



Universidade de São Paulo

Lucas de Almeida Louzada

**Segmentação da Articulação Sacroilíaca em Imagens de Ressonância Magnética
Utilizando a Rede Neural Convolutacional U-Net no Framework nnU-Net**

RIBEIRÃO PRETO
2023

Lucas de Almeida Louzada

**Segmentação da Articulação Sacroilíaca em Imagens de Ressonância Magnética
Utilizando a Rede Neural Convolutacional U-Net no Framework nnU-Net**

Trabalho de Conclusão de Curso de Graduação, apresentado à disciplina de Trabalho de Conclusão de Curso (RIB0114), do Curso Superior de Informática Biomédica, Faculdade de Medicina, da Universidade de São Paulo – Campus Ribeirão Preto, como requisito parcial para obtenção do título de Bacharel em Informática Biomédica.

Orientador: Prof. Dr. Paulo Mazzoncini de Azevedo Marques

RIBEIRÃO PRETO
2023

AGRADECIMENTOS

Primeiramente agradeço a minha família, por todo o apoio, suporte e paciência durante essa jornada. Obrigado por todos os conselhos e por me motivar a chegar até aqui.

Agradeço também ao Prof. Dr Paulo Mazzoncini de Azevedo Marques, pela orientação, motivação e entusiasmo no desenvolvimento deste trabalho.

Não posso deixar de agradecer aos meus colegas e amigos, pela companhia e pelas memórias. E a todos que de alguma forma tornaram possível o desenvolvimento deste trabalho.

RESUMO

A avaliação por imagem é um processo fundamental no diagnóstico de pacientes com espondiloartropatias soronegativas, grupo de doenças inflamatórias sistêmicas crônicas que, caracteristicamente, apresentam manifestações clínicas precoces nas articulações sacroilíacas e podem causar um grande impacto na vida de portadores. Portanto, a detecção em estágios iniciais é de extrema importância.

Este projeto propõe a aplicação da arquitetura de rede neural convolucional U-Net, utilizando-se do *framework* nnU-Net, para identificar e segmentar a articulação sacroilíaca em imagens de ressonância magnética (RM). A correta identificação e segmentação dessa articulação são passos cruciais para que modelos de aprendizado de máquina possam ser empregados de maneira eficaz no auxílio ao diagnóstico da sacroiliíte e de outras condições associadas à inflamação desta articulação.

Para o desenvolvimento deste projeto, foram utilizadas imagens de ressonância magnética previamente selecionadas, que foram segmentadas manualmente por especialistas. A utilização dessas imagens, juntamente com a arquitetura UNet, tem como objetivo realizar uma segmentação precisa das articulações em foco.

Palavras-chave: U-Net. nnU-Net. Aprendizado de Máquina. sacroiliíte. Articulações Sacroilíacas.

ABSTRACT

Imaging assessment is a fundamental process in the diagnosis of patients with spondyloarthropathies, diseases that present early manifestations in the sacroiliac joints and can have a major impact on the lives of sufferers. Therefore, detection at early stages is of utmost importance.

This project proposes the application of the UNet convolutional neural network architecture, using the framework nnU-Net, to identify and segment the sacroiliac joint in magnetic resonance images (MRI). The correct identification and segmentation of this joint are crucial steps so that machine learning models can be used effectively to aid in the diagnosis of sacroiliitis and other conditions associated with inflammation of this joint.

For the development of this project, previously selected magnetic resonance images were used, which were manually segmented by experts. The use of these images, together with the UNet architecture, it aims to perform an accurate segmentation of the joints in focus.

Keywords: U-Net. nnU-Net. Machine Learning. Sacroiliitis. Sacroiliac Joints.

LISTA DE FIGURAS

Figura 1 – Arquitetura da U-Net.	14
Figura 2 – Exemplo das imagens após o pré-processamento.	25
Figura 3 – Exemplos de imagens resultantes após o processo de <i>data augmentation</i> . . .	26
Figura 4 – Estrutura de diretórios para uso da nnU-Net.	27
Figura 5 – Duração dos <i>epochs</i>	29
Figura 6 – Progressão da taxa de aprendizagem durante o treinamento do modelo. . . .	30
Figura 7 – Evolução das perdas de treinamento e validação ao longo dos <i>epochs</i>	30
Figura 8 – Exemplo de predição 1.	32
Figura 9 – Exemplo de predição 2.	33
Figura 10 – Exemplo de predição 3.	33
Figura 11 – Exemplo de predição arredondada 1.	33
Figura 12 – Exemplo de predição arredondada 2.	34
Figura 13 – Exemplo de predição com falhas 1.	34
Figura 14 – Exemplo de predição com falhas 2.	34
Figura 15 – Box-plot da distribuição dos coeficientes Dice obtidos.	35
Figura 16 – Box-plot da distribuição dos coeficientes IoU obtidos.	35
Figura 17 – Exemplos de imagens com sobreposição da região de interesse.	37

LISTA DE TABELAS

Tabela 1 – Médias dos coeficientes Dice para cada modelo após treinamento.	29
Tabela 2 – Metricas obtidas após a validação cruzada.	31
Tabela 3 – Resumo dos Coeficientes Dice Obtidos.	36
Tabela 4 – Resumo dos IoU obtidos.	36
Tabela 5 – Intervalo de Confiança para a média da amostra de Coeficientes Dice, com 95% de confiança.	36
Tabela 6 – Intervalo de Confiança para a média da amostra de Coeficientes IoU, com 95% de confiança.	36

LISTA DE ABREVIATURAS E SIGLAS

RM	Ressonância Magnética
EAS	Espondiloartropatias Soronegativas
AS	Articulações Sacroilíacas
ML	<i>Machine Learning</i>
CNN	<i>Deep Convolutional Neural Networks</i>
nnU-Net	<i>No-new-U-Net</i>
SPAIR	<i>Spectral Attenuated Inversion Recovery</i>
STIR	<i>Short Tau Inversion Recovery</i>
CEP-HCRP	Comitê de Ética em Pesquisa do Hospital das Clínicas de Ribeirão Preto
CAD	<i>Computer Aided Diagnosis</i> (Diagnóstico Auxiliado por Computador)
FCNs	<i>Fully Convolutional Networks</i> (redes totalmente convolucionais)
ReLU	<i>Rectified Linear Units</i> (Unidade Linear Retificada)
IoU	<i>Intersection over Union</i> (Intersecção sobre União)
ROI	<i>Regions of Interest</i> (Regiões de Interesse)

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Contextualização	11
1.2	U-Net e nnU-Net	13
1.2.1	<i>A Rede U-Net</i>	13
1.2.1.1	<i>Detalhes da Arquitetura U-Net</i>	14
1.2.2	<i>O framework nnU-Net</i>	16
1.3	Coeficiente Dice	18
1.4	<i>Intersection over Union</i>	19
1.5	Comentários Finais	19
2	OBJETIVOS	21
2.1	Objetivos Gerais	21
2.2	Objetivos Específicos	21
2.3	Comentários Finais	22
3	MATERIAIS E MÉTODOS	23
3.1	Conjunto de Imagens	23
3.2	Pré-processamento	23
3.3	<i>Data Augmentation</i>	25
3.4	Implementação da Arquitetura	27
3.4.1	<i>Configuração e Testes</i>	27
3.4.2	<i>Treinamento</i>	28
3.4.3	<i>5-Fold Cross Validation</i>	31
3.4.4	<i>Pós-processamento</i>	31
3.5	Comentários Finais	31
4	RESULTADOS	32
4.1	Análise Visual dos Resultado	32
4.2	Análise Quantitativa dos Resultados	35
4.3	Resultados Adicionais	37

4.4	Comentários Finais	38
5	CONCLUSÃO E DISCUSSÃO	39
5.1	Conclusão e Discussão	39
	REFERÊNCIAS	41

1 INTRODUÇÃO

Neste Capítulo serão introduzidos os conceitos principais abordados no presente Trabalho de Conclusão de Curso. Será contextualizada a problemática da sacroiliíte, trazendo um panorama da doença e de suas dificuldades diagnósticas. Também introduzimos a rede U-Net e suas inovações bem como o *framework* nnU-Net e suas ferramentas. Ao final do capítulo, conceitualizamos as principais métricas que serão usadas ao longo do desenvolvimento do trabalho para a avaliação do resultado final, incluindo o coeficiente Dice e IoU.

1.1 Contextualização

Espondiloartropatias soronegativas (EAS) são um grupo de doenças inflamatórias sistêmicas crônicas que, caracteristicamente, apresentam manifestações clínicas precoces nas articulações sacroilíacas (AS) (CASTRO et al., 2017). Essas doenças podem causar um grande impacto na qualidade de vida dos portadores, afetando significativamente sua mobilidade e conforto diário, apresentando muitas vezes dor na região lombar que pode se estender pelas pernas e em casos mais graves acometer a coluna vertebral, podendo evoluir com rigidez e limitação funcional progressiva do esqueleto axial, como no caso da espondilite anquilosante (EA) (SAMPAIO-BARROS et al., 2007). As EAS englobam condições como a EA, artrite psoriásica, artrite reativa, entre outras. Segundo Sampaio-Barros et al. (2007), elas compartilham aspectos clínicos, radiológicos e laboratoriais comuns, tais como:

- Dor axial inflamatória associada à artrite;
- Aspectos radiológicos característicos, como a sacroiliíte;
- Aspectos laboratoriais específicos, incluindo a ausência de anticorpos reumatóides no sangue, caracterizando-as como "soronegativas".

As manifestações clínicas precoces nas AS são de lenta evolução e sem marcadores laboratoriais específicos que demonstrem sua atividade (CASTRO et al., 2017), sendo caracterizadas pelo processo inflamatório não-infeccioso das mesmas, nomeado Sacroiliíte. Tal processo é critério diagnóstico das EAS (MONTANDON et al., 2007). Por esta razão, na prática clínica, o diagnóstico e a avaliação de sacroiliíte são baseados principalmente em exames de imagem, mais precisamente a ressonância magnética (RM), por ser extremamente sensível a mudanças

na composição dos tecidos moles e na presença de líquido, o que a torna especialmente útil na detecção precoce de processos inflamatórios, permitindo a identificação em estágios iniciais (CASTRO et al., 2017).

Diante da complexidade dessas doenças, a detecção precoce é crítica, pois o tratamento oportuno pode retardar ou até mesmo impedir a progressão do quadro patológico (SAMPAIO-BARROS et al., 2007). Embora as técnicas de imagem sejam valiosas, a sua interpretação pode ser complexa, o que possibilita a ocorrência de variações diagnósticas tanto intra quanto interpessoais (FALEIROS, 2018). Mais especificamente, o potencial para diagnóstico falso-positivo de sacroileíte por RM é relativamente alto, exigindo experiência e perícia radiológica profunda do padrão de sacroiliíte por RM e seu diagnóstico diferencial (SLOBODIN et al., 2018). Para um resultado mais robusto e preciso, modelos de aprendizado de máquina podem ser empregados como ferramenta de auxílio ao diagnóstico da sacroiliíte e de outras condições associadas à inflamação desta articulação (SHENKMAN et al., 2019).

É neste contexto que o presente trabalho busca explorar a eficácia do emprego de modelos de *Machine Learning*¹ (ML), em específico, o uso de *Deep Convolutional Neural Networks*² (CNNs), propondo a aplicação da arquitetura U-Net, através da implementação do *framework* nnU-Net, para segmentar imagens de RM da AS. Assim, procurando trazer *insights* à respeito do potencial da aplicação de tais técnicas na automação do processo diagnóstico da sacroiliíte e das EAS.

A arquitetura U-Net foi escolhida devido ao seu destacado desempenho em várias aplicações de segmentação de imagens biomédicas, conforme evidenciado pelos trabalhos na literatura. A eficácia dessa abordagem foi inicialmente demonstrada por Ronneberger, Fischer e Brox (2015), que introduziram a arquitetura U-Net e relataram resultados promissores. Posteriormente, Isensee et al. (2018b) alcançaram o mais alto valor médio do coeficiente Dice, uma métrica de precisão de segmentação, no *Medical Segmentation Decathlon*³, utilizando o *framework* nnU-Net. Além disso, a pesquisa de Zhang et al. (2023) adotou um fluxo de trabalho semelhante ao proposto neste estudo, integrando o nnU-Net com técnicas adicionais para classificar a sacroiliíte ativa em imagens de tomografia computadorizada, demonstrando a precisão e a versatilidade da U-Net em contextos desafiadores de processamento de imagem.

¹ Aprendizado de Máquina

² Redes Neurais Convolucionais Profundas

³ Decatlo de Segmentação Médica

1.2 U-Net e nnU-Net

Nos últimos anos, as redes neurais convolucionais (CNNs) superaram o estado da arte em diversas tarefas de reconhecimento visual, transformando substancialmente campos que vão desde a visão computacional até o processamento de linguagem natural (GIRSHICK et al., 2014). No domínio da assistência médica, essa transformação é evidente no diagnóstico auxiliado por computador (CAD), onde métodos baseados em aprendizado de máquina têm sido aplicados por mais de duas décadas. Entretanto, as limitações dessas técnicas tradicionais em tarefas complexas abriram caminho para o advento das CNNs como ferramentas mais eficazes e robustas (HADJIISKI et al., 2023).

As redes CNNs têm sido fundamentais no avanço da segmentação, classificação e detecção em imagens médicas, provando ser uma ferramenta essencial no CAD para diversas modalidades como tomografia computadorizada, ultrassom, raio-x e RM (ALOM et al., 2019).

A crescente disponibilidade de grandes conjuntos de dados de imagem médica e o desenvolvimento contínuo de arquiteturas de rede neural mais avançadas têm impulsionado essas melhorias. Em particular, a U-Net e suas iterações subsequentes, como o *framework* nnU-Net, representam marcos significativos nessa trajetória. A U-Net, originalmente projetada para a segmentação de imagens biomédicas, destaca-se pela sua eficiência em aprender a partir de um número limitado de amostras de treinamento, mantendo ao mesmo tempo um desempenho de alto nível (RONNEBERGER; FISCHER; BROX, 2015).

1.2.1 A Rede U-Net

Diferentemente das tarefas de classificação de imagem onde um rótulo de classe único é atribuído à imagem inteira, o processamento de imagem biomédica frequentemente exige localização, necessitando que um rótulo de classe seja atribuído a cada pixel (RONNEBERGER; FISCHER; BROX, 2015). Neste contexto, técnicas como a abordagem de janela deslizante tornaram-se essenciais, conforme explorado por Ciresan et al. (2012), mas tal método enfrentava desafios significativos relacionados à velocidade de processamento e eficiência devido à redundâncias no processamento. Além disso, a técnica de janela deslizante apresentava um compromisso entre a precisão da localização e a extensão do contexto capturado (RONNEBERGER; FISCHER; BROX, 2015).

Para superar essas limitações, as redes totalmente convolucionais (FCNs) (LONG; SHE-

LHAMER; DARRELL, 2015) emergiram como uma arquitetura elegante, levando a modificações e extensões significativas que se materializaram na forma da U-Net. A U-Net, uma variação das FCNs, foi especialmente projetada para segmentação de imagens biomédicas, oferecendo uma estrutura que permitia a segmentação de imagens precisas com uma quantidade relativamente pequena de dados de treinamento (SIDDIQUE et al., 2021).

1.2.1.1 Detalhes da Arquitetura U-Net

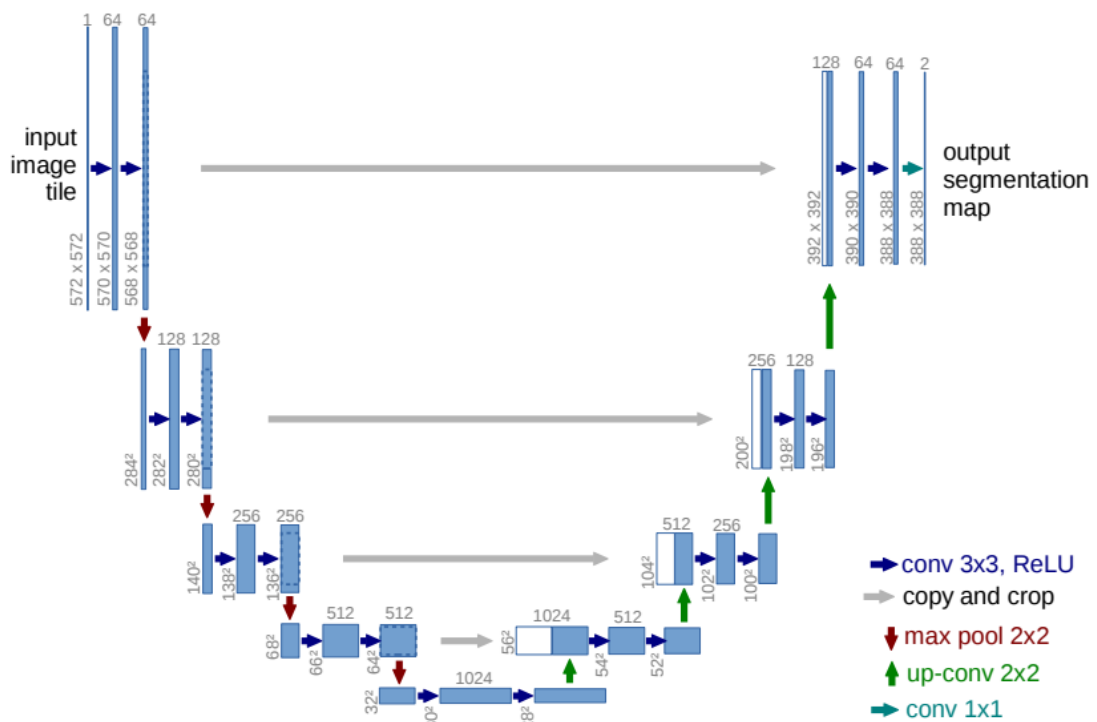


Figura 1 – Arquitetura da U-Net (exemplo para imagens de resolução mais baixa de 32x32 pixels). Cada caixa azul corresponde a um mapa de características multi-canal, com o número de canais indicado acima da caixa. O tamanho em x-y é fornecido no canto inferior esquerdo. Caixas brancas representam mapas de características que são copiados, e as setas denotam as diferentes operações. Fonte: Adaptado de (RONNEBERGER; FISCHER; BROX, 2015).

Apresentada por Ronneberger, Fischer e Brox (2015), a arquitetura da U-Net é projetada para ser eficiente em termos de memória e adequada para o processamento de imagens médicas, onde os dados de treinamento podem ser limitados. A rede é composta por um caminho contrativo (*downsampling*) e um caminho expansivo (*upsampling*). A **Figura 1** ilustra esta estrutura.

O caminho contrativo segue a arquitetura convencional de uma rede convolucional. Este segmento da U-Net começa com a aplicação repetida de duas convoluções 3x3 sem preenchimento, seguidas por uma unidade linear retificada (ReLU). A ReLU é fundamental no aprendizado de características não-lineares das imagens; ela opera passando o valor de ativação se for positivo e atribuindo zero se for negativo. Essa simplicidade torna a ReLU uma função de ativação eficiente e rápida, permitindo que as redes neurais convolucionais aprendam mais rapidamente e reduzam a probabilidade de o gradiente desaparecer durante o treinamento (AGARAP, 2018). A função ReLU é definida matematicamente como:

$$\text{ReLU}(x) = \max(0, x). \quad (1.1)$$

Após a ReLU, é aplicada uma operação de *max pooling* 2x2, que é uma forma poderosa de redução de dimensionalidade. A operação de *max pooling* pode ser definida matematicamente para uma sub-região R da entrada como:

$$\text{MaxPooling}(R) = \max_{i,j \in R} X_{i,j}, \quad (1.2)$$

onde $X_{i,j}$ representa o valor do elemento na posição (i, j) da entrada e o máximo é tomado sobre todos os valores na região R . Como explicado por Goodfellow, Bengio e Courville (2016), este processo reduz a sensibilidade da saída para variações exatas de localização da característica e contribui para a generalização ao reduzir o número de parâmetros envolvidos no aprendizado, ajudando a reduzir a ocorrência de sobreajuste (*overfitting*). Sobreajuste acontece quando um modelo de aprendizado de máquina se adapta não apenas aos aspectos gerais dos dados de treinamento, mas também às suas peculiaridades aleatórias. Isso geralmente leva a um desempenho insatisfatório em novos dados, pois o modelo falha em generalizar a partir de sua experiência de treinamento. Goodfellow, Bengio e Courville (2016) descrevem o sobreajuste como um fenômeno central no aprendizado de máquina, ressaltando que é um dos principais obstáculos na criação de modelos robustos e generalizáveis.

A cada etapa de *downsampling*, o número de canais de características é dobrado, permitindo que a rede capture contextos mais complexos e variações de características nas imagens de entrada. No caminho expansivo, cada passo inicia com um *upsampling* do mapa de características, seguido por uma convolução 2x2, conhecida como "*up-convolution*", que reduz pela metade o número de canais de características. Este é então concatenado com o mapa de características correspondente, adequadamente recortado, do caminho contrativo. Seguem-se duas convoluções 3x3, cada uma seguida por uma ReLU. O recorte é necessário para compensar a perda de

pixels nas bordas devido às convoluções. Este processo de *upsampling* e concatenação permite que a rede recupere as informações de localização para segmentação precisa. Na camada final, uma convolução 1x1 é utilizada para mapear cada vetor de características de 64 componentes para o número desejado de classes.

Para possibilitar um mapeamento contínuo do mapa de segmentação de saída, é crucial escolher o tamanho de entrada de tal maneira que todas as operações de *max pooling* 2x2 sejam aplicadas em camadas com tamanhos pares em x e y.

Esses detalhes estruturais da U-Net permitem que ela se destaque na segmentação de imagens médicas, fornecendo uma saída de alta resolução e permitindo a segmentação precisa de estruturas anatômicas e patológicas em diversas modalidades de imagens médicas.

Além das inovações arquitetônicas da U-Net, Ronneberger, Fischer e Brox (2015) reforçam a importância da utilização de técnicas de *data augmentation* (aumento da base de dados) para ensinar a rede as propriedades de invariância e robustez desejadas, especialmente quando há poucas amostras de treinamento disponíveis, ressaltando que a aplicação de transformações aleatórias nas imagens, como transformações elásticas, de intensidade, geométricas e outras, parecem ser o conceito chave para atingir melhores resultados com a U-Net.

1.2.2 O framework nnU-Net

Como citado anteriormente, a área de segmentação de imagens médicas está tomada pelo uso de CNNs, no entanto, parece que cada padrão de segmentação exige arquiteturas específicas e ajustes nos métodos de treinamento para obter um desempenho de alto nível (ISENSEE et al., 2018a; LI et al., 2018). Tal limitação constantemente dificulta o desenvolvimento do trabalho de pesquisa, limitando novas publicações, muitas vezes, a apenas um conjunto de dados e tornando-se uma barreira na identificação de novos métodos (ISENSEE et al., 2018b). Nesse contexto, surge o *framework* nnU-Net (“no-new-U-Net”), desenvolvido inicialmente para a participação de Isensee et al. (2018b) no *Medical Segmentation Decathlon*⁴, desafio internacional dedicado à identificação de um algoritmo de uso geral para segmentação de imagens médicas (ANTONELLI et al., 2022), buscando soluções para a problemática apresentada inicialmente nesta seção. A implementação hoje encontra-se em sua segunda versão e pode ser encontrada no GitHub⁵.

⁴ Decatlo de Segmentação Médica

⁵ <<https://github.com/MIC-DKFZ/nnUNet>>

O nnU-Net baseia-se diretamente em torno da arquitetura U-Net original de Ronneberger, Fischer e Brox (2015). Seu nome é uma abreviação de *no-new-U-Net* (não nova U-Net, em tradução livre). É um *framework* que foi desenvolvido para automatizar a configuração de arquiteturas de rede neural convolucional para tarefas de segmentação em imagens médicas.

O nome nnU-Net carrega uma espécie de mensagem implícita: em vez de projetar uma nova arquitetura de rede neural para cada novo problema de segmentação, o *framework* se adapta automaticamente às características dos dados de entrada (por exemplo, espaçamento entre os *pixels*, tamanho da imagem, formato da imagem, entre outros), selecionando e configurando uma arquitetura de rede neural convolucional já existente, que é a U-Net ou suas variantes. O mesmo expõe as principais configurações em forma de variáveis, permitindo o fino ajuste dos parâmetros de treino. Além das ferramentas para treino de modelos, o nnU-Net fornece ferramentas de pré-processamento, *data augmentation*, validação e pós-processamento.

A base do *framework* consiste em um conjunto de três modelos baseados no U-Net original, adaptando-se automaticamente à geometria do conjunto de imagens, de forma a permitir o treino de modelos utilizando-se de imagens 2D e 3D. Os três modelos base são:

- U-Net 2D;
- U-Net 3D;
- U-Net *cascade*.

Enquanto os modelos 2D e 3D geram segmentações em resolução cheia, o modelo *cascade* gera primeiro segmentações em baixa resolução e vai subsequentemente as incrementando (ISENSE et al., 2018b). Para este trabalho, foi escolhido a abordagem do modelo 2D, devido à sua capacidade de treinar imagens em formato PNG.

O treinamento realizado pelo nnU-Net é baseado no *five-fold cross-validation*, técnica de validação cruzada usada para avaliar a habilidade de generalização de um modelo estatístico, isto é, quão bem o modelo performa em um conjunto de dados independente do que foi usado para treiná-lo. Ela é especialmente útil em situações onde o conjunto de dados é limitado (WONG; YEH, 2019). Para tanto, o conjunto de imagens de treino é separado em cinco *folds* (subconjuntos geralmente de mesmo tamanho). Cinco modelos são treinados, utilizando-se em cada um deles um dos *folds* como conjunto de validação e os outros quatro *folds* para treino. Após o treino dos modelos, o *framework* faz a avaliação a partir dos parâmetros e métricas dos treinamentos de cada modelo, buscando identificar a configuração final ideal para ser usada na ge-

ração das predições. Ao final do processo, o *framework* nnU-Net aplica um pós-processamento nas imagens segmentadas (previsões), buscando otimizar o resultado final.

Além disso, durante o treinamento, o *framework* gera diversas métricas que ajudam na análise dos resultados, como tempo de *epoch*, métricas de perda de validação e treinamento, cálculo de pseudo-dice em tempo real, entre outros. Opcionalmente, o mesmo permite a integração do treinamento com a biblioteca *hiddenlayer*, disponível no GitHub⁶, a qual fornece utilidades gráficas para CNNs, como *PyTorch*, *Tensorflow* e *Keras*, principais bibliotecas para implementação de CNNs. Tais utilitários foram usadas no presente trabalho para gerar os gráficos das métricas dos treinamentos e uma visualização da arquitetura da rede gerada a partir das configurações do projeto.

1.3 Coeficiente Dice

O Coeficiente Dice é uma métrica essencial para a avaliação de modelos de segmentação de imagens, particularmente na área de processamento de imagens médicas. Ele mede a similaridade entre duas amostras e é definido como o tamanho do corte entre duas conjuntos dividido pelo total tamanho dos dois conjuntos. Matematicamente, o Coeficiente Dice pode ser representado pela seguinte fórmula para duas amostras, A e B:

$$DICE(A,B) = \frac{2 \times |A \cap B|}{|A| + |B|}, \quad (1.3)$$

onde $|A \cap B|$ é o número de elementos comuns a ambos os conjuntos, e $|A|$ e $|B|$ são, respectivamente, o número de elementos em cada conjunto. Um valor de 1 indica perfeita concordância entre as duas amostras, enquanto um valor de 0 indica nenhuma sobreposição (DICE, 1945).

A utilização desta métrica é crucial para validar a precisão das técnicas de segmentação, assegurando que os contornos das estruturas de interesse sejam precisamente delineados. A capacidade de quantificar a sobreposição entre a segmentação proposta pelo algoritmo e a verdade de base oferece uma avaliação objetiva da performance do modelo.

No presente trabalho, o Coeficiente Dice será adotado como métrica de avaliação para o modelo de segmentação de imagens desenvolvido. Sua aplicação visa quantificar a exatidão das segmentações em relação às referências padrão, permitindo uma avaliação objetiva e a pri-

⁶ <<https://github.com/FabianIsensee/hiddenlayer>>

moramento contínuo do desempenho dos modelos. A meta é obter um alto Coeficiente Dice, evidenciando a precisão e a confiabilidade das técnicas de segmentação empregadas

1.4 *Intersection over Union*

A métrica *Intersection over Union* (IoU), também conhecida como *Jaccard Index*, é outra ferramenta estatística vital para avaliar a precisão de modelos de segmentação de imagens. Similar ao Coeficiente Dice, o IoU oferece uma medida quantitativa da sobreposição entre duas áreas segmentadas. A métrica IoU é calculada como a área de interseção entre a previsão do modelo e a verdade de base, dividida pela área da união das duas segmentações:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (1.4)$$

onde $|A \cap B|$ representa a interseção e $|A \cup B|$ a união dos conjuntos de *pixels* segmentados da verdade de base A e da previsão B . Valores de IoU mais próximos de 1 indicam uma maior precisão na segmentação, enquanto valores mais baixos indicam uma precisão menor (FLETCHER; ISLAM et al., 2018).

No contexto deste estudo, a IoU será empregada como uma métrica complementar ao Coeficiente Dice para avaliar o desempenho dos modelos de segmentação.. Almeja-se alcançar valores de IoU que demonstrem a eficácia das abordagens de segmentação propostas.

1.5 Comentários Finais

Neste capítulo foram introduzidos os principais conceitos abordados no presente trabalho. Essa base teórica estabelece o alicerce para nossa investigação, destacando a relevância e a necessidade de avanços na área. Isso incluiu a contextualização da sacroilíte, processo inflamatório não-infeccioso da articulação sacroilíaca, e as dificuldades diagnósticas relacionadas a tal patologia. Foi ressaltado como a identificação precoce é fundamental e como buscamos usar ferramentas de aprendizado de máquina para explorar a aplicabilidade das mesmas no auxílio diagnóstico da sacroilíte.

Ademais, foi apresentado a rede neural U-Net e sua arquitetura inovadora juntamente com o *framework* nnU-Net e suas ferramentas para implementação de um fluxo completo de treinamento para modelos utilizando variantes da rede U-Net original.

Ao final do capítulo, conceitualizamos as principais métricas avaliativas utilizadas como indicadores da exatidão das segmentações em relação às referências padrão. Tais métricas incluem o coeficiente Dice e IoU, ambos apresentando capacidade de quantificar a sobreposição entre a segmentação proposta pelo algoritmo e a verdade da base.

No capítulo seguinte, delinearemos os objetivos específicos deste estudo, estabelecendo um caminho claro para a análise e experimentação desenvolvidos.

2 OBJETIVOS

Neste capítulo, vamos estabelecer os objetivos específicos do presente trabalho, definindo assim uma rota clara para a análise e experimentação que se seguirá.

2.1 Objetivos Gerais

O objetivo geral deste Trabalho de Conclusão de Curso é implementar, treinar e analisar os resultados da aplicação da rede U-Net, implementada através do *framework* nnU-Net, para a identificação e segmentação automática das articulações sacroilíacas (AS) em imagens de ressonância magnética (RM). Busca-se, através deste estudo, endereçar o desafio de automatizar o processo de detecção das referidas articulações.

2.2 Objetivos Específicos

Os principais objetivos específicos deste trabalho, são divididos nos seguintes tópicos:

- O Pré-processamento da base de imagens, no qual incluem a padronização das mesmas (tamanhos, intensidades de *pixels*, entre outros);
- O enriquecimento da base de dados, através do uso de técnicas de *data augmentation*;
- A organização da base, onde a paridade das imagens e máscaras é verificada, junto com a presença de características necessárias nas imagens. Essa etapa também inclui a organização da base de imagens no formato de diretórios correto para a aplicação da nnU-Net;
- A implementação do *framework* e seu treinamento, incluindo ajustes realizados aos parâmetros e testes;
- O pós-processamento das previsões do modelo final.
- A validação da eficácia e precisão do modelo de segmentação proposto através de uma análise dos resultados.

2.3 Comentários Finais

Ao final deste capítulo, destacamos a importância de cada objetivo delineado, tanto os gerais quanto os específicos, no desenvolvimento do estudo. A implementação, treinamento e análise da rede U-Net, por meio do *framework* nnU-Net, representam passos fundamentais para enfrentar o desafio da detecção automática das articulações sacroilíacas em imagens de ressonância magnética. Cada objetivo específico, desde o pré-processamento da base de imagens até a validação final do modelo, contribui significativamente para a precisão e eficácia do processo de segmentação. Estas etapas não apenas viabilizam a aplicação prática do modelo, mas também asseguram a confiabilidade dos resultados obtidos.

No capítulo seguinte, serão apresentados os materiais e métodos utilizados neste trabalho, oferecendo um aprofundamento detalhado nas técnicas e abordagens adotadas para alcançar os objetivos aqui estabelecidos.

3 MATERIAIS E MÉTODOS

Este capítulo detalha os materiais e métodos empregados no estudo, proporcionando uma visão profunda das técnicas e estratégias utilizadas para atingir os objetivos previamente definidos. A importância deste capítulo reside na compreensão abrangente da implementação prática do projeto, incluindo desde a escolha dos materiais até as metodologias específicas adotadas no processamento das imagens e no treinamento dos modelos.

3.1 Conjunto de Imagens

O material principal deste trabalho é o conjunto de imagens de ressonância magnética (RM) da articulação sacroilíaca, junto com suas respectivas máscaras de segmentação manual. Estas imagens são de 51 pacientes obtidas com sequência Spectral Attenuated Inversion Recovery (SPAIR) e de 46 pacientes obtidas com sequência Short Tau Inversion Recovery (STIR), cada exame foi classificado por um especialista como positivo ou negativo para sacroilíte ativa, totalizando 22 exames positivos e 29 negativos para SPAIR e 20 exames positivos e 26 negativos para STIR. Todas as imagens foram segmentadas manualmente por especialistas em um fundo escuro para separar a região de interesse.

Este conjunto foi anteriormente estruturada e anonimizada para uso público, com aprovação do Comitê de Ética em Pesquisa do Hospital das Clínicas de Ribeirão Preto (CEP-HCRP), sob parecer número 1.951.052.

3.2 Pré-processamento

Para o desenvolvimento deste trabalho, o primeiro passo necessário foi o pré-processamento das imagens de RM obtidas no formato *DICOM*. Inicialmente, todas as imagens foram iteradas, com uma verificação detalhada da presença de atributos essenciais para o processamento adequado, sendo eles *WindowCenter*, *WindowWidth*, *RescaleSlope* e *RescaleIntercept*. Seguindo a detecção desses atributos, as imagens foram convertidas para o formato PNG, resultando em imagens em tons de cinza.

Foi empregada a técnica de *windowing* (GONZALEZ, 2009), um procedimento que realça faixas específicas de intensidades nas imagens médicas. Isso foi realizado ajustando-se os va-

lores dos *pixels* da imagem DICOM usando os parâmetros de *WindowCenter* e *WindowWidth* para definir o ponto médio e a gama de intensidades visuais, respectivamente. A inclinação (*RescaleSlope*) e o intercepto (*RescaleIntercept*) foram aplicados para transformar os valores brutos dos *pixels* em intensidades de imagem normalizadas. A aplicação de *windowing* permitiu mapear os valores de intensidade dos *pixels* dentro da janela especificada para o espectro completo de tons de cinza (0 a 255), enquanto valores fora desta janela eram convertidos para o preto ou branco absoluto, melhorando assim o contraste e destacando as estruturas de interesse.

Ao final desse processo, constatou-se que 18 imagens STIR tiveram que ser removidas do conjunto de dados devido à ausência dos atributos necessários, os quais são imprescindíveis para a realização do pré-processamento com a técnica de *windowing*.

Antes de seguir, foi verificado a paridade das imagens e máscaras de segmentação, onde todas imagens foram iteradas, assegurando-se a existência, no conjunto de imagens, da respectiva máscara. Uma imagem SPAIR foi removida da base de imagens por falta de máscara de segmentação.

As máscaras também foram pré-processadas, através da conversão de qualquer *pixel* com valor diferente de 0 (preto) em 255 (branco), com a finalidade de garantir a consistência das anotações das máscaras, sendo 0 o fundo e 255 a área de interesse. Ainda, passaram por um processo de erosão e dilatação, afim de remover quaisquer ruídos contidos nas mesmas e, após uma confêrencia manual para checar o correto pré-processamento, foram binarizadas, convertendo a área de interesse no valor 1.

Por fim, os tamanhos de ambas foram padronizados, em 384x384, realizando-se um corte nas poucas imagens que eram maiores que essas dimensões e colocando as que eram menores dentro de um fundo escuro, resultando em imagens como o exemplo na **Figura 2**. O conjunto final de imagens conta com um total de 569 pares de imagens e máscaras. Segundo Hadjiiski et al. (2023), idealmente os conjuntos de imagens devem ser separados em três sub-conjuntos: conjunto de treino, usado para o treinamento do modelo; Conjunto de validação, usado para guiar a otimização do modelo (*tunning*); E conjunto de testes, usado para testar o modelo após o treinamento. Por essa razão, o conjunto utilizado nesse trabalho foi separado em 80% para treino e 20% para teste, resultando em 455 imagens para treino e 114 para testes. O conjunto validação será discutido na seção **Implementação da Arquitetura**, pois é tratado pela implementação do nnU-Net.

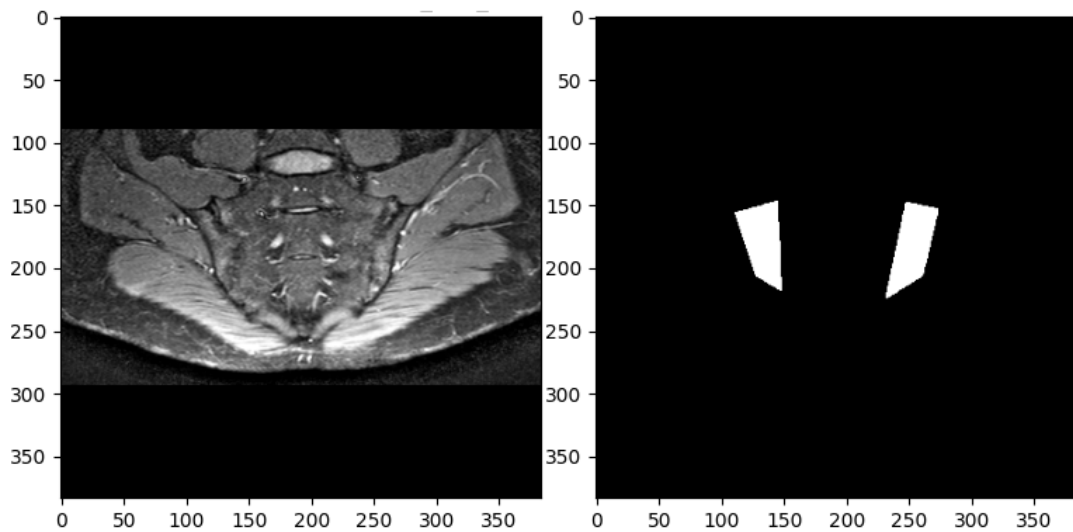


Figura 2 – Exemplo das imagens após o pré-processamento.

3.3 *Data Augmentation*

Como citado anteriormente, o enriquecimento do conjunto de imagens a partir de técnicas de *data augmentation*¹ é um conceito fundamental para atingir melhores resultados no treinamento de modelos U-Net. Apesar do nnU-Net possuir ferramentas próprias para tanto, neste trabalho, tais ferramentas não foram exploradas. Primeiramente, por falta de documentação adequada nesse sentido, dificultando a implementação, como também o fato de *data augmentation* já ter sido aplicado anteriormente ao conjunto, utilizando a biblioteca de *Python Albumentations* (BUSLAEV et al., 2020). Foram aplicadas transformações apenas ao conjunto de treino, inicialmente com a intenção de aumentar em 5 vezes o conjunto original, mas que, após testes (que serão discutidos na próxima seção), foi necessário manter apenas um aumento de 2 vezes do conjunto original, resultando em 910 imagens e máscaras para treino. Todas as transformações são aplicadas igualmente entre as imagens e máscara, gerando máscaras que correspondam às imagens transformadas.

As transformações aplicadas foram:

- **HorizontalFlip:** com uma probabilidade de 0.5, esta transformação espelha a imagem horizontalmente.
- **VerticalFlip:** com uma probabilidade de 0.5, espelha a imagem verticalmente;

¹ Aumento de Dados

- **Rotate**: rotaciona a imagem dentro de um limite de 30 graus com uma probabilidade de 0.5;
- **RandomSizedCrop**: recorta a imagem aleatoriamente em um tamanho entre 80% e 100% da altura original, mantendo a altura e largura especificadas, com uma probabilidade de 0.5;
- **GridDistortion**: distorce a imagem em uma grade com uma probabilidade de 0.4;
- **GaussNoise**: adiciona ruído gaussiano à imagem com uma probabilidade de 0.4.

A **Figura 3** contém exemplos das transformações aplicadas durante o processo de *data augmentation*.

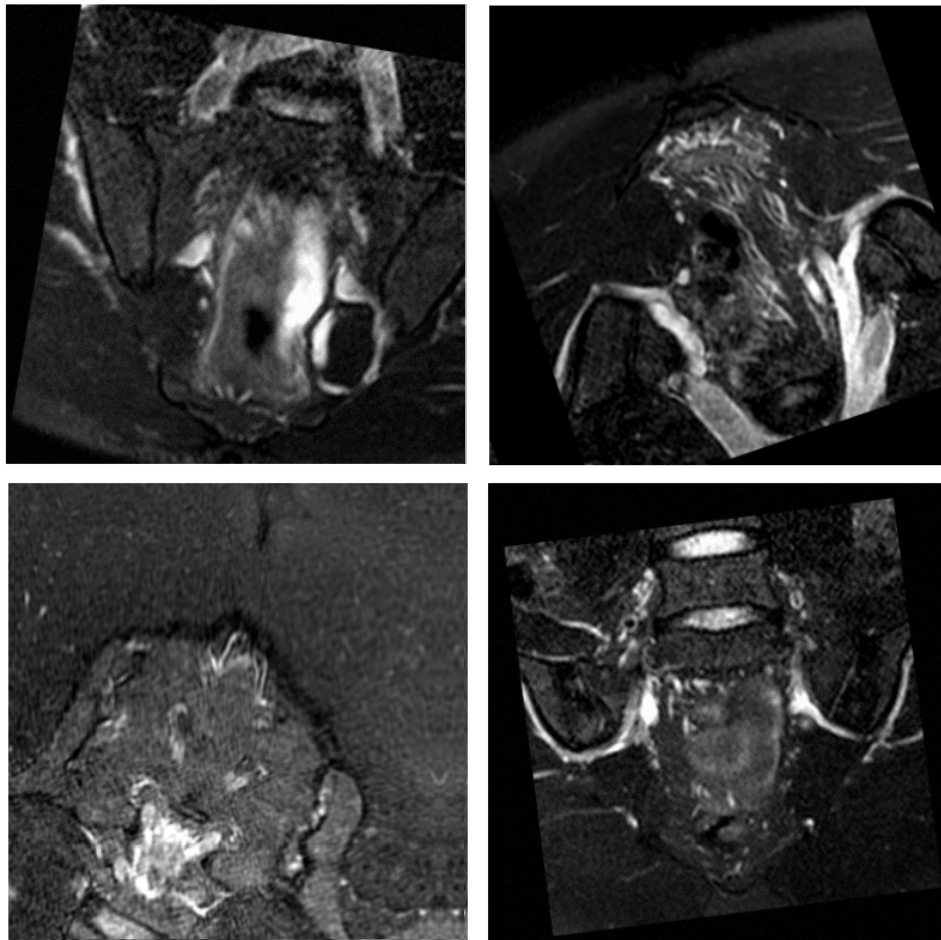


Figura 3 – Exemplos de imagens resultantes após o processo de *data augmentation*.

3.4 Implementação da Arquitetura

3.4.1 Configuração e Testes

O primeiro passo para o treinamento, foi a organização do conjunto de imagens. Para a utilização do mesmo pelo nnU-Net, foi necessário organiza-lo em um formato de diretório específico, baseado no formato utilizado no *Medical Segmentation Decathlon*, com uma pasta raiz, nomeada *nnUNet_raw*, de onde o nnU-Net consegue ler as imagens e máscaras; Uma pasta específica para cada base de imagens, contendo as subpastas *ImagesTr*, *LabesTr* com as imagens e máscaras para treino, respectivamente e opcionalmente, a subpasta *LabelsTs*, com as máscaras para teste; Em conjunto com um arquivo no formato JSON, contendo as configuração desejadas para cada base de imagens, como ilustrado na **Figura 4**.

```
nnUNet_raw/Dataset001_SacroRM/
├── dataset.json
├── imagesTr
│   ├── SacroRM_001_0000.png
│   ├── SacroRM_001_0001.png
│   ├── SacroRM_001_0002.png
│   ├── SacroRM_001_0003.png
│   ├── SacroRM_002_0000.png
│   ├── SacroRM_002_0001.png
│   ├── SacroRM_002_0002.png
│   ├── SacroRM_002_0003.png
│   └── ...
├── imagesTs
│   ├── SacroRM_485_0000.png
│   ├── SacroRM_485_0001.png
│   ├── SacroRM_485_0002.png
│   ├── SacroRM_485_0003.png
│   ├── SacroRM_486_0000.png
│   ├── SacroRM_486_0001.png
│   ├── SacroRM_486_0002.png
│   ├── SacroRM_486_0003.png
│   └── ...
└── labelsTr
    ├── SacroRM_001.png
    ├── SacroRM_002.png
    └── ...
```

Figura 4 – Estrutura de diretórios para uso da nnU-Net.

Para o treinamento, optou-se pelo uso do *Google Colab* devido à necessidade intensiva de GPU. Assim, fez-se necessário a subscrição do seu plano mensal para atender a esta demanda. O nnU-Net tem por padrão algumas configurações, como o número de iterações por *epoch*², numero total de *epochs* para o treinamento, tamanho dos *kernels* de convolução e *max-pooling*,

² Passagem completa de todo o conjunto de dados de treinamento pela rede durante o processo de treinamento

*learning rate*³ inicial, entre outros. A grande maioria dessas configurações foi mantida em seu valor padrão, modificando apenas o número total de *epochs* e o número de iterações por *epochs* com base no tempo médio do mesmo. O Objetivo foi buscar um tempo médio de tal forma que fosse possível logisticamente realizar o treinamento. Com os valores padrão do nnU-Net, o tempo médio estava em torno de 8 minutos, e com o padrão de 1000 *epochs* por treinamento, tornaria invável a aplicação do mesmo. Por tanto, chegamos ao valor ideal de 250 iterações por *epoch*, baixando seu tempo médio para 5 minutos. Quanto ao número total de *epochs*, o valor escolhido foi de 80. Tal valor foi escolhido devido à observação de que próximo desse número, o treino começou a evoluir muito lentamente, como será discutido na próxima subseção.

3.4.2 *Treinamento*

O treinamento foi aplicado a cinco modelos, seguindo a técnica de validação cruzada *5-Fold Cross Validation*. O conjunto de testes, com 910 imagens, foi dividido em 5 subconjuntos, e em cada modelo, um desses subconjuntos foi utilizado para validar o treinamento e os outros quatro subconjuntos foram utilizados para treino. Portanto, cada um dos cinco modelos foi treinado com 728 imagens e validado com 182.

Como discutido anteriormente, os parâmetros dos modelos foram ajustas buscando otimizar o tempo de processamento. Como pode ser observado na **Figura 5**, o tempo médio dos *epochs* ficou em torno de 300 segundos (5 minutos).

A decisão do número total de *epochs* se deve à observação de que a taxa de aprendizagem do modelo se degrada muito quando se aproxima de 80 *epochs*, como pode ser observado na **Figura 6**. Isso indica que o modelo alcança um ponto em que ajustes adicionais nos parâmetros não resultam em melhorias significativas na performance de validação. A literatura sugere que o uso de decaimento da taxa de aprendizagem ajuda a minimizar o sobreajuste e permite que os parâmetros da rede encontrem um ponto ótimo mais eficazmente (VALOVA et al., 2020), (BUKHARI; DIN, 2021)). A implementação dessa estratégia no treinamento do modelo atual contribuiu para a decisão de limitar o número de *epochs* e evitar o prolongamento desnecessário do treinamento sem ganhos de desempenho.

Outros fatores que foram decisivos para se chegar a esse número de *epochs* incluem a análise da progressão das métricas de perda durante o treinamento. Como pode ser observado

³ Taxa de aprendizagem - determina o tamanho dos passos que o algoritmo de otimização dá na direção do mínimo da função de custo (ou de perda) durante o treinamento (ISENSE et al., 2018b)

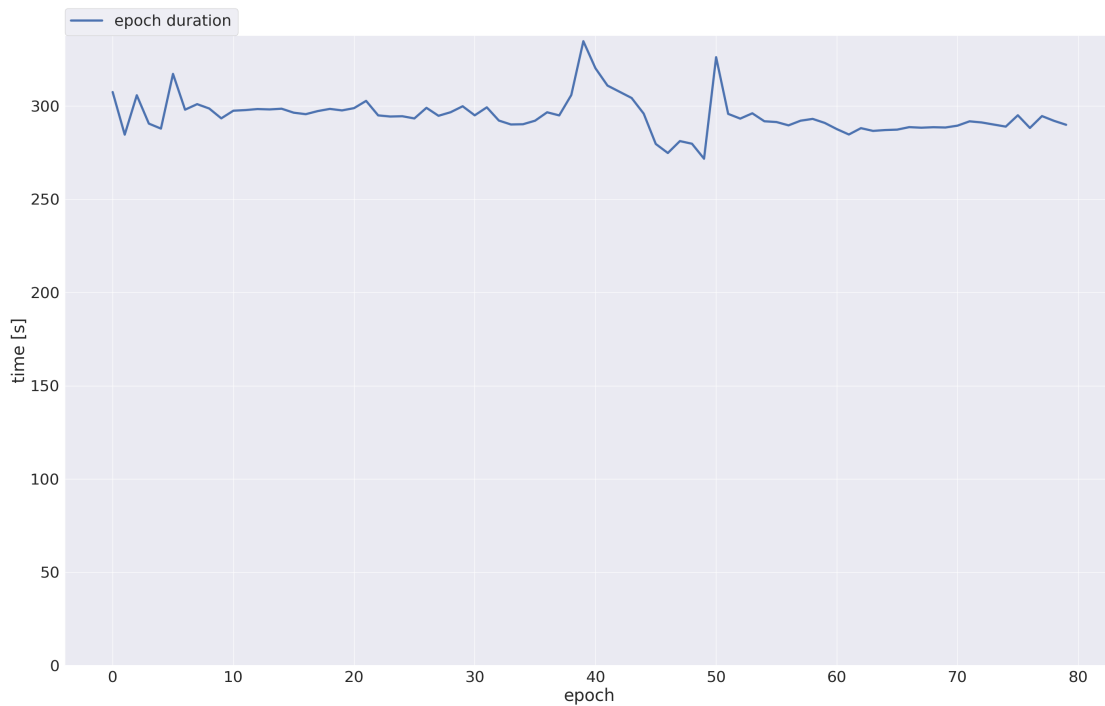


Figura 5 – Duração dos *epochs*.

na **Figura 7**, que ilustra a evolução das perdas de treinamento e validação, ambas as métricas mostram uma tendência de diminuição à medida que o número de *epochs* aumenta, com uma convergência para valores estáveis após aproximadamente 20 *epochs*. Este comportamento indica um bom ajuste do modelo aos dados, sem evidências de overfitting significativo (SHI et al., 2021). Além disso, a métrica Pseudo Dice, representada no gráfico, alcança um patamar elevado e mantém-se estável, refletindo a alta precisão do modelo na tarefa de segmentação. A média móvel do Pseudo Dice também demonstra pouca variação, o que sugere que o modelo mantém um desempenho consistente ao longo das épocas. Estes resultados foram fundamentais para definir o número total de *epochs* para o treinamento dos modelos.

Ao final, o treinamento dos cinco modelos teve a duração de aproximadamente 35 horas. As média dos coeficientes Dice aplicado ao conjunto de validação obtidas em cada modelo pode ser observada na **Tabela 1**.

Tabela 1 – Médias dos coeficientes Dice para cada modelo após treinamento.

Modelo	Coefficiente Dice
Fold_0	0.9177
Fold_1	0.9144
Fold_2	0.9126
Fold_3	0.9078
Fold_4	0.9136

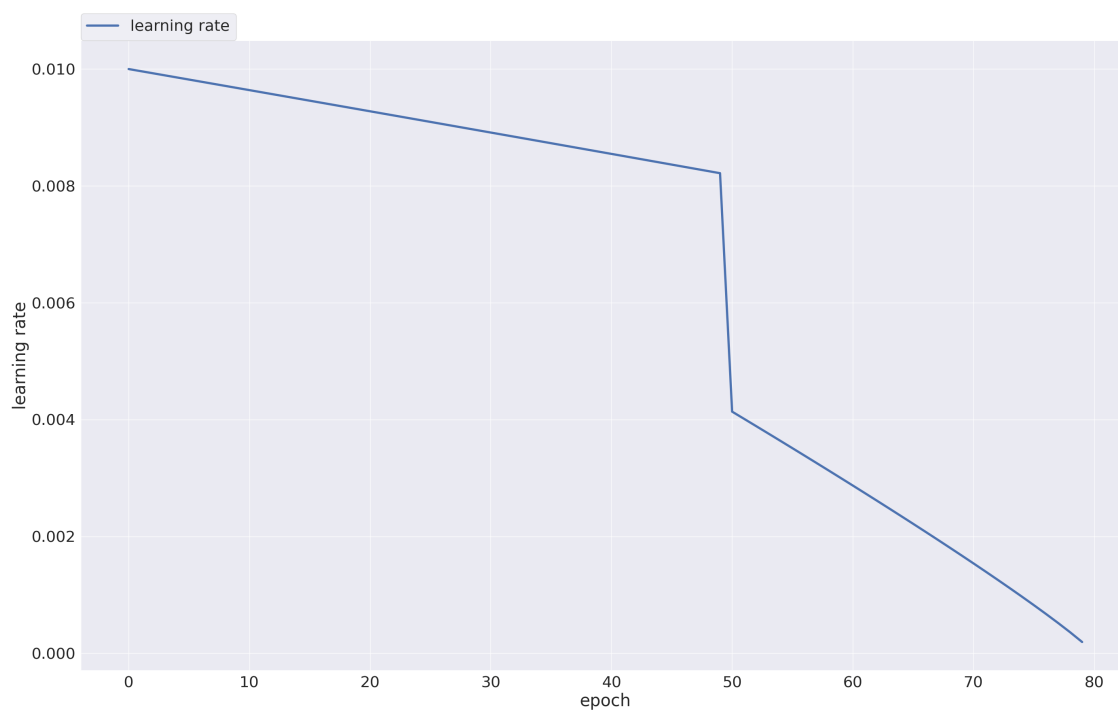


Figura 6 – Progressão da taxa de aprendizagem durante o treinamento do modelo.

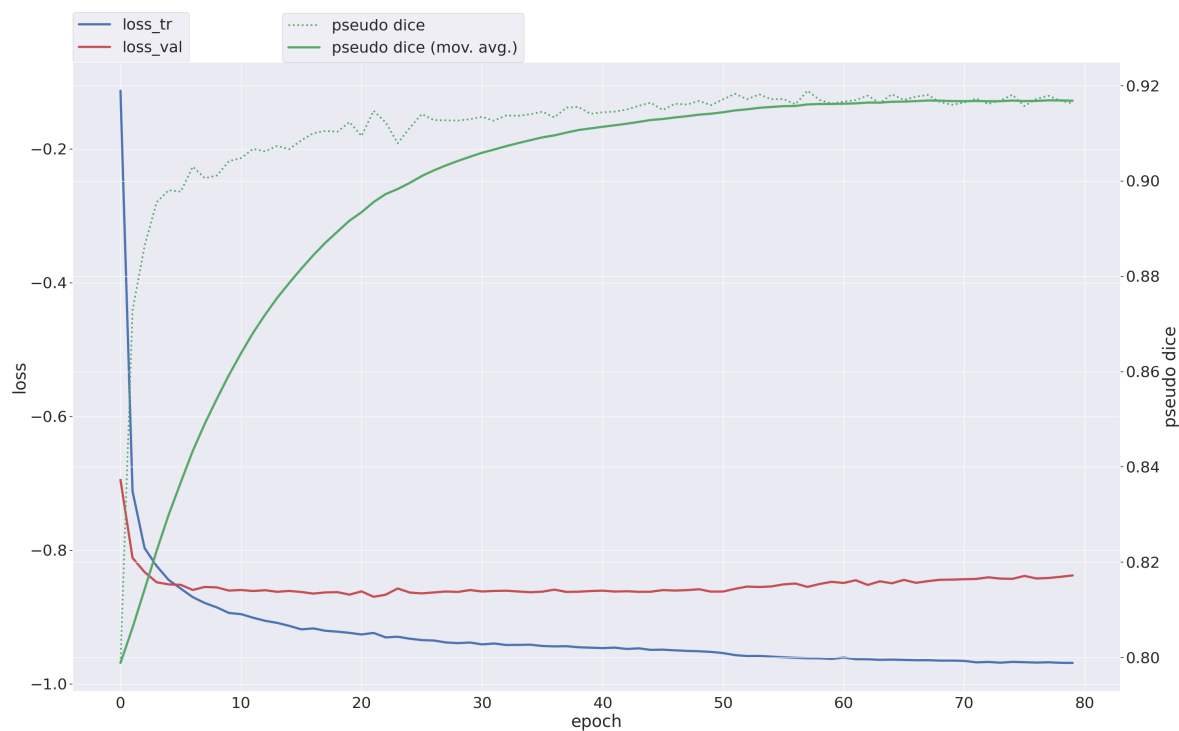


Figura 7 – Evolução das perdas de treinamento e validação ao longo dos *epochs*.

3.4.3 5-Fold Cross Validation

Após o treinamento, foi aplicada a validação cruzada, onde a performance média do modelo é calculada a partir dos resultados dos treinamentos com cada um dos *folds* como conjunto de validação. Isso fornece uma estimativa mais robusta da performance do modelo, pois é baseada em múltiplas divisões dos dados e reduz a variância dos resultados em comparação com a validação cruzada com uma única divisão (WONG; YEH, 2019). A partir desse resultado, o nnU-Net busca a configuração ideal para realizar as previsões e inferências. Os Valores obtidos a partir da validação cruzada podem ser observados na **Tabela 2**.

Tabela 2 – Metricas obtidas após a validação cruzada.

Coeficiente Dice	IoU
0.9132	0.8417

3.4.4 Pós-processamento

Para concluir, o nnU-Net aplica o pós-processamento. Se ele identificar múltiplos componentes conectados para uma classe que, segundo a análise, geralmente forma um único componente, ele automaticamente remove todos os componentes menores, deixando apenas o maior. Esse passo é feito na suposição de que os componentes menores são provavelmente erros de segmentação (ruídos), e que a estrutura real que está sendo segmentada é o componente maior (ISENSEE et al., 2018b).

3.5 Comentários Finais

Ao final deste capítulo, foi introduzido o principal material utilizado no trabalho, o conjunto de imagens de RM da articulação sacroilíaca.

Ademais, foi apresentado a metodologia e técnicas utilizadas, desde o pré-processamento das imagens, *data augmentation*, até detalhes técnicos da implementação do treinamento. Nas metodologias e técnicas incluem-se as nuances de configuração e testes para determinar os parâmetros utilizados; O processo de treinamento dos modelos, incluindo a técnica de validação cruzada aplicada e o pós-processamento realizado.

No capítulo seguinte serão apresentados os resultados obtidos após a implementação dos modelos, incluindo uma análise do mesmo.

4 RESULTADOS

O presente capítulo apresentará os resultados obtidos ao longo do desenvolvimento do projeto, primeiramente analisando os resultados de forma visual, e posteriormente, fazendo uma análise quantitativa dos resultados, usando ferramentas estatísticas. Ao final, incluímos uma seção com resultados adicionais, demonstrando a aplicabilidade prática do projeto.

4.1 Análise Visual dos Resultado

As figuras a seguir são exemplos das predições realizadas pelo modelo final, e serão usadas para discutir os resultados obtidos. Todas foram geradas a partir do treinamento de 5 modelos, seguindo a técnica de *5-Fold cross validation*. A performance de cada modelo treinado em um *fold* foi avaliada com base em métricas relevantes para a tarefa de segmentação, como a Dice ou a IoU. Isso forneceu uma estimativa robusta da performance do modelo. Após a validação, o nnU-Net aplicou um pós-processamento, buscando a melhor configuração. Com isso foram removidos componentes pequenos previstos, com base nas métricas coletadas, ajudando a garantir que o modelo seja geral e robusto, características essenciais para aplicações médicas onde a precisão e a confiabilidade são de extrema importância.

Primeiramente, podemos notar que uma grande parte do conjunto de testes foi predito com sucesso, obtendo *scores* de similaridade com as segmentações originais relativamente altos (discutido na próxima subseção). Como exemplo deste fato, podemos observar nas **Figuras 8, 9 e 10**. É apresentado a imagem original, seguida da máscara original, a máscara predita e a máscara predita pós-processada (PP). A **Figura 8** foi o nosso melhor resultado, obtendo Dice 0.953 e IoU 0.909.

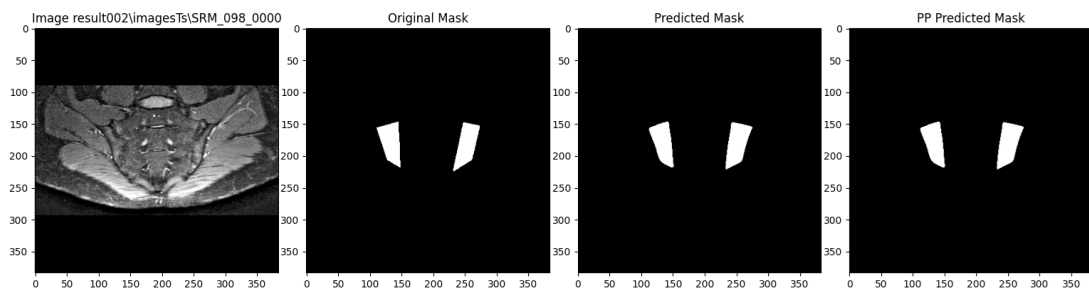


Figura 8 – Exemplo de predição 1.

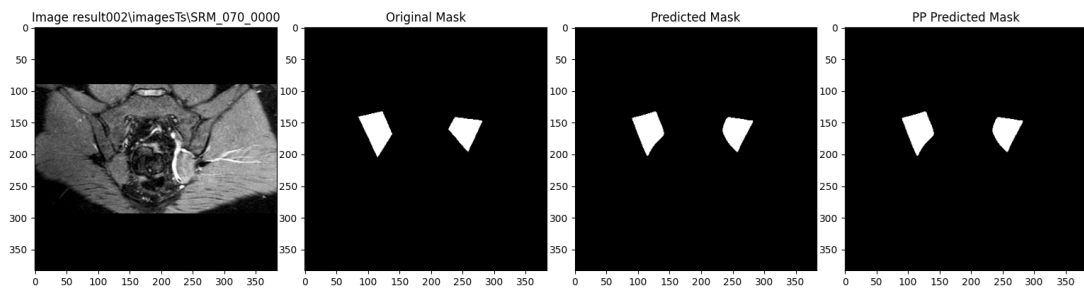


Figura 9 – Exemplo de predição 2.

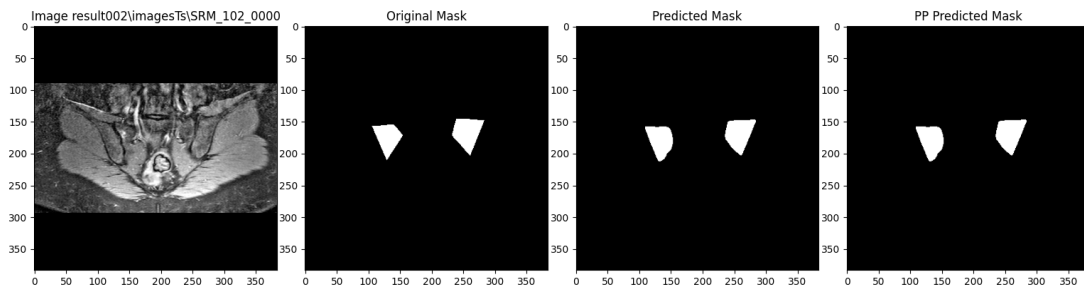


Figura 10 – Exemplo de predição 3.

É válido destacar também que é possível notar uma tendencia do modelo de arredondar as bordas das predições, e como as segmentações originais foram feitas de tal forma que apresentem vértices, isso parece ter alguma relação com o resultado final. Podemos observar esse fato nas **Figuras 11 e 12**.

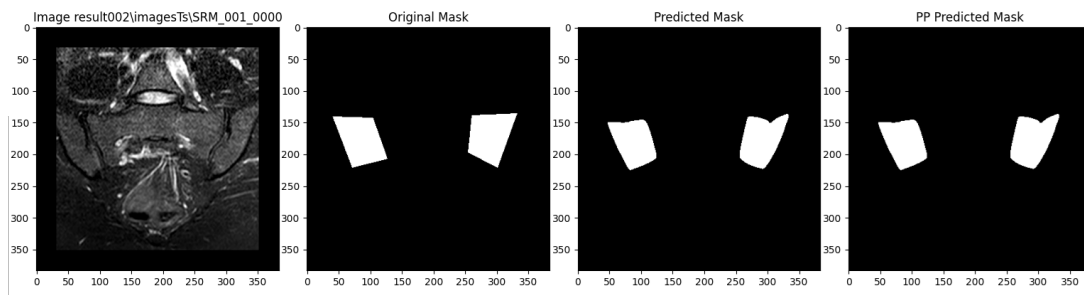


Figura 11 – Exemplo de predição arredondada 1.

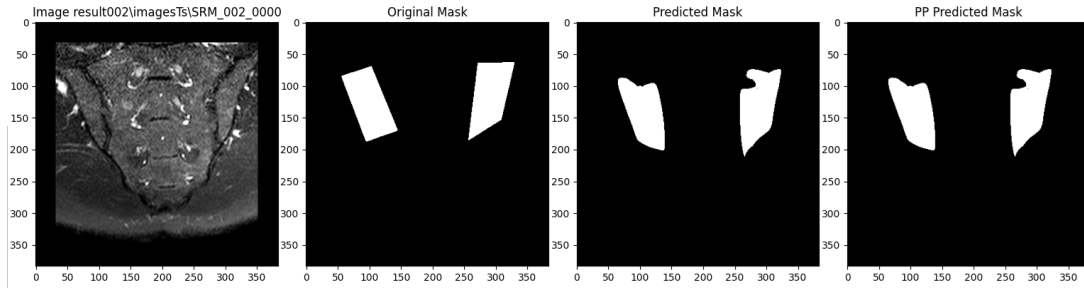


Figura 12 – Exemplo de predição arredondada 2.

Algumas poucas imagens tiveram baixos *scores* de similaridade. Especula-se que há a possibilidade de melhorar tais resultados, trabalhando com uma gama mais diversificada de transformações no conjunto de imagens e explorando os parâmetros de configuração. Exemplos desses resultados podem ser observados nas **Figuras 13 e 14**. A **Figuras 14** em específico foi o nosso pior resultado, atingindo Dice 0.399 e IoU 0.249. Na mesma conseguimos visualizar que o modelo buscou uma região incorreta (na parte inferior da imagem), mas também é notável um certa semelhança da região incorretamente apontada pelo modelo com as características da verdadeira região de interesse.

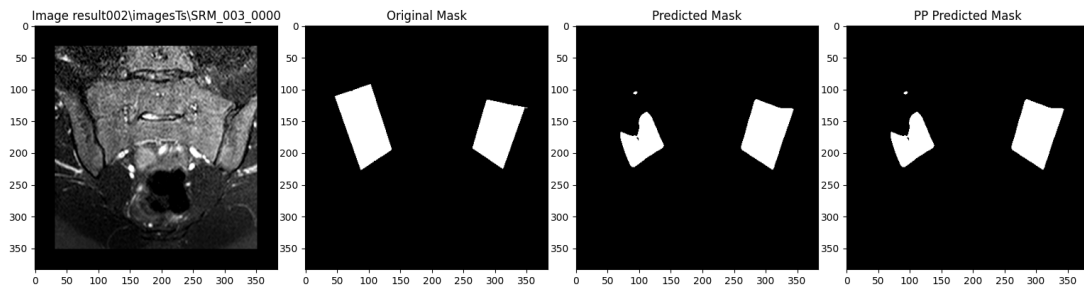


Figura 13 – Exemplo de e predição com falhas 1.

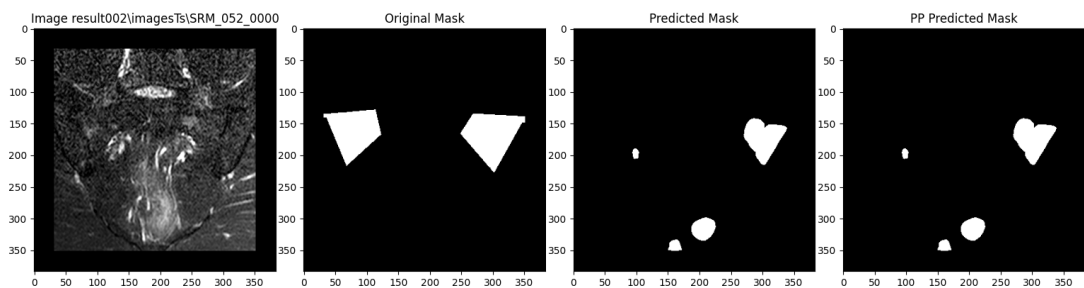


Figura 14 – Exemplo de e predição com falhas 2.

4.2 Análise Quantitativa dos Resultados

As figuras a seguir são os *Box-plots* das distribuições tanto dos Coeficientes Dice obtidos (**Figura 15**) quanto dos IoU obtidos (**Figura 16**) no conjunto de testes.

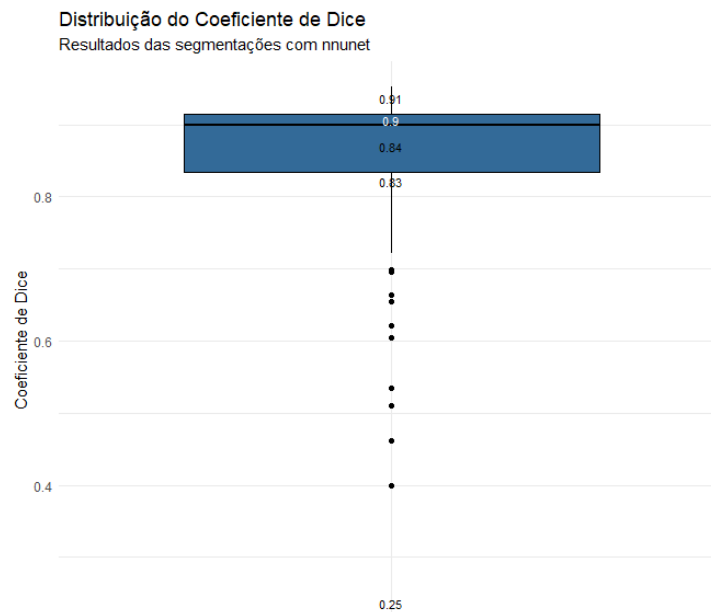


Figura 15 – Box-plot da distribuição dos coeficientes Dice obtidos.

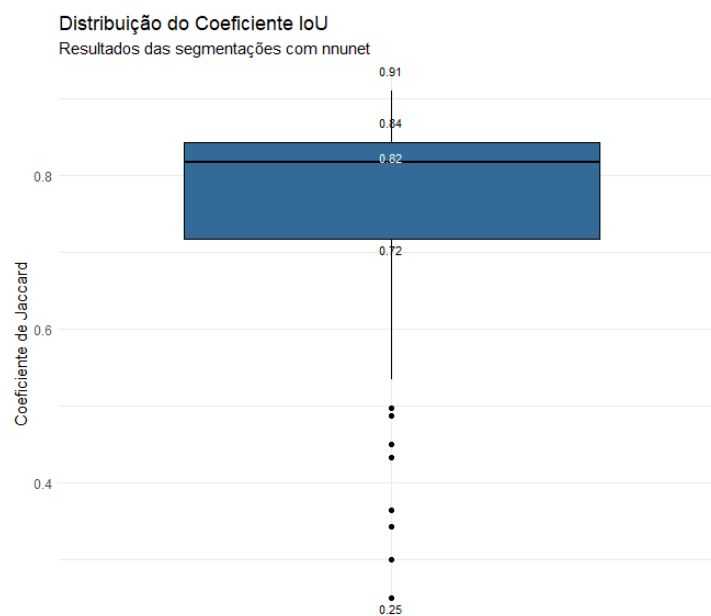


Figura 16 – Box-plot da distribuição dos coeficientes IoU obtidos.

O resumo dos valores mais relevantes encontram-se na **Tabela 3** (Dice) e na **Tabela 4** (IoU).

Tabela 3 – Resumo dos Coeficientes Dice Obtidos.

	Min.	1º Quadrante	Mediana	Média	3º Quadrante	Max.
Dice	0.3996	0.8348	0.8992	0.8542	0.9147	0.9528

Tabela 4 – Resumo dos IoU obtidos.

	Min.	1º Quadrante	Mediana	Média	3º Quadrante	Max.
IoU	0.2497	0.7164	0.8168	0.7576	0.8429	0.9099

Além disso, também foi construído um Intervalo de Confiança para a média, tanto da Coeficiente Dice (**Tabela 5**), quanto do IoU (**Tabela 6**), utilizando-se do teste *t student*, para 95% de confiança.

Tabela 5 – Intervalo de Confiança para a média da amostra de Coeficientes Dice, com 95% de confiança.

Limite Inferior	Limite Superior
0.835	0.873

Tabela 6 – Intervalo de Confiança para a média da amostra de Coeficientes IoU, com 95% de confiança.

Limite Inferior	Limite Superior
0.732	0.782

Podemos inferir a partir dos resultados estatísticos apresentados que:

- A variação entre o mínimo e o máximo é relativamente grande. Isso indica que o desempenho do modelo pode variar consideravelmente dependendo da imagem. Alguns casos são segmentados com alta precisão, enquanto outros apresentam resultados insatisfatórios;
- A mediana e a média sugerem que, em geral, o modelo está realizando uma boa segmentação na maioria das imagens, já que valores mais próximos de 1 indicam uma alta sobreposição entre a segmentação prevista e a verdadeira;
- Para os resultados do coeficiente Dice, o fato do 1º quadrante estar em 0.8348 e o 3º quadrante em 0.9147 mostra que 50% dos dados têm scores entre esses dois valores. Isso sugere uma consistência razoável na performance para a maioria dos dados, com a metade inferior dos resultados ainda obtendo um score acima de 0.8348. Os resultados dos Intervalos de Confiança construídos também corroboram com essa inferência;

- O mínimo de 0.3996 é um *outlier* em relação aos outros valores, que são todos substancialmente mais altos. Isso pode indicar que há casos particulares nos quais o modelo está tendo dificuldades significativas. Esses casos podem ser devido a características particulares das imagens, como ruído, variações de contraste ou formas atípicas que não foram bem representadas no conjunto de treinamento, conforme sugerido na análise visual realizada anteriormente

4.3 Resultados Adicionais

Como um extra aos resultados obtidos ao longo do desenvolvimento deste trabalho, foi adicionado uma funcionalidade para gerar *bounding box* (Caixa delimitadora) às imagens de teste, como pode ser visualizado na **Figura 17**. O objetivo dessa funcionalidade é demonstrar como, ao final de todo o processo realizado, seria possível aplicar o que foi desenvolvido ao contexto clínico. Dessa forma, é aplicada às predições uma função que busca os pontos mais externos de cada região anotada como de interesse, identificando tais pontos e aplicando à imagem na forma de sobreposição.

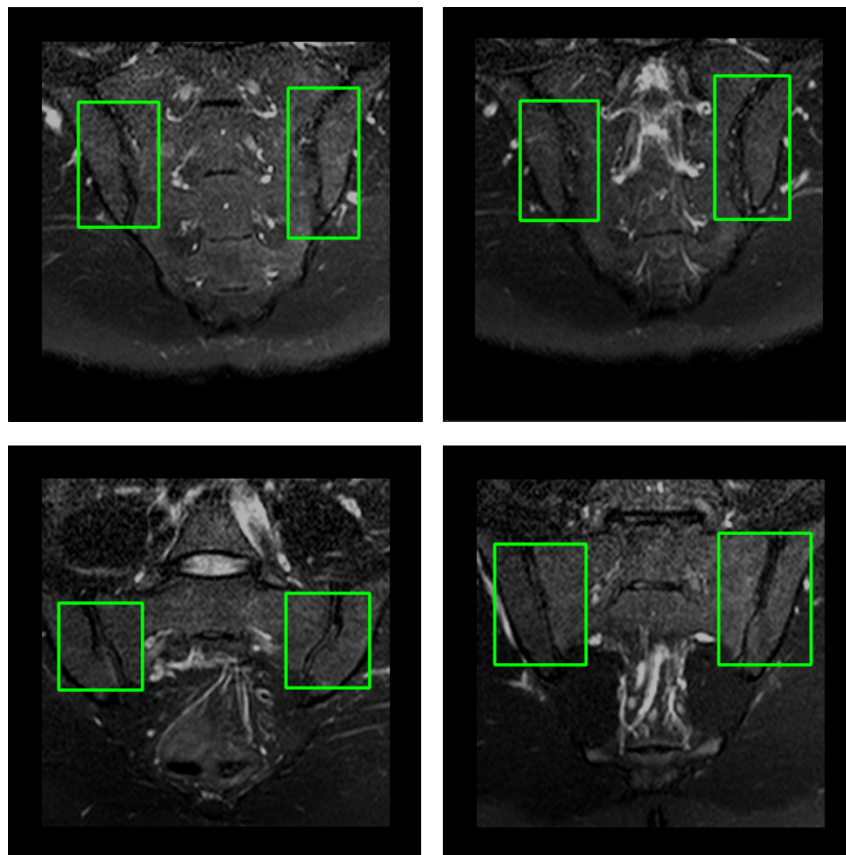


Figura 17 – Exemplos de imagens com sobreposição da região de interesse.

4.4 Comentários Finais

Ao final deste capítulo, foram apresentados os resultados obtidos ao longo do desenvolvimento. Fizemos uma análise visual dos resultados, identificando características da rede, como o arredondamento de bordas, e explorando os resultados, tanto os melhores quanto os piores, tentando de alguma forma fazer inferências do que pode ser melhorado e o que pode ter afetado os piores resultados. Além disso, exploramos de forma quantitativa os resultados, usando ferramentas estatísticas para embasar as inferências obtidas ao final. Como adicional, foi demonstrado a aplicação final do projeto, onde foi mostrado uma maneira efetiva de aplicação das predições.

No próximo capítulo, encerramos o presente trabalho com uma discussão, avaliando os resultados obtidos e a aplicabilidade dos mesmos. Também prospectamos a continuidade do projeto, visualizando os passos seguintes que seriam de interesse para dar segmento ao mesmo.

5 CONCLUSÃO E DISCUSSÃO

Neste capítulo final, refletiremos sobre os resultados alcançados e sua aplicabilidade no contexto clínico.. Além disso, abordaremos o conjunto de dados utilizado neste estudo, suas peculiaridades e como eles influenciaram os resultados. Refletiremos sobre como a escolha e a qualidade dos dados impactam diretamente a performance do modelo e discutiremos possíveis direções para pesquisas futuras, incluindo a expansão do escopo do projeto para abranger a classificação da sacroilíte ativa, um objetivo inicial que permanece como um campo fértil para investigações subsequentes.

Este capítulo visa não apenas concluir essa jornada de pesquisa, mas também abrir caminhos para novas indagações e desenvolvimentos, vislumbrando um futuro onde o aprendizado de máquina pode desempenhar um papel ainda mais significativo no diagnóstico médico assistido por computador.

5.1 Conclusão e Discussão

Em um cenário clínico, ferramentas de aprendizado de máquina devem ser precisos para que possam auxiliar efetivamente no processo de auxílio diagnóstico de radiologistas e outros profissionais da saúde. Por conta disso, a U-Net se mostrou um candidato de interesse para este trabalho.

A partir dos resultados obtidos, evidencia-se que a U-Net, implementada através do framework nnU-Net apresenta grande potencial para a aplicabilidade no contexto clínico. Porém ainda há muito espaço para melhora, incluindo a possibilidade de trabalhar com uma gama mais ampla de transformações nas imagens, gerando ainda mais robustez ao modelo. Também seria de grande interesse explorar mais configurações da rede, buscando os parâmetros que tragam os melhores resultados.

Não podemos deixar de comentar sobre o conjunto de dados utilizado no trabalho, no qual, as segmentações manuais foram feitas com outro propósito, anteriormente explorado por Faleiros (2018), que era de interesse que as mesmas apresentassem vértices. Para este trabalho seria de grande valor poder explorar a rede U-Net com segmentações mais específicas da articulação sacroilíaca (AS), e não de uma área ao seu redor. Assim poderíamos realmente dizer que o modelo segmenta a AS. No resultado apresentado neste trabalho, estamos detectando a região a

qual a articulação pertence. Daí o interesse em gerar os bounding boxes ao redor das regiões de interesse (ROIs) contendo as AS. Isso facilitaria a posterior segmentação das ROIs para serem utilizadas no processo de identificação da presença de inflamação ativa (sacroiliíte).

Por fim, é válido salientar que este projeto teve por objetivo iniciais, além de detectar e segmentar a AS, ainda realizar a classificação da mesma em positivo ou negativo para sacroiliíte ativa, mas que, por questões de logística e tempo, não foi possível implementar. Então fica como um desejo de continuar o projeto, explorando ainda mais a arquitetura proposta e ampliando o cenário ao qual o projeto se encontra, com o objetivo final de fornecer um fluxo completo de automação para a localização da AS e classificação da sacroiliíte, que possa servir de fundamentação para a aplicabilidade do mesmo como ferramenta de diagnóstico auxiliado por computador.

REFERÊNCIAS

- AGARAP, A. F. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.
- ALOM, M. Z. et al. Recurrent residual u-net for medical image segmentation. *Journal of Medical Imaging*, Society of Photo-Optical Instrumentation Engineers, v. 6, n. 1, p. 014006–014006, 2019.
- ANTONELLI, M. et al. The medical segmentation decathlon. *Nature communications*, Nature Publishing Group UK London, v. 13, n. 1, p. 4128, 2022.
- BUKHARI, S. T.; DIN, H. Mohy-ud. A systematic evaluation of learning rate policies in training cnns for brain tumor segmentation. *Physics in Medicine & Biology*, IOP Publishing, v. 66, n. 10, p. 105004, 2021.
- BUSLAEV, A. et al. Alumentations: fast and flexible image augmentations. *Information*, Multidisciplinary Digital Publishing Institute, v. 11, n. 2, p. 125, 2020.
- CASTRO, M. R. d. et al. Espondiloartropatias: critérios de ressonância magnética na detecção da sacroileíte. *Radiologia Brasileira*, SciELO Brasil, v. 50, p. 258–262, 2017.
- CIRESAN, D. et al. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, v. 25, 2012.
- DICE, L. R. Measures of the amount of ecologic association between species. *Ecology*, JSTOR, v. 26, n. 3, p. 297–302, 1945.
- FALEIROS, M. C. *Avaliação computadorizada de sacroileíte em imagens de ressonância magnética*. Tese (Doutorado) — Universidade de São Paulo, 2018.
- FLETCHER, S.; ISLAM, M. Z. et al. Comparing sets of patterns with the jaccard index. *Australasian Journal of Information Systems*, Australian Computer Society, v. 22, 2018.
- GIRSHICK, R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 580–587.
- GONZALEZ, R. C. *Digital image processing*. [S.l.]: Pearson education india, 2009.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016.
- HADJIISKI, L. et al. Aapm task group report 273: Recommendations on best practices for ai and machine learning for computer-aided diagnosis in medical imaging. *Medical Physics*, Wiley Online Library, v. 50, n. 2, p. e1–e24, 2023.
- ISENSEE, F. et al. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: SPRINGER. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*. [S.l.], 2018. p. 287–297.

- ISENSEE, F. et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.
- LI, X. et al. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 12, p. 2663–2674, 2018.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 3431–3440.
- MONTANDON, C. et al. Sacroiliitis: imaging evaluation. *Radiologia Brasileira*, SciELO Brasil, v. 40, p. 53–60, 2007.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. [S.l.], 2015. p. 234–241.
- SAMPAIO-BARROS, P. D. et al. Consenso brasileiro de espondiloartropatias: espondilite anquilosante e artrite psoriásica diagnóstico e tratamento-primeira revisão. *Revista Brasileira de Reumatologia*, SciELO Brasil, v. 47, p. 233–242, 2007.
- SHENKMAN, Y. et al. Automatic detection and diagnosis of sacroiliitis in ct scans as incidental findings. *Medical image analysis*, Elsevier, v. 57, p. 165–175, 2019.
- SHI, X. et al. Evaluating the learning procedure of cnns through a sequence of prognostic tests utilising information theoretical measures. *Entropy*, MDPI, v. 24, n. 1, p. 67, 2021.
- SIDDIQUE, N. et al. U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, IEEE, v. 9, p. 82031–82057, 2021.
- SLOBODIN, G. et al. Sacroiliitis—early diagnosis is key. *Journal of inflammation research*, Taylor & Francis, p. 339–344, 2018.
- VALOVA, I. et al. Optimization of convolutional neural networks for imbalanced set classification. *Procedia Computer Science*, Elsevier, v. 176, p. 660–669, 2020.
- WONG, T.-T.; YEH, P.-Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 32, n. 8, p. 1586–1594, 2019.
- ZHANG, K. et al. Automatic image segmentation and grading diagnosis of sacroiliitis associated with as using a deep convolutional neural network on ct images. *Journal of Digital Imaging*, Springer, p. 1–10, 2023.