.16)
Similaridade Volumétrica (%)	ResNet (DP)	104.8 (23.0)	101.2 (14.2)	108.7 (34.8)	104.7 (14.7)
	U-Net (DP)	108.5 (13.2)	90.1 (11.9)	105.0 (32.9)	100.3 (25.1)
	AH-Net (DP)	108.3 (13.3)	99.4 (30.9)	101.2 (24.3)	96.3 (18.9)

Fonte: elaborado pelo autor (2020)

As médias das pontuações dos revisores para todas as estruturas-chave apresentaram similaridade (4,2 comparado com 4,3, p=91) independente da forma de segmentação (manual ou automatizada), conforme tabela 2. Separadamente, nenhuma estrutura apresentou diferença estatística entre os métodos avaliados.

[19] DP: Desvio padrão

Tabela 2 - Resultado da análise subjetiva por especialistas de conjunto de dados do Exp. 1

	Método	***Manual***		***Automatizado***		
Estrutura		Média de pontuação (1-5)	DP[20]	Média de pontuação (1-5)	DP	Teste T
Cápsula ótica		4.3	0.9	4.3	0.5	1
Orelha interna		3.8	0.9	4.1	0.9	0.37
Ossículos		3.8	1	4	1	0.55
Nervo facial		4.6	0.5	4.4	0.5	0.44
Seio sigmoide		4.8	0.5	4.5	0.7	0.17
Média		4.2		4.3		0.91

Fonte: elaborado pelo autor (2020)

[20] DP: Desvio padrão

4.2. EXPERIMENTO 2 – MODELOS DE MÚLTIPLAS ESTRUTURAS

Neste experimento, novas estruturas-chave foram adicionadas à tarefa de segmentação por aprendizado profundo, incluindo a ACI, o NCoT, o CAE, o CAI e a CO. Todas as nove estruturas do conjunto de dados (n=325) foram manualmente segmentadas pelo pesquisador utilizando a plataforma 3D Slicer. Todas as segmentações foram revisadas antes do início do treinamento para garantir a precisão anatômica das segmentações manuais.

Os achados clínico-radiológicos estão expostos no quadro 2.

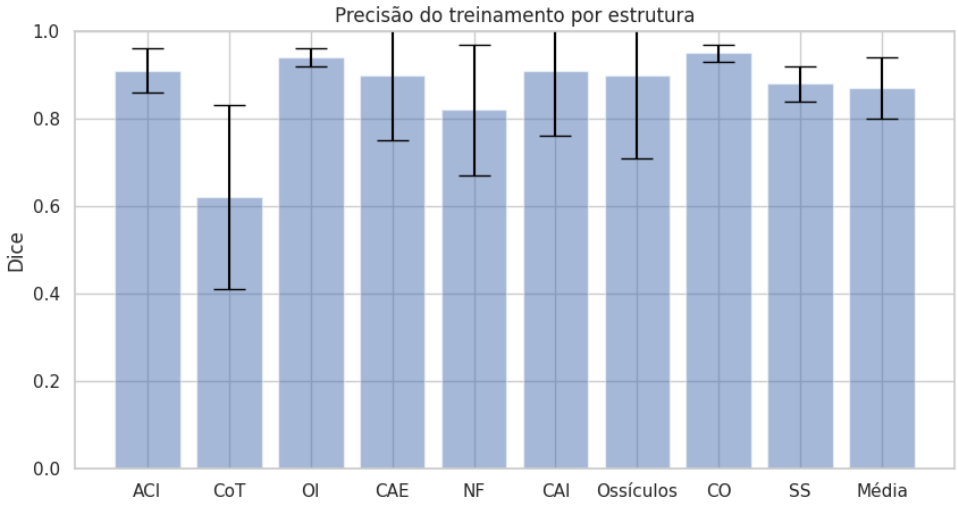
Quadro 2- Achados radiológicos do conjunto de dados do Exp. 2

Normal		228	70%
Otosclerose		18	6%
Prótese estapedotomia		3	1%
Prótese de ossiculoplastia		5	2%
Mastoidectomia		18	6%
Implante coclear		1	0%
Total pós-operatório *		24	7%
Opacificação da orelha média		22	7%
Opacificação de células da mastoide		31	10%
Esclerose da mastoide		19	6%
Total de achados inflamatórios *		40	13%
Deiscência de CSS		18	6%
Outros		11	3%

Fonte: elaborado pelo autor (2023)
Legenda: CSS – Canal semicircular superior

Ao final do treinamento do algoritmo, a precisão do modelo alcançou Dice de 0,87 para todas as estruturas em conjunto. Para cada estrutura, a precisão do modelo ao final do treinamento foi de ACI (0.91, 0.05) (Dice, DP); NCoT (0.62, 0.21); OI (0.94, 0.02); CAE (0.90, 0.15); FN (0.82, 0.15); CAI (0.91, 0.15); Ossículos (0.90, 0.19); CO (0.95, 0.02); SS (0.88, 0.04), demonstrado na figura 24.

Figura 24 - Gráfico do coeficiente de Dice por estrutura ao final do treinamento dos modelos do experimento 2, medidos em validação cruzada.



Fonte: elaborado pelo autor (2023)

Os resultados da análise objetiva das segmentações do conjunto de teste realizado de forma automatizada pelo modelo de predição construído pelo treinamento do algoritmo, comparado com a segmentação manual realizado por especialista estão representados na tabela 3.

Tabela 3 - Resultado da análise objetiva do conjunto de teste (n = 60/325) do Exp. 2 para as diferentes estruturas

Estrutura	***ACI***	***Nervo da corda do tímpano***	***Orelha interna***	***CAE***	***Nervo facial***	***CAI***	***Ossículos***	***Cápsula ótica***	***Seio Sigmoide***	***Média***
Dice (DP)	0.90 (0.06)	0,59 (0,18)	0,95 (0,02)	0,90 (0,10)	0,83 (0,06)	0,93 (0,03)	0,88 (0,15)	0,95 (0,03)	0,86 (0,07)	0,87 (0,08)
AHD[21](DP) mm	0,28 (0,22)	0,41 (0,89)	0,06 (0,05)	0,29 (0,43)	0,14 (0,05)	0,13 (0,05)	0,17 (0,46)	0,07 (0,03)	0,46 (0,34)	0,22 (0,28)
HD95[22] (DP) mm	1,03 (0,95)	1,06 (1,75)	0,32 (0.41)	1,05 (1,05)	0,46 (0,22)	0,52 (0,26)	0,50 (0,82)	0,26 (0,03)	1,91 (1.47)	0,79 (0,77)
PB[23]	0,96 (0,05)	0,83 (0,10)	0,98 (0,02)	0,94 (0,06)	0,95 (0,03)	0,97 (0,03)	0,96 (0,04)	0,98 (0,03)	0,92 (0,06)	0,94 (0,05)
Sim. Volum.[24] (%)	102 (14,8)	175 (328)	101 (6,3)	94,5 (14,3)	118 (23)	100 (9,9)	122 (79)	102 (6,9)	95 (17,6)	112 (56)

Fonte: Elaborado pelo autor (2023)

O modelo de estruturas múltiplas processou as tomografias do conjunto de teste em um tempo médio de 9,1 segundos por exame, e um exemplo do resultado da segmentação automatizada pode ser visto na figura 25.

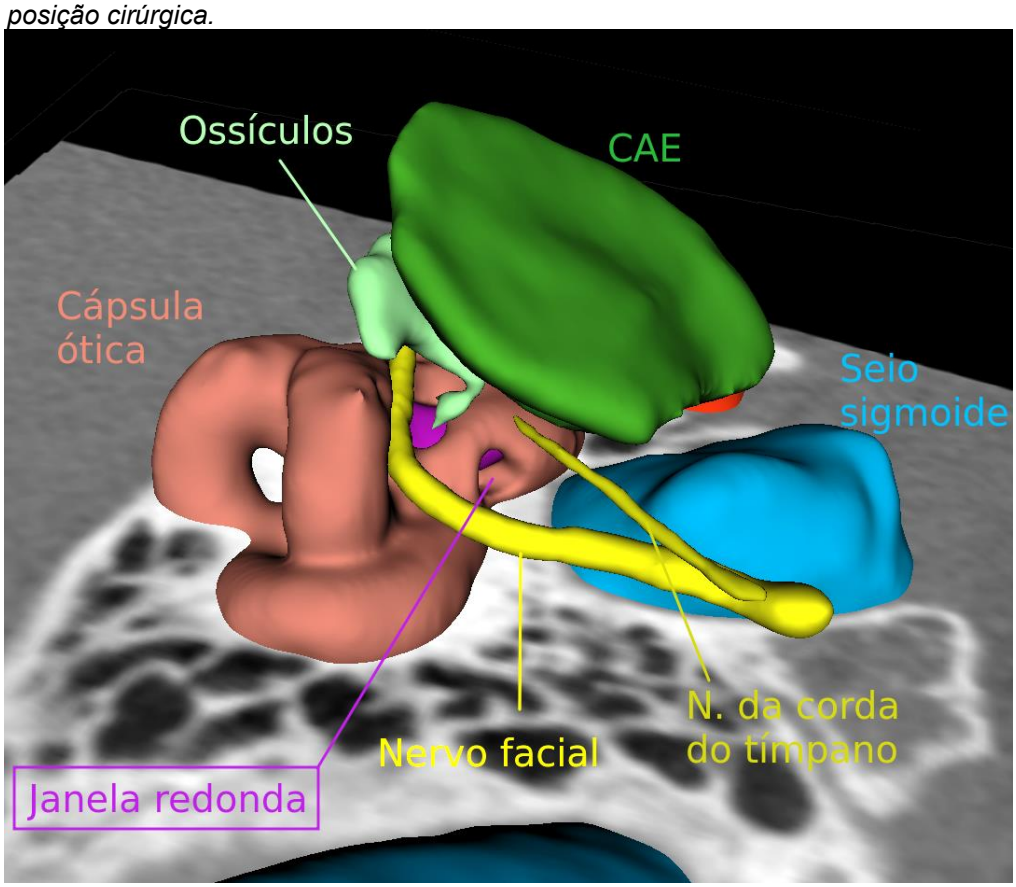
[21] AHD: Distância de Hausdorff média

[22] HD95: Percentil 95º da distância de Hausdorff

[23] PB: Precisão balanceada

[24] Similaridade volumétrica

Figura 25 - Renderização de estruturas do osso temporal segmentadas pelo modelo automatizado. Vista de posição cirúrgica.



Fonte: elaborado pelo autor (2023)

5 DISCUSSÃO

Este estudo descreve a construção de modelos para a segmentação automatizada de estruturas-chaves do osso temporal através da análise de tomografias por algoritmos de aprendizado profundo. O processo visa facilitar o preparo pré-operatório, assim como alavancar projetos de simulação cirúrgica e realidade mista.

Este trabalho, que apresenta dois experimentos em um contínuo de evolução, mostra que o aprimoramento dos algoritmos e o aumento do conjunto de testes foram associados a uma melhor performance dos modelos quando comparamos os resultados de Dice, que será exposto a seguir.

No experimento 1, os modelos treinados com as diversas arquiteturas demonstraram precisão semelhante em geral, e o modelo com a rede AH-Net se destacou pelo menor tempo de processamento. Este resultado pressupõe a eficiência aprimorada desta arquitetura para o trabalho, e esta característica pode facilitar o processamento de grandes volumes de dados. Para a predição da OI e dos Ossículos, porém, esta arquitetura mostrou um desempenho levemente menor quando comparada às outras, o que pode sugerir uma troca do equilíbrio entre precisão e velocidade. Independente desta diferença na avaliação objetiva, as segmentações provenientes das diferentes redes foram similares e com valores de Dice elevados no conjunto de testes.

No Experimento 2 o resultado alcançado pode ser considerado ainda mais promissor. Em primeiro lugar, a construção de um modelo de segmentação simultânea de múltiplas estruturas facilita a implementação do servidor de predições e demonstra tanto o aprimoramento das ferramentas de desenvolvimento quanto a capacidade de adaptá-las para este projeto. Em segundo lugar, a precisão do modelo foi ainda maior quando se comparam o resultado das mesmas estruturas entre os dois experimentos (Dice OI 0,91 em comparação com 0,95 no experimento 2; NF 0,75 contra 0,83; Ossículos 0,85 contra 0,88; e SS 0,86 contra 0,86). Este resultado pode ter contribuição do aumento do conjunto de dados, pois permite ao algoritmo exposição a uma maior variação de características das TC, aumentando a generalização e performance do modelo.

A utilização do algoritmo híbrido com a incorporação de *transformers* também pode ter contribuído para este resultado. Por fim, o tempo de inferência no modelo de múltiplas estruturas foi ainda menor que no experimento 1. A nova arquitetura de algoritmo somado à utilização de hardware mais eficiente são os fatores que devem ter contribuído para o resultado.

A comparação de desempenho entre o primeiro e segundo experimento pode ser ampliada na discussão da construção de modelos de estrutura única em contraste com modelos de múltiplas estruturas. No primeiro caso, as pré-transformações podem ser ajustadas para o treino de cada estrutura, na tentativa de favorecer a identificação das características específicas para um treinamento mais eficaz. Em contrapartida, diversos ciclos de treinamento são exigidos para alcançar um resultado satisfatório, considerando a necessidade de ajuste dos hiperparâmetros do algoritmo.

Já na construção do modelo de múltiplas estruturas, as pré-transformações devem ser genéricas, que abarquem a possibilidade de identificação de todas as estruturas, porém com a vantagem de treinar um modelo único, reduzindo a quantidade de ciclos de treinamento e ajuste de hiperparâmetros. Neste contexto, taxas de aprendizado, decaída e tamanho do lote (*batch size*) foram ajustados para apenas um modelo no Exp. 2, em comparação com quatro rodadas no Exp. 1, uma para cada estrutura.

No Experimento 2, o conjunto de dados foi ampliado pela segmentação manual de 9 estruturas anatômicas em um número consideravel maior de TC. Esta ampliação, realizada de maneira sistemática ao longo de meses, permitiu o treinamento do novo modelo com uma base de dados mais variada em termos de variações anatômicas. Para este experimento também já havia maior sedimentação de conhecimentos e estratégias a serem colocadas em prática que pode ter contribuído para a melhora do desempenho.

Neste sentido, a segmentação manual, a adaptação e a incorporação de pré-transformações e o ajuste dos hiperparâmetros foram mais eficientes no Experimento 2, em comparação com o Experimento 1, no qual a implementação foi o resultado de um longo processo de aprendizado. No segundo experimento, o uso de pré-transformações de aumento de dados foi ampliado, com a incorporação de mais técnicas de deformação e adição de ruído às imagens, objetivando aumentar a variabilidade do conjunto de dados.

O uso de hardware moderno também pode contribuir para o aprimoramento dos modelos de predição anatômica. Embora as configurações de *hardware* de ambos os experimentos contemplem a mesma quantidade de memória da placa de vídeo (24GB), os *hardwares* diferem em modelo e versão. No Exp. 1, duas placas NVIDIA TitanXP em paralelo somam 7680 núcleos CUDA (unidades de processamento), enquanto no Exp. 2, uma NVIDIA RTX 4090 possui 16384 núcleos CUDA e atualmente pontua como líder nos testes de comparação de GPUs (unidades de processamento gráfico), considerando o mercado consumidor. Associado a isto, o uso de algoritmos do estado da arte pode contribuir para o aprimoramento dos resultados.

O aprendizado de máquina, particularmente as técnicas de aprendizado profundo, têm demonstrado sistematicamente resultados precisos e promissores na anotação automática de estruturas anatômicas a partir de imagens clínicas. Isto porque é capaz de extrair padrões que nem sempre são facilmente aparentes à visão humana. A evolução dos algoritmos, como o desenvolvimento dos *transformers,* reforça a importância destas técnicas, que ampliam a capacidade de análise de grande quantidade de dados e possibilitam obter informações importantes para a prática clínica.

O uso dos *transformers* em aplicações de aprendizado profundo tem se expandido nos últimos anos pelo grande impacto de desempenho que esta técnica proporcionou no desempenho dos grandes modelos de linguagem. (BROWN *et al.*, 2020; LEWKOWYCZ *et al.*, 2022; VASWANI *et al.*, 2017). Desde então, variantes têm sido desenvolvidas e testadas em áreas como a visão de computador. Esta arquitetura se caracteriza por aprender a relação entre as diferentes partes do objeto (palavras, frases e imagens), mesmo que distantes, como em longos textos ou grandes imagens. A capacidade de atenção é a característica principal deste tipo de algoritmo.

Modelos baseados em *transformers* desenvolvidos especificamente para analisar imagens, como o Swin Transformer (LIU *et al.*, 2021), dividem as imagens de várias formas, e podem aprender características delas em várias escalas. Estes mecanismos podem favorecer o treinamento de um modelo de múltiplas estruturas de melhor performance, pois tanto as características das estruturas são aprendidas em vários níveis como a inter-relação entre as estruturas são incorporadas ao modelo. Nesta tese adaptamos um algoritmo híbrido, chamado SwinUNETR, que combina

transformers com CNN, que pode associar os benefícios das diferentes arquiteturas para a construção do modelo.

Estes fatores em conjunto podem ter contribuído para o aumento da velocidade de inferência nos dois experimentos em sequência, superando em muitas vezes o tempo de segmentação anatômica manual realizada por um especialista treinado. A segmentação manual exigiu atenção focada por uma média de 211s para cada estrutura, enquanto o método automatizado levou apenas 2,7 s por estrutura no Exp. 1 e 9,1s para as nove estruturas simultaneamente no Exp. 2. Este resultado é particularmente importante no potencial do uso clínico desta tecnologia.

A segmentação automatizada de forma rápida (redução de 200 vezes no tempo de segmentação) pode possibilitar o uso destes modelos em grande escala, e permitir a integração desta abordagem em sistemas de simulação, planejamento e navegação cirúrgicos, fomentando pesquisadores e cirurgiões. (WON *et al.*, 2018, 2019).

O osso compacto que envolve a OI é denominado cápsula ótica. Esta estrutura possui um aspecto mais radiodenso quando comparado com o restante do osso temporal. Entretanto, a diferenciação e segmentação da cápsula ótica não é trivial na prática clínica das TC de osso temporal, a despeito da possível manipulação do contraste (níveis de exposição da TC). Dentre as estruturas-chave, a cápsula ótica apresentou um resultado de precisão elevado e pode prover informações importantes no planejamento pré-operatório.

No primeiro experimento utilizamos a técnica de propagação da cápsula ótica a partir da OI conforme descrito por Neves et al. (NEVES *et al.*, 2021). Esta abordagem expande a cápsula ótica pelos voxels com densidade de osso compacto que são adjacentes à OI. A técnica produziu resultados visuais satisfatórios naquele experimento, e produziu informações espaciais valiosas da cápsula ótica.

No Experimento 2, a CO foi incluída como estrutura-chave no treinamento do modelo de estruturas múltiplas, e este alcançou um ótimo resultado no conjunto de testes para esta estrutura. A menor variabilidade anatômica tridimensional associada à marcada fronteira da CO com a orelha interna devem ter contribuído para a boa performance de segmentação desta estrutura.

A importância da segmentação da CO inclui a identificação de referências anatômicas importantes desta estrutura como a janela redonda, que pode aprimorar o planejamento pré-operatório (NEVES *et al.*, 2022) e o desenvolvimento da cirurgia

de implante coclear assistida por robô. (CAVERSACCIO *et al.*, 2022; MUELLER *et al.*, 2021). Outra vantagem do método é a capacidade de identificar e avaliar defeitos patológicos da cápsula ótica, como os que ocorrem na deiscência do canal semicircular superior. A identificação e quantificação destes defeitos na CO pode ajudar na seleção entre os acessos via fossa média ou transmastoideo para procedimentos de fechamento destas deiscências.

Nos dois experimentos observamos a uniformidade da precisão dos modelos de predição no conjunto de treinamento e de teste, conforme descrito nas figuras Isto sugere que os modelos construídos não estão erroneamente sobreajustados aos dados, e que generalizam bem a sua precisão. A eventual ampliação do conjunto de testes pode fornecer dados sobre quais alterações, variações anatômicas e outras características devem ser incorporadas para o refinamento do modelo. Neste estudo foi implementado um mecanismo de predição e pós-processamento de novas tomografias na plataforma 3DSlicer, que infere a segmentação da estrutura, maximiza seus contornos (Exp. 1) e remove voxels estranhos (falsos positivos). A etapa de pós-processamento tem como objetivo aprimorar e refinar as predições realizadas pelo modelo, e deve ser adaptada individualmente para determinada aplicação. (LITJENS *et al.*, 2017).

Neste estudo obtivemos resultados encorajadores na avaliação objetiva das segmentações automatizadas, comparáveis e em muitos casos superiores comparados com a literatura atual. (DING *et al.*, 2023; KE *et al.*, 2023). É provável que a segmentação automatizada da orelha interna seja facilitada pelo fato de ser uma estrutura cheia de fluido quase inteiramente cercada pelo osso radiodenso da cápsula ótica, proporcionando um contraste consistente com seus arredores. Isto não ocorre com o nervo facial, que tem múltiplas interfaces com tecidos moles, ar e osso heterogêneo ao longo de seu trajeto tortuoso pelo osso temporal. Isto, juntamente com seu longo caminho e pequeno volume, provavelmente reduz o valor do resultado da análise objetiva desta estrutura.

Ao serem comparados com os resultados da literatura atual acerca do assunto, observamos a grande vantagem do modelo desenvolvido neste trabalho quanto à quantidade de estruturas analisadas simultaneamente e sobretudo no tamanho da amostra. Com uma amostra maior, é de se esperar uma maior variedade de achados radiológicos no treinamento do algoritmo, e supor uma maior robustez e generalização

do modelo. A tabela 4 situa o resultado deste trabalho entre estudos comparáveis publicados até junho de 2023.

Tabela 4 – Posicionamento dos achados deste trabalho de segmentação anatômica do osso temporal com estudos similares

Autor	Ano	Método	Tipo de imagem	Entrada	Amostra	Carotida interna	NCoT	Orelha interna	Conduto auditivo externo	Nervo facial	Conduto auditivo interno	Ossículos	Cápsula ótica	Seio sigmóide
Fauser et al.	2019	Híbrido AP (2D UNet) + regularização de forma	Clínico	n/a	24	0.84	0.39	0,85	0.61	0.63	0.68	0.79	-	0.38
Heutink et al.	2020	AP (2D Unet)	Clínico alta resolução	150x150	123	-		0.9	-	-	-	-	-	-
Ke et al.	2020	AP (3D W-Net)	Clínico	64x64x80	30	-	-	0.71	-	0.49	-	0.64	-	-
Li et al.	2020	AP(UNet)	Clínico	48x48x48	64	-	-	0.83	-	-	0.81	0.82	-	-
Neves et al. (Exp. 1)	**2021**	**AP (3D AH-Net, 3D Unet, 3D ResNet)**	**Clínico**	**128x128x128**	**150**	**-**	**-**	**0.91**	**-**	**0.75**	**-**	**0.85**	**0.91**	**0.86**
Nikan et al.	2021	AP (PWD 3DNet)	MicroTC cadáver	144x144x144	39	0.81	-	0.9	-	0.74	0.89	0.85	-	0.86
Lv et al.	2021	AP (3D W-Net)	Clínico	64x64x80	30	-	-	0.9	-	0.77	-	0.85	-	-
Hussain et al.	2021	AP (2D Unet)	Micro TC cadáver	256x256	17	-	-	0.9	-	-	-	-	-	-
Neves et al.	2021	AP (3D ResNet)	Clínico	96x96x96	150	0.76	-		-	-	-	-	-	-
Wang et al.	2021	AP (3D W-Net)	Clínico	64x64x80	58	-	-	0.91	-	0.7	-	0.86	-	-
Zhenhua et al.	2023	AP (UneTr)	Clínico	64x64x32	92	-	-	0.92	-	-	-	-	-	-
Ding et al.	2023	AP (3D Unet)	Clínico	não descrito	15	0.92	0.61	-	0.84	0.86	0.91	-	0.95	-
Ke et al.	2023	AP (W-Net)	Clínico	64x64x80	80	0.87	-	0.91	0.86	0.75	0.85	0.89	-	-
Zhou et al.	2023	AP (UneTr)	Clínico	48x48x48	147	-	-	0.93	-	0.69	0.88	0.81	-	-
Neves et al. (Exp. 2)	**2023**	**AP (SwinUneTr)**	**Clínico**	**96x96x96**	**325**	**0.90**	**0.59**	**0.95**	**0.90**	**0.83**	**0.93**	**0.88**	**0.95**	**0.86**

Fonte: elaborado pelo autor (2023)
Legenda: AP - Aprendizado profundo

Os volumes das estruturas foram semelhantes entre os métodos de segmentação do conjunto de testes. A similaridade volumétrica foi maior para a CO e mostrou-se alta para a OI, CAE e SS. A maior variabilidade foi observada na segmentação do NCoT, para o qual o coeficiente de Dice apresentou um resultado modesto de precisão. Contudo, a métrica de similaridade volumétrica é sensível a extremos de volume, e tende a apresentar valores mais expressivos em pequenos volumes, como o caso do NCoT.

A proposta de segmentação do nervo corda do tímpano merece destaque pela pequena magnitude volumétrica e grande significado no planejamento de cirurgias otológicas como o implante coclear. Isto tem particular importância no desenvolvimento de sistemas de análise automatizada de trajetória. (CAVERSACCIO *et al.*, 2022; MUELLER *et al.*, 2021; TOPSAKAL *et al.*, 2020, 2022). O NCoT é o limite lateral do recesso do facial, área de acesso à janela redonda em cirurgias de implante coclear por acesso transmastoideo. Em TC com o recesso do facial pouco aerado, o NCoT pode ser mais facilmente identificado nos cortes coronais e sagitais entre a sua emergência do NF até a orelha média. O modelo de segmentação automatizado desenvolvido no Exp. 2, apesar da modesta precisão média, tem potencial promissor de possibilitar meios de identificação automatizada do recesso do facial e impulsionar projetos de automação cirúrgica.

Nos experimentos, os valores da distância de Hausdorff indicaram erros mínimos para as estruturas analisadas, e destacam o potencial deste modelo para projetos de desenvolvimento de navegadores cirúrgicos aprimorados com segmentação anatômica.

Na avaliação subjetiva por especialistas, cegos ao método utilizado, realizado do Exp.1, as segmentações manuais e automatizadas foram classificadas como altamente precisas (4,3/5). O valor atribuído às segmentações manuais e automatizadas das estruturas-chave foram semelhantes e sem diferença estatística significativa no teste T. Isto sugere que as segmentações automatizadas do método proposto podem alcançar bons resultados e apoia a potencial translação deste sistema para aplicações clínicas e de pesquisa, antes limitados pela necessidade de segmentação manual.

Neste sentido, é razoável supor que este sistema seja uma base para a ampliação da base de dados de TC com estruturas anatômicas segmentadas, que

após verificação e refinamento, podem ser incorporados a novas etapas de treinamento, e com isso, aumentar a robustez e precisão dos modelos de segmentação. A incorporação de outras modalidades de imagem como ressonância magnética e exames contrastados pode ampliar as aplicações e fornecer novas ferramentas para uso clínico e de pesquisa. (NEVES *et al.*, 2023).

Entre as limitações do nosso estudo podemos identificar o tamanho relativamente pequeno da amostra, mesmo com a inclusão de novos conjuntos de dados anotados no segundo experimento. Apesar deste trabalho apresentar, até onde o conhecimento do autor alcança, a maior amostra de dados na construção de um modelo de segmentação de estruturas do osso temporal por algoritmos de aprendizado profundo, a quantidade de TC no conjunto de dados ainda pode limitar a generalização dos resultados. A necessidade de segmentação manual responde por esta limitação, e é esperado que o próprio sistema auxilie na anotação de dados novos.

Outra limitação deste estudo é a presença de quantidade relativamente alta de exames com anatomia normal e que as TC são provenientes de uma única instituição. Isto poderia contribuir para uma menor robustez do modelo, mas há que se ressaltar que as variações anatômicas interpessoais são numerosas e ampliam a complexidade do modelo, e que uma instituição de referência possui diversos modelos de aparelhos de aquisição, além de incorporar exames realizados em outros locais. De qualquer forma, a incorporação de novos casos alterados por doença e a ampliação do conjunto de dados adquiridos com equipamentos e protocolos diferentes podem aumentar a generalização do modelo.

A possibilidade da distribuição aleatória de orelhas de um mesmo indivíduo no grupo de treinamento e teste pode ter ocorrido e isto poderia reduzir a independência dos grupos. Por outro lado, diversos estudos mostram haver diferença considerável entre as duas orelhas de um determinado indivíduo. (ASLAN *et al.*, 2001; SINGH *et al.*, 2019; VAN OSCH *et al.*, 2019). Então, consideramos que a interferência desta possível ocorrência não afete a generalização do modelo.

O presente estudo demonstra métodos de construção de modelos de segmentação automatizada por meio de algoritmos de aprendizado profundo com diversas vantagens em comparação a estudos anteriores, conforme discutido por Neves et al 2021:

- O sistema não requer nenhuma anotação do usuário que possa introduzir variabilidade. Isto se contrasta a métodos que utilizam estratégias semiautomatizadas, em que o usuário precisa apontar para um ou vários pontos da estrutura em questão para iniciar o processo de segmentação. (NOBLE *et al.*, 2008; POWELL *et al.*, 2017).
- A precisão das segmentações geradas por computador alcançou elevado índice de similaridade com as segmentações manuais realizadas por especialista treinado;
- O tempo de segmentação por computador é dezenas de vezes inferior comparado com a alternativa manual, o que o torna mais factível para o cenário clínico;
- A construção deste modelo a partir de técnicas e algoritmos do estado da arte na área de aprendizado profundo, sobre plataformas de código aberto e disponíveis gratuitamente, garantem o aprimoramento futuro e colaboração de outras equipes;
- Este modelo foi treinado com uso de TCs clínicas, em contraste com propostas de segmentação a partir de TCs de espécimes cadavéricos. (BARTLING *et al.*, 2021; NIKAN *et al.*, 2021; VAN OSCH *et al.*, 2019). O uso de exames de pacientes reais possibilita a construção de modelo voltado para o uso clínico e com a possibilidade de acréscimo de novos dados com diferentes achados.

Dessa forma os achados desta pesquisa se traduzem em uma nova e relevante ferramenta para a identificação e segmentação de estruturas do osso temporal e pode contribuir para impulsionar o desenvolvimento de modelos mais amplos e precisos.

6 CONCLUSÃO

Este trabalho constituiu-se no desenvolvimento e validação de um sistema de segmentação anatômica de estruturas do osso temporal em TC com o uso de técnicas de aprendizado profundo. O modelo de segmentação automatizado produzido apresentou elevado grau de precisão no conjunto de teste e robustez para a generalização em dados novos.

As etapas de construção da base de dados de TCs manualmente anotadas, adaptação e treinamento dos algoritmos foram concluídas e permitiram a implementação do sistema de segmentação. O desenvolvimento de servidor para acesso remoto e a utilização de métricas objetivas possibilitaram a verificação da precisão dos modelos construídos e demonstram o potencial do sistema para autoalimentação e refinamento.

Os dados obtidos destacam o potencial do uso de técnicas de aprendizado profundo como relevante ferramenta de segmentação anatômica de TCs do osso temporal, para o planejamento pré-operatório personalizado e o desenvolvimento de sistemas de navegação intraoperatório aprimorados.

7 REFERÊNCIAS

AHMAD, S Danial *et al.* A computer vision approach for analyzing label free leukocyte trafficking dynamics on a microvascular mimetic. **Frontiers in immunology**, Switzerland, v. 14, p. 1140395, 2023.

AMODEI, Dario *et al.* Concrete Problems in AI Safety. [*s. l.*], p. 1–29, 2016. Disponível em: http://arxiv.org/abs/1606.06565.

ANDERSEN, Steven Arild Wuyts *et al.* Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. **Laryngoscope**, [*s. l.*], v. 126, n. 2, p. E74–E79, 2016.

ANDERSEN, Steven Arild Wuyts *et al.* Learning curves of virtual mastoidectomy in distributed and massed practice. **JAMA Otolaryngology - Head and Neck Surgery**, [*s. l.*], v. 141, n. 10, p. 913–918, 2015.

ANDERSEN, Steven Arild Wuyts *et al.* Patient-specific Virtual Temporal Bone Simulation Based on Clinical Cone-beam Computed Tomography. **Laryngoscope**, [*s. l.*], v. 131, n. 8, p. 1855–1862, 2021.

ASLAN, A *et al.* Morphometric analysis of anatomical relationships of the facial nerve for mastoid surgery. **The Journal of laryngology and otology**, England, v. 115, n. 6, p. 447–449, 2001.

BARBER, Samuel R *et al.* Augmented Reality, Surgical Navigation, and 3D Printing for Transcanal Endoscopic Approach to the Petrous Apex. **OTO open**, [*s. l.*], v. 2, n. 4, p. 2473974X18804492, 2018.

BARTLING, Mandolin L. *et al.* Micro-CT of the human ossicular chain: Statistical shape modeling and implications for otologic surgery. **Journal of Anatomy**, [*s. l.*], v. 239, n. 4, p. 771–781, 2021.

BISHOP, Christopher. **Pattern Recognition and Machine Learning`**. [*S. l.*]: Springer New York, NY, 2006. *E-book*. Disponível em: https://link.springer.com/book/9780387310732.

BROCKMEYER, Phillipp; WIECHENS, Bernhard. The Role of Augmented Reality in the Advancement of Minimally Invasive Surgery Procedures : A Scoping Review. [*s. l.*], 2023.

BROWN, Tom B. *et al.* Language models are few-shot learners -- special version. **Advances in Neural Information Processing Systems**, [*s. l.*], v. 2020-Decem, n. NeurIPS, 2020.

BUTLER, Alexander J *et al.* Augmented reality in minimally invasive spine surgery: early efficiency and complications of percutaneous pedicle screw instrumentation. **The spine journal : official journal of the North American Spine Society**, United States, v. 23, n. 1, p. 27–33, 2023.

CARDOSO, M. Jorge *et al.* MONAI: An open-source framework for deep learning in healthcare. [*s. l.*], 2022. Disponível em: http://arxiv.org/abs/2211.02701.

CAVERSACCIO, Marco *et al.* Robotic Cochlear Implantation for Direct Cochlear Access. **Journal of Visualized Experiments**, [*s. l.*], v. 2022, n. 184, p. 1–10, 2022.

CHAN, Sonny *et al.* High-fidelity haptic and visual rendering for patient-specific simulation of temporal bone surgery. **Computer assisted surgery (Abingdon, England)**, England, v. 21, n. 1, p. 85–101, 2016.

CHEN, Hao *et al.* DCAN: Deep contour-aware networks for object instance segmentation from histology images. **Medical image analysis**, Netherlands, v. 36, p. 135–146, 2017.

CHEN, Jenny X. *et al.* Predicting Resident Competence for Otolaryngology Key Indicator Procedures. **Laryngoscope**, [*s. l.*], p. 1–5, 2023.

ÇIÇEK, Özgün *et al.* 3D U-net: Learning dense volumetric segmentation from sparse annotation. *In*: , 2016. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [*S. l.: s. n.*], 2016. p. 424–432.

DALY, M. J. *et al.* Fusion of intraoperative cone-beam CT and endoscopic video for image-guided procedures. **Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling**, [*s. l.*], v. 7625, p. 762503, 2010.

DE OLIVEIRA, Henrique Fernandes *et al.* A feasible, low-cost, reproducible lamb's head model for endoscopic sinus surgery training. **PLoS ONE**, [*s. l.*], v. 12, n. 6, p. 1–9, 2017.

DE OLIVEIRA, Henrique F. *et al.* Acquisition of endoscopic nasal surgery skills with a lamb's head model. **Brazilian Journal of Otorhinolaryngology**, [*s. l.*], v. 88, p. S119–S125, 2022.

DING, Andy S. *et al.* A Self-Configuring Deep Learning Network for Segmentation of Temporal Bone Anatomy in Cone-Beam CT Imaging. **Otolaryngology–Head and Neck Surgery**, [*s. l.*], 2023.

DING, Andy S. *et al.* Automated Registration-Based Temporal Bone Computed Tomography Segmentation for Applications in Neurotologic Surgery. **Otolaryngology - Head and Neck Surgery (United States)**, [*s. l.*], v. 167, n. 1, p. 133–140, 2022.

DOSOVITSKIY, Alexey *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. [*s. l.*], 2020. Disponível em: http://arxiv.org/abs/2010.11929.

DRAKE, Richard L. **Gray's: atlas de anatomia**. 1ªed. [*S. l.*]: Elsevier, 2011.

DUBUISSON, M -.; JAIN, A K. A modified Hausdorff distance for object matching. *In*: , 1994. **Proceedings of 12th International Conference on Pattern Recognition**. [*S. l.: s. n.*], 1994. p. 566–568 v.1.

ERICSSON, K Anders. Deliberate practice and acquisition of expert performance: a general overview. **Academic emergency medicine : official journal of the Society for Academic Emergency Medicine**, United States, v. 15, n. 11, p. 988–994, 2008.

ESTEVA, Andre *et al.* A guide to deep learning in healthcare. **Nature Medicine**, [*s. l.*], v. 25, n. 1, p. 24–29, 2019. Disponível em: https://doi.org/10.1038/s41591-018-0316-z.

ESTEVA, Andre *et al.* Deep learning-enabled medical computer vision. **npj Digital Medicine**, [*s. l.*], v. 4, n. 1, p. 1–9, 2021. Disponível em: http://dx.doi.org/10.1038/s41746-020-00376-2.

FAUSER, Johannes *et al.* Toward an automatic preoperative pipeline for image-guided temporal bone surgery. **International Journal of Computer Assisted Radiology and Surgery**, [*s. l.*], 2019. Disponível em: https://doi.org/10.1007/s11548-019-01937-x.

FEDOROV, Andriy *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. **Magnetic resonance imaging**, [*s. l.*], v. 30, n. 9, p. 1323–1341, 2012.

FRENDØ, Martin *et al.* Cochlear implant surgery: Learning curve in virtual reality simulation training and transfer of skills to a 3D-printed temporal bone–A prospective trial. **Cochlear Implants International**, [*s. l.*], v. 22, n. 6, p. 330–337, 2021. Disponível em: https://doi.org/10.1080/14670100.2021.1940629.

FRITHIOFF, Andreas *et al.* 3D - printing a cost - effective model for mastoidectomy training. [*s. l.*], p. 1–8, 2023.

FRITHIOFF, Andreas; SØRENSEN, Mads Sølvsten; ANDERSEN, Steven Arild Wuyts. European status on temporal bone training: a questionnaire study. **European Archives of Oto-Rhino-Laryngology**, [*s. l.*], v. 275, n. 2, p. 357–363, 2018. Disponível em: http://dx.doi.org/10.1007/s00405-017-4824-0.

GARE, Bradley M. *et al.* Multi-atlas segmentation of the facial nerve from clinical CT for virtual reality simulators. **International Journal of Computer Assisted Radiology and Surgery**, [*s. l.*], v. 15, n. 2, p. 259–267, 2020. Disponível em: https://doi.org/10.1007/s11548-019-02091-0.

GEORGE-JONES, Nicholas A. *et al.* Automated Detection of Vestibular Schwannoma Growth Using a Two-Dimensional U-Net Convolutional Neural Network. **Laryngoscope**, [*s. l.*], v. 131, n. 2, p. E619–E624, 2021.

GONZALEZ, Rafael C.; WOODS, Richard E. **Digital image processing**. 4ª Ediçãoed. [*S. l.: s. n.*], 2018.

GULSHAN, Varun *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. **JAMA**, [*s. l.*], v. 316, n. 22, p. 2402–2410, 2016. Disponível em: https://doi.org/10.1001/jama.2016.17216.

HADDAD, Alexander F.; AGHI, Manish K.; BUTOWSKI, Nicholas. Novel intraoperative strategies for enhancing tumor control: Future directions. **Neuro-Oncology**, [*s. l.*], v.

24, p. S25–S32, 2022.

HATAMIZADEH, Ali *et al.* Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*], v. 12962 LNCS, p. 272–284, 2022.

HE, Kaiming *et al.* Deep residual learning for image recognition. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, [*s. l.*], v. 2016-Decem, p. 770–778, 2016.

HEUTINK, Floris *et al.* Multi-Scale deep learning framework for cochlea localization, segmentation and analysis on clinical ultra-high-resolution CT images. **Computer Methods and Programs in Biomedicine**, [*s. l.*], v. 191, p. 105387, 2020. Disponível em: https://doi.org/10.1016/j.cmpb.2020.105387.

HUDSON, Thomas J. *et al.* Intrinsic Measures and Shape Analysis of the Intratemporal Facial Nerve. **Otology and Neurotology**, [*s. l.*], v. 41, n. 3, p. e378–e386, 2020.

HUSSAIN, Raabid *et al.* Automatic segmentation of inner ear on CT-scan using auto-context convolutional neural network. **Scientific Reports**, [*s. l.*], v. 11, n. 1, p. 1–10, 2021. Disponível em: https://doi.org/10.1038/s41598-021-83955-x.

JAMES, Joel *et al.* Simulation training in endoscopic skull base surgery: A scoping review. **World Journal of Otorhinolaryngology - Head and Neck Surgery**, [*s. l.*], v. 8, n. 1, p. 73–81, 2022.

JAVED, F M *et al.* High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. **Computers in Biology and Medicine**, [*s. l.*], v. 155, n. June 2022, p. 106646, 2023. Disponível em: https://doi.org/10.1016/j.compbiomed.2023.106646.

KAMNITSAS, Konstantinos *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. **Medical Image Analysis**, [*s. l.*], v. 36, p. 61–78, 2017. Disponível em: https://www.sciencedirect.com/science/article/pii/S1361841516301839.

KE, Jia *et al.* Application of 3DU-net in automatic segmentation of middle ear surgery structures in temporal bone CT. [*s. l.*], p. 6–9,

KE, Jia *et al.* Deep learning-based approach for the automatic segmentation of adult and pediatric temporal bone computed tomography images. **Quantitative Imaging in Medicine and Surgery**, [*s. l.*], v. 13, n. 3, p. 1577–1591, 2023.

KINGMA, Diederik P.; BA, Jimmy Lei. Adam: A method for stochastic optimization. **3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings**, [*s. l.*], p. 1–15, 2015.

KOHAVI, Ron; EDU, Stanford. A study of cross-validation and bootstrap for accuracy estimation and model selection. **Proceedings of the 14th International Joint Conference on Artificial intelligence**, [*s. l.*], v. 2, p. 1137–1143, 1993.

KOTSIANTIS, S B. Supervised Machine Learning: A Review of Classification Techniques. *In*: , 2007, NLD. **Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies**. NLD: IOS Press, 2007. p. 3–24.

LAI, Carolyn *et al.* High-Fidelity Virtual Reality Simulation for the Middle Cranial Fossa Approach-Modules for Surgical Rehearsal and Education. **Operative neurosurgery (Hagerstown, Md.)**, [*s. l.*], v. 23, n. 6, p. 505–513, 2022.

LECUN, Y *et al.* Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [*s. l.*], v. 86, n. 11, p. 2278–2324, 1998.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, [*s. l.*], v. 521, n. 7553, p. 436–444, 2015. Disponível em: https://doi.org/10.1038/nature14539.

LEONARD, Simon *et al.* Image-based navigation for functional endoscopic sinus surgery using structure from motion. **Medical Imaging 2016: Image Processing**, [*s. l.*], v. 9784, p. 97840V, 2016.

LEUZE, Christoph *et al.* Augmented Reality for Retrosigmoid Craniotomy Planning. **J Neurol Surg B Skull Base**, [*s. l.*], n. EFirst, 2021.

LEWKOWYCZ, Aitor *et al.* Solving Quantitative Reasoning Problems with Language Models. [*s. l.*], p. 1–54, 2022. Disponível em: http://arxiv.org/abs/2206.14858.

LI, Xiaoguang *et al.* A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images. **Neural Networks**, [*s. l.*], v. 124, p. 75–85, 2020. Disponível em: https://doi.org/10.1016/j.neunet.2020.01.005.

LI, Zhenhua *et al.* Application of UNETR for automatic cochlear segmentation in temporal bone CTs. **Auris Nasus Larynx**, [*s. l.*], v. 50, n. 2, p. 212–217, 2022. Disponível em: https://doi.org/10.1016/j.anl.2022.06.008.

LITJENS, Geert *et al.* A survey on deep learning in medical image analysis. **Medical Image Analysis**, [*s. l.*], v. 42, n. December 2012, p. 60–88, 2017.

LIU, Siqi *et al.* **3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes**. [*S. l.*]: Springer International Publishing, 2018-. ISSN 16113349.v. 11071 LNCS Disponível em: http://dx.doi.org/10.1007/978-3-030-00934-2_94.

LIU, Ze *et al.* Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. **Proceedings of the IEEE International Conference on Computer Vision**, [*s. l.*], p. 9992–10002, 2021.

LOCKETZ, Garrett D *et al.* Anatomy-Specific Virtual Reality Simulation in Temporal Bone Dissection: Perceived Utility and Impact on Surgeon Confidence. **Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery**, England, v. 156, n. 6, p. 1142–1149, 2017.

LV, Yi *et al.* Automatic segmentation of temporal bone structures from clinical conventional CT using a CNN approach. **International Journal of Medical Robotics and Computer Assisted Surgery**, [*s. l.*], v. 17, n. 2, p. 1–9, 2021.

MENG, Lu; TIAN, Yaoyu; BU, Sihang. Liver tumor segmentation based on 3D convolutional neural network with dual scale. **Journal of Applied Clinical Medical Physics**, [*s. l.*], v. 21, n. 1, p. 144–157, 2020.

MOAWAD, Ahmed W *et al.* Artificial Intelligence in Diagnostic Radiology : Where Do We Stand , Challenges , and Opportunities. [*s. l.*], v. 00, n. 00, p. 1–13, 2022.

MUELLER, Fabian *et al.* Image-Based Planning of Minimally Traumatic Inner Ear Access for Robotic Cochlear Implantation. **Frontiers in Surgery**, [*s. l.*], v. 8, n. November, p. 1–12, 2021.

MYRONENKO, Andriy; HATAMIZADEH, Ali. Robust Semantic Segmentation of Brain Tumor Regions from 3D MRIs BT - Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. *In*: , 2020, Cham. (Alessandro Crimi & Spyridon Bakas, Org.)**Anais [...]**. Cham: Springer International Publishing, 2020. p. 82–89.

NAKASHIMA, S *et al.* Computer-aided 3-D reconstruction and measurement of the facial canal and facial nerve. I. Cross-sectional area and diameter: preliminary report. **The Laryngoscope**, United States, v. 103, n. 10, p. 1150–1156, 1993.

NEVES, Caio A *et al.* Application of holographic augmented reality for external approaches to the frontal sinus. **International forum of allergy & rhinology**, United States, v. 10, n. 7, p. 920–925, 2020.

NEVES, C. A. *et al.* Automated Radiomic Analysis of Vestibular Schwannomas and Inner Ears using Contrast-enhanced T1-weighted and T2-weighted MRI Sequences and Artificial Intelligence. **Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology**, [*s. l.*], 2023.

NEVES, Caio A *et al.* Deep learning automated segmentation of middle skull-base structures for enhanced navigation. **International forum of allergy & rhinology**, United States, v. 11, n. 12, p. 1694–1697, 2021.

NEVES, Caio A. *et al.* Fully Automated Measurement of Cochlear Duct Length From Clinical Temporal Bone Computed Tomography. **Laryngoscope**, [*s. l.*], v. 132, n. 2, p. 449–458, 2022.

NEVES, C. A. *et al.* Fully automated preoperative segmentation of temporal bone structures from clinical CT scans. **Scientific Reports**, [*s. l.*], v. 11, n. 1, p. 1–11, 2021. Disponível em: https://doi.org/10.1038/s41598-020-80619-0.

NIKAN, Soodeh *et al.* Pwd-3dnet: A deep learning-based fully-automated segmentation of multiple structures on temporal bone ct scans. **IEEE Transactions on Image Processing**, [*s. l.*], v. 30, p. 739–753, 2021.

NISHIO, Mizuho *et al.* Automatic classification between COVID - 19 pneumonia , non

- COVID - 19 pneumonia , and the healthy on chest X - ray image : combination of data augmentation methods. **Scientific Reports**, [*s. l.*], n. 0123456789, p. 1–6, 2020. Disponível em: https://doi.org/10.1038/s41598-020-74539-2.

NOBLE, Jack H. *et al.* Automatic identification and 3D rendering of temporal bone anatomy. **Otology and Neurotology**, [*s. l.*], v. 30, n. 4, p. 436–442, 2009.

NOBLE, Jack H. *et al.* Automatic segmentation of intracochlear anatomy in conventional CT. **IEEE Transactions on Biomedical Engineering**, [*s. l.*], v. 58, n. 9, p. 2625–2632, 2011.

NOBLE, Jack H. *et al.* Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values. **Medical Physics**, [*s. l.*], v. 35, n. 12, p. 5375–5384, 2008.

NVIDIA CLARA IMAGING. [*S. l.*], 2022. Disponível em: https://developer.nvidia.com/clara-medical-imaging (2022). .

PATIL, Aseem; RANE, Milind. Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition. **Smart Innovation, Systems and Technologies**, [*s. l.*], v. 195, p. 21–30, 2021.

PIROMCHAI, Patorn *et al.* Virtual reality training for improving the skills needed for performing surgery of the ear, nose or throat. **Cochrane Database of Systematic Reviews**, [*s. l.*], v. 2015, n. 9, 2015.

POWELL, Kimerly A. *et al.* Atlas-Based Segmentation of Temporal Bone Anatomy. **International Journal of Computer Assisted Radiology and Surgery**, [*s. l.*], v. 12, n. 11, p. 1937–1944, 2017.

POWELL, Kimerly A. *et al.* Atlas-based segmentation of temporal bone surface structures. **International Journal of Computer Assisted Radiology and Surgery**, [*s. l.*], v. 14, n. 8, p. 1267–1273, 2019. Disponível em: https://doi.org/10.1007/s11548-019-01978-2.

PRISMAN, Eitan *et al.* Real-time tracking and virtual endoscopy in cone-beam CT-guided surgery of the sinuses and skull base in a cadaver model. **International Forum of Allergy and Rhinology**, [*s. l.*], v. 1, n. 1, p. 70–77, 2011.

RAJPURKAR, Pranav *et al.* CheXNet : Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. [*s. l.*], p. 3–9, 2017.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*], v. 9351, p. 234–241, 2015.

SACHIDANAND, Alle *et al.* **Nvidia AI-assisted annotation (AIAA) for 3D Slicer**. [*S. l.*], 2019. Disponível em: https://github.com/NVIDIA/ai-assisted-annotation-client/blob/master/slicer-plugin/README.md. .

SHAO, Xuefei *et al.* Virtual reality technology for teaching neurosurgery of skull base tumor. **BMC Medical Education**, [*s. l.*], v. 20, n. 1, p. 1–7, 2020.

SHERER, Michael V *et al.* Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: a critical review. **Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology**, Ireland, 2021.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, [*s. l.*], v. 6, n. 1, 2019. Disponível em: https://doi.org/10.1186/s40537-019-0197-0.

SILVER, David *et al.* Mastering the game of Go without human knowledge. **Nature**, England, v. 550, n. 7676, p. 354–359, 2017.

SINGH, Anup *et al.* Study of Sigmoid Sinus Variations in the Temporal Bone by Micro Dissection and its Classification - A Cadaveric Study. **International archives of otorhinolaryngology**, [*s. l.*], v. 23, n. 3, p. e311–e316, 2019.

SPEICHER, Maximilian; HALL, Brian D.; NEBELING, Michael. What is mixed reality?. **Conference on Human Factors in Computing Systems - Proceedings**, [*s. l.*], n. April, 2019.

SUN, Jiawei *et al.* DRRNet: Dense Residual Refine Networks for Automatic Brain Tumor Segmentation. **Journal of medical systems**, United States, v. 43, n. 7, p. 221, 2019.

TOPSAKAL, Vedat *et al.* Comparison of the Surgical Techniques and Robotic Techniques for Cochlear Implantation in Terms of the Trajectories Toward the Inner Ear. **The journal of international advanced otology**, Turkey, v. 16, n. 1, p. 3–7, 2020.

TOPSAKAL, Vedat *et al.* **First Study in Men Evaluating a Surgical Robotic Tool Providing Autonomous Inner Ear Access for Cochlear Implantation.** Switzerland: [*s. n.*], 2022.

VAN OSCH, Kylen *et al.* Morphological analysis of sigmoid sinus anatomy: clinical applications to neurotological surgery. **Journal of Otolaryngology - Head & Neck Surgery**, [*s. l.*], v. 48, n. 1, p. 2, 2019. Disponível em: https://doi.org/10.1186/s40463-019-0324-0.

VASWANI, Ashish *et al.* Attention is all you need. **Advances in Neural Information Processing Systems**, [*s. l.*], v. 2017-Decem, n. Nips, p. 5999–6009, 2017.

WANG, Jiang *et al.* Fully automated segmentation in temporal bone CT with neural network: a preliminary assessment study. **BMC Medical Imaging**, [*s. l.*], v. 21, n. 1, p. 1–11, 2021.

WIET, Gregory J.; SØRENSEN, Mads Sølvsten; ANDERSEN, Steven Arild Wuyts. Otologic Skills Training. **Otolaryngologic Clinics of North America**, [*s. l.*], v. 50, n. 5, p. 933–945, 2017.

WINNE, Christian *et al.* Overlay visualization in endoscopic ENT surgery. **International Journal of Computer Assisted Radiology and Surgery**, [*s. l.*], v. 6, n. 3, p. 401–406, 2011.

WON, Tae-Bin *et al.* Early experience with a patient-specific virtual surgical simulation for rehearsal of endoscopic skull-base surgery. **International forum of allergy & rhinology**, United States, v. 8, n. 1, p. 54–63, 2018.

WON, Tae-Bin *et al.* Validation of a rhinologic virtual surgical simulator for performing a Draf 3 endoscopic frontal sinusotomy. **International forum of allergy & rhinology**, United States, v. 9, n. 8, p. 910–917, 2019.

XIAO, Zhibo *et al.* Deep-Gamma: deep low-excitation fluorescence imaging global enhancement. **Optics letters**, United States, v. 48, n. 9, p. 2496–2499, 2023.

YUAN, Jie *et al.* **Extended reality for biomedicine**. [*S. l.: s. n.*], 2023. v. 3

APÊNDICE A – Informação sobre plataformas de projetos de aprendizado profundo

A expansão dos projetos de segmentação automatizada por DL foi acompanhada e alavancada pela disponibilidade de kits de ferramentas de software, como o kit de ferramentas Clara SDK (NVIDIA CLARA IMAGING, 2022) e mais recentemente a plataforma MONAI (GEORGE-JONES *et al.*, 2021), assim como a evolução da plataforma 3D Slicer (FEDOROV *et al.*, 2012). MONAI é uma plataforma de desenvolvimento e suporte para a implementação de projetos de modelos de aprendizado profundo para a área biomédica. Já a plataforma 3D Slicer e suas extensões fornecem uma interface amigável para as predições e manipulação dos dados de segmentação em software robusto orientado para pesquisa (FEDOROV *et al.*, 2012).

APÊNDICE B – Detalhes da análise objetiva do modelo do Exp. 2 sobre o conjunto de testes

Estrutura		Dice	VP	VN	FP	FN	Vol manual	Vol auto	AHD	HD95	Especificidade	Sensibilidade	PB	Similaridade volumérica
Art. Carótida interna	Média	0.900	0.247	99.696	0.030	0.027	0.662	0.667	0.284	1.025	0.9997	0.914	0.957	103%
	DP	0.057	0.096	0.129	0.030	0.034	0.396	0.391	0.225	0.954	0.0003	0.092	0.046	15%
N. corda do timpano	Média	0.591	0.001	99.997	0.002	0.001	0.003	0.006	0.405	1.064	1.0000	0.657	0.828	175%
	DP	0.177	0.001	0.006	0.006	0.000	0.002	0.011	0.887	1.753	0.0001	0.206	0.103	328%
Orelha interna	Média	0.948	0.072	99.920	0.004	0.004	0.176	0.178	0.063	0.319	1.0000	0.953	0.976	101%
	DP	0.021	0.026	0.029	0.004	0.003	0.032	0.032	0.046	0.412	0.0000	0.039	0.020	6%
Cond. aud. ext.	Média	0.904	0.307	99.611	0.020	0.062	0.818	0.733	0.287	1.050	0.9998	0.885	0.942	94%
	DP	0.096	0.132	0.241	0.017	0.188	0.410	0.169	0.434	1.049	0.0002	0.121	0.060	14%
Nervo facial	Média	0.830	0.021	99.971	0.006	0.002	0.055	0.062	0.138	0.460	0.9999	0.900	0.950	118%
	DP	0.059	0.007	0.009	0.003	0.002	0.016	0.014	0.045	0.216	0.0000	0.062	0.031	23%
Cond., aud, interno	Média	0.931	0.085	99.902	0.006	0.007	0.205	0.204	0.126	0.516	0.9999	0.934	0.967	101%
	DP	0.027	0.051	0.061	0.005	0.012	0.077	0.072	0.051	0.258	0.0000	0.061	0.030	10%
Ossículos	Média	0.885	0.012	99.983	0.003	0.001	0.031	0.036	0.170	0.502	1.0000	0.928	0.964	122%
	DP	0.150	0.005	0.015	0.013	0.001	0.009	0.024	0.456	0.823	0.0001	0.088	0.044	79%
Cápsula ótica	Média	0.949	0.302	99.666	0.018	0.014	0.738	0.750	0.074	0.258	0.9998	0.957	0.978	102%
	DP	0.031	0.090	0.100	0.013	0.017	0.092	0.103	0.033	0.030	0.0001	0.050	0.025	7%
Seio sigmoide	Média	0.860	0.854	98.891	0.092	0.164	2.314	2.149	0.465	1.906	0.9991	0.842	0.920	95%
	DP	0.071	0.490	0.616	0.070	0.152	0.980	0.798	0.343	1.474	0.0007	0.111	0.055	18%
	Média	0.866	0.211	99.738	0.020	0.031	0.556	0.532	0.224	0.789	0.9998	0.885	0.943	112%
	DP	0.077	0.100	0.134	0.018	0.046	0.224	0.179	0.280	0.774	0.0002	0.092	0.046	56%

APÊNDICE C – Configuração do algoritmo do Exp. 2


```
{
    "imports": [
        "$import glob",
        "$import os",
        "$import ignite"
    ],
    "bundle_root": "/workspace/models/v2_multilabel/",
    "ckpt_dir": "$@bundle_root + '/models'",
    "output_dir": "$@bundle_root + '/eval'",
    "dataset_dir": "/workspace/dataset/train/",
    "images": "$list(sorted(glob.glob(@dataset_dir + '/img2/*.nii.gz')))",
    "labels": "$list(sorted(glob.glob(@dataset_dir + '/labels2/*.nii.gz')))",
    "val_interval": 5,
    "device": "$torch.device('cuda:0' if torch.cuda.is_available() else 'cpu')",
    "network_def": {
        "_target_": "SwinUNETR",
        "spatial_dims": 3,
        "img_size": 128,
        "in_channels": 1,
        "out_channels": 10,
        "feature_size": 96,
        "use_checkpoint": true
    },
    "network": "$@network_def.to(@device)",
    "loss": {
        "_target_": "DiceCELoss",
        "to_onehot_y": true,
        "softmax": true,
        "squared_pred": true,
        "batch": true
    },
    "optimizer": {
        "_target_": "torch.optim.Adam",
        "params": "$@network.parameters()",
        "lr": 0.0002
    },
    "train": {
        "deterministic_transforms": [
            {
                "_target_": "LoadImaged",
                "keys": [
                    "image",
                    "label"
                ],
                "reader": "ITKReader"
            },
```

```
        {
            "_target_": "EnsureChannelFirstd",
            "keys": [
                "image",
                "label"
            ]
        },
        {
            "_target_": "Orientationd",
            "keys": [
                "image",
                "label"
            ],
            "axcodes": "RAS"
        },
        {
            "_target_": "Spacingd",
            "keys": [
                "image",
                "label"
            ],
            "pixdim": [
                0.25,
                0.25,
                0.25
            ],
            "mode": [
                "bilinear",
                "nearest"
            ]
        },
        {
            "_target_": "ScaleIntensityRanged",
            "keys": "image",
            "a_min": -500,
            "a_max": 1800,
            "b_min": 0.0,
            "b_max": 1.0,
            "clip": true
        },
        {
            "_target_": "EnsureTyped",
            "keys": [
                "image",
                "label"
            ]
        }
    ],
    "random_transforms": [
```

```
{
    "_target_": "RandCropByPosNegLabeld",
    "keys": [
        "image",
        "label"
    ],
    "label_key": "label",
    "spatial_size": [
        96,
        96,
        96
    ],
    "pos": 1,
    "neg": 1,
    "num_samples": 2,
    "image_key": "image",
    "image_threshold": 0
},
{
    "_target_": "RandFlipd",
    "keys": [
        "image",
        "label"
    ],
    "spatial_axis": [
        0
    ],
    "prob": 0.4
},
{
    "_target_": "RandZoomd",
    "keys": [
        "image",
        "label"
    ],
    "min_zoom": 0.9,
    "max_zoom": 1.1,
    "prob": 0.3
},
{
    "_target_": "RandGaussianNoised",
    "keys": [
        "image"
    ],
    "mean": 0.0,
    "std": 0.1,
    "prob": 0.2
},
{
```

```
            "_target_": "RandRotated",
            "keys": [
                "image",
                "label"
            ],
            "range_x": 0.2,
            "range_y": 0.2,
            "range_z": 0.2,
            "prob": 0.33
        },
        {
            "_target_": "RandShiftIntensityd",
            "keys": "image",
            "offsets": 0.1,
            "prob": 0.5
        }
    ],
    "preprocessing": {
        "_target_": "Compose",
        "transforms": "$@train#deterministic_transforms +
@train#random_transforms"
    },
    "dataset": {
        "_target_": "Dataset",
        "data": "$[{'image': i, 'label': l} for i, l in zip(@images[:-65], @labels[:-65])]",
        "transform": "@train#preprocessing"
    },
    "dataloader": {
        "_target_": "DataLoader",
        "dataset": "@train#dataset",
        "batch_size": 2,
        "shuffle": true,
        "num_workers": 4
    },
    "inferer": {
        "_target_": "SimpleInferer"
    },
    "postprocessing": {
        "_target_": "Compose",
        "transforms": [
            {
                "_target_": "Activationsd",
                "keys": "pred",
                "softmax": true
            },
            {
                "_target_": "AsDiscreted",
                "keys": [
                    "pred",
```

```
            "label"
          ],
          "argmax": [
            true,
            false
          ],
          "to_onehot": 10
        }
    ]
},
"handlers": [
    {
        "_target_": "ValidationHandler",
        "validator": "@validate#evaluator",
        "epoch_level": true,
        "interval": "@val_interval"
    },
    {
        "_target_": "StatsHandler",
        "tag_name": "train_loss",
        "output_transform": "$monai.handlers.from_engine(['loss'], first=True)"
    },
    {
        "_target_": "TensorBoardStatsHandler",
        "log_dir": "@output_dir",
        "tag_name": "train_loss",
        "output_transform": "$monai.handlers.from_engine(['loss'], first=True)"
    }
],
"key_metric": {
    "train_accuracy": {
        "_target_": "ignite.metrics.Accuracy",
        "output_transform": "$monai.handlers.from_engine(['pred', 'label'])"
    }
},
"trainer": {
    "_target_": "SupervisedTrainer",
    "max_epochs": 1000,
    "device": "@device",
    "train_data_loader": "@train#dataloader",
    "network": "@network",
    "loss_function": "@loss",
    "optimizer": "@optimizer",
    "inferer": "@train#inferer",
    "postprocessing": "@train#postprocessing",
    "key_train_metric": "@train#key_metric",
    "train_handlers": "@train#handlers",
    "amp": true
}
```

```
},
"validate": {
    "preprocessing": {
        "_target_": "Compose",
        "transforms": "%train#deterministic_transforms"
    },
    "dataset": {
        "_target_": "CacheDataset",
        "data": "$[{'image': i, 'label': l} for i, l in zip(@images[-65:], @labels[-65:])]",
        "transform": "@validate#preprocessing",
        "cache_rate": 1.0
    },
    "dataloader": {
        "_target_": "DataLoader",
        "dataset": "@validate#dataset",
        "batch_size": 1,
        "shuffle": false,
        "num_workers": 4
    },
    "inferer": {
        "_target_": "SlidingWindowInferer",
        "roi_size": [
            128,
            128,
            128
        ],
        "sw_batch_size": 1,
        "overlap": 0.33
    },
    "postprocessing": "%train#postprocessing",
    "handlers": [
        {
            "_target_": "StatsHandler",
            "iteration_log": false
        },
        {
            "_target_": "TensorBoardStatsHandler",
            "log_dir": "@output_dir",
            "iteration_log": false
        },
        {
            "_target_": "CheckpointSaver",
            "save_dir": "@ckpt_dir",
            "save_dict": {
                "model": "@network"
            },
            "save_key_metric": true,
            "key_metric_filename": "model.pt"
        }
```

```
        ],
        "key_metric": {
            "val_mean_dice": {
                "_target_": "MeanDice",
                "include_background": false,
                "output_transform": "$monai.handlers.from_engine(['pred', 'label'])"
            }
        },
        "additional_metrics": {
            "val_accuracy": {
                "_target_": "ignite.metrics.Accuracy",
                "output_transform": "$monai.handlers.from_engine(['pred', 'label'])"
            }
        },
        "evaluator": {
            "_target_": "SupervisedEvaluator",
            "device": "@device",
            "val_data_loader": "@validate#dataloader",
            "network": "@network",
            "inferer": "@validate#inferer",
            "postprocessing": "@validate#postprocessing",
            "key_val_metric": "@validate#key_metric",
            "additional_metrics": "@validate#additional_metrics",
            "val_handlers": "@validate#handlers",
            "amp": true
        }
    },
    "training": [
        "$monai.utils.set_determinism(seed=123)",
        "$setattr(torch.backends.cudnn, 'benchmark', True)",
        "$@train#trainer.run()"
    ]
}
```

ANEXO 1 – Carta de aprovação do estudo pelo comitê de ética da Universidade de Stanford.

STANFORD UNIVERSITY

Stanford, CA 94305 [Mail Code 5579]

David D Oakes, M.D. (650) 724-6695
CHAIR, PANEL ON MEDICAL HUMAN SUBJECTS (650) 725-8013

Certification of Human Subjects Approvals

Date: October 15, 2019

To: Nikolas Blevins, MD, OHNS/Otology & Neurotology Division
Homer I. Abaya BS, Ksenia Aviella Aaron MD, Caio Athayde Neves MD, Davood Hosseini, Jennifer C Alyono, Steven Domenic Losorelli BS, Haiying Sun, Mrudula Penta MD, Nancy Fischbein M.D., Bradley J Girod, Stephen Creig Marcott, Yohan Song MD, Dr. Yilai Shu

From: David D Oakes, M.D., Administrative Panel on Human Subjects in Medical Research

eProtocol Imaging Characterization Of Middle Ear Ligaments, Sigmoid sinus, cochlea and Internal carotid dehiscences

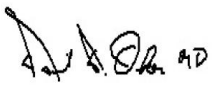
eProtocol #: 38946 **IRB 6 (Registration** 6)

The IRB approved human subjects involvement in your research project on 10/15/2019. **'Prior to subject recruitment and enrollment, if this is: a Cancer-related study, you must obtain Cancer Center Scientific Review Committee (SRC) approval;** a CTRU study, you must obtain CTRU approval; a VA study, you must obtain VA R and D Committee approval; and **if a contract is involved, it must be signed.'**

This protocol has been approved under the Extended Approval Process and **approval does not expire**. Proposed changes to approved research must still be reviewed and approved prospectively by the IRB. No changes may be initiated without prior approval by the IRB, except where necessary to eliminate apparent immediate hazards to subjects. (Any such exceptions must be reported to the IRB within 10 working days.) Unanticipated problems involving risks to participants or others and other events or information, as defined and listed in the Report Form, must be submitted promptly to the IRB. (See Events and Information that Require Prompt Reporting to the IRB at http://humansubjects.stanford.edu.) It is your responsibility to report the completion of the protocol to the IRB within 30 days.

Please remember that all data, including all signed consent form documents, must be retained for a minimum of three years past the completion of this research. Additional requirements may be imposed by your funding agency, your department, HIPAA, or other entities. (See Policy 1.9 on Retention of and Access to Research Data at http://doresearch.stanford.edu/policies/research-policy-handbook)

This institution is in compliance with requirements for protection of human subjects, including 45 CFR 46, 21 CFR 50 and 56, and 38 CFR 16.



David D Oakes, M.D., Chair

Approval Period: 10/15/2019 - (Does Not Expire)
Review Type: EXPEDITED - MODIFICATION
Funding: None
Expedited Under Category: 5
Assurance #: FWA00000935 (SU), FWA00000934 (SHC), FWA00000933 (LPCH)

ANEXO 2 – Artigo publicado com os resultados do Exp.1

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Fully automated preoperative segmentation of temporal bone structures from clinical CT scans

C. A. Neves[1✉], E. D. Tran[2], I. M. Kessler[1] & N. H. Blevins[2]

Middle- and inner-ear surgery is a vital treatment option in hearing loss, infections, and tumors of the lateral skull base. Segmentation of otologic structures from computed tomography (CT) has many potential applications for improving surgical planning but can be an arduous and time-consuming task. We propose an end-to-end solution for the automated segmentation of temporal bone CT using convolutional neural networks (CNN). Using 150 manually segmented CT scans, a comparison of 3 CNN models (AH-Net, U-Net, ResNet) was conducted to compare Dice coefficient, Hausdorff distance, and speed of segmentation of the inner ear, ossicles, facial nerve and sigmoid sinus. Using AH-Net, the Dice coefficient was 0.91 for the inner ear; 0.85 for the ossicles; 0.75 for the facial nerve; and 0.86 for the sigmoid sinus. The average Hausdorff distance was 0.25, 0.21, 0.24 and 0.45 mm, respectively. Blinded experts assessed the accuracy of both techniques, and there was no statistical difference between the ratings for the two methods (p = 0.93). Objective and subjective assessment confirm good correlation between automated segmentation of otologic structures and manual segmentation performed by a specialist. This end-to-end automated segmentation pipeline can help to advance the systematic application of augmented reality, simulation, and automation in otologic procedures.

Safe and effective middle- and inner-ear surgery requires extensive training and knowledge of radiological and surgical anatomy. Procedures such as cochlear implantation, tympanomastoidectomy, and superior semicircular canal dehiscence repair depend on the pre- and intra-operative identification of critical structures and an appreciation of their complex interrelationships[1]. Individualized preoperative planning and the implementation of augmented reality systems may assist in such surgery given the intricacy and variability of anatomy involved. Such efforts require specialized anatomical and radiological knowledge of the key structures, which takes considerable time and effort to acquire. A method for the rapid and accurate generation of patient-specific, high-fidelity 3D models for preoperative planning[2] and intraoperative navigation[3,4] would offer a variety of potential benefits to both patient and surgeon.

Computed tomography (CT) imaging of the temporal bone is critical to provide otologists insights into a patient's unique anatomy for pre-operative planning. However, identifying structures of interest and subtle developmental or pathologic variations may be challenging for both surgeons and radiologists due to the structures' small size and inherent complexity. However, understanding their orientation and geometry is essential for successful otologic procedures such as cochlear implantation or tumor removal[5]. In addition, although CT datasets are inherently volumetric, surgeons routinely review them as multiplanar two dimensional (2D) representations. This necessitates a mental translation of the data back into the three-dimensional (3D) relationships expected at the time of surgery. Efforts to enhance preoperative planning using innovative tools such as 3D simulations and augmented reality offer promise for improving operative safety and efficiency. However, these efforts are limited by the labor intensive step of manual segmentation of imaging data[6,7] by highly trained specialists (Fig. 1). An automated pipeline of medical image segmentation for temporal bone CT (TBCT) scans might expand the application of simulation, planning, and procedural automation.

Cochlear implantation is an example of an otologic procedure that is both commonly performed and highly influenced by anatomic variability. As such, it has motivated a number of studies to integrate computer-assisted segmentation to increase safety and efficacy. Early works include computer-aided analysis of human temporal bone histopathology specimens by Nakashima et al.[8]. Noble et al.[9–11] published a series of papers using atlas-based approaches and other customized solutions for automated identification of the facial nerve, ossicles and intracochlear anatomy. Recently, Powell et al.[12] and Gare[5] also showed strong correlation of atlas-based auto-segmentation of the temporal bone with ground truth. Hudson et al. used atlas-based models registered to

[1]Faculty of Medicine, University of Brasilia, Brasília, DF, Brazil. [2]Otolaryngology Head & Neck Surgery, Stanford University School of Medicine, Stanford, CA, USA. ✉email: caioath@gmail.com

ANEXO 3 – Carta de aceitação de poster com resultados do Exp. 2

Gmail

Caio Athayde Neves <caioath@gmail.com>

{Poster Acceptance Notification} AAO-HNSF 2023 Annual Meeting & OTO Experience

posters@entnet.org <posters@entnet.org> 4 de maio de 2023 às 17:36
Para: caioath@gmail.com

Dear Dr. Neves,

Congratulations!

On behalf of the Annual Meeting Program Committee, the following proposal(s) were reviewed and accepted for presentation at the AAO-HNSF 2023 Annual Meeting & OTO Experience.

Deep Learning Method for Rapid Simultaneous Multi-structure Temporal Bone Segmentation
Specialty: Otology/Neurotology

This official notification will be followed up with an invitation to the **AAO-HNSF Poster Presenter Portal** that will expand on the information provided below.

Online Poster Gallery
In addition to displaying a physical copy your poster in Nashville, you are required to upload an electronic version so it can be made accessible to all attendees. You will also have the option to create an audio file introducing your poster in the online gallery.

Daniel C. Chelius, Jr., MD, FAAP, FACS
Coordinator, Annual Meeting Program Committee
AAO-HNSF
www.entannualmeeting.org