

Universidade Federal do Maranhão
Centro de Ciências Exatas e Tecnologia
Programa de Pós-Graduação em Engenharia Elétrica

JOÃO OTÁVIO BANDEIRA DINIZ

Métodos para segmentação de medula
espinhal e esôfago em tomografia
computadorizada de planejamento à
radioterapia

São Luís

2021

Universidade Federal do Maranhão
Centro de Ciências Exatas e Tecnologia
Programa de Pós-Graduação em Engenharia Elétrica

JOÃO OTÁVIO BANDEIRA DINIZ

Métodos para segmentação de medula espinhal e
esôfago em tomografia computadorizada de
planejamento à radioterapia

Tese de Doutorado apresentada ao Programa
de Pós-Graduação em Engenharia de
Eletricidade da UFMA como requisito parcial
para a obtenção do grau de Doutor em
Engenharia Elétrica.

Orientador: Prof. Dr. Aristófanês Corrêa Silva
Coorientador: Prof. Dr. Anselmo Cardoso de
Paiva

São Luís
2021

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Diniz, João Otávio Bandeira.

Métodos para segmentação de medula espinhal e esôfago
em tomografia computadorizada de planejamento à
radioterapia / João Otávio Bandeira Diniz. - 2021.
132 f.

Coorientador(a): Anselmo Cardoso de Paiva.

Orientador(a): Aristófanês Corrêa Silva.

Tese (Doutorado) - Programa de Pós-graduação em
Engenharia Elétrica/ccet, Universidade Federal do
Maranhão, Videoconferência, 2021.

1. Esôfago. 2. Imagens médicas. 3. Medula espinhal.
4. Redes neurais convolucionais. 5. TC de planejamento.
I. de Paiva, Anselmo Cardoso. II. Silva, Aristófanês
Corrêa. III. Título.

Tese de Doutorado de autoria de João Otávio Bandeira Diniz, sob o título “**Métodos para segmentação de medula espinhal e esôfago em tomografia computadorizada de planejamento à radioterapia**”, apresentado ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Maranhão, como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica, aprovado em 10 de março de 2021 pela comissão examinadora constituída pelos doutores:

Prof. Dr. Aristófanês Corrêa Silva

Orientador

Universidade Federal do Maranhão

Prof. Dr. Anselmo Cardoso de Paiva

Co-orientador

Universidade Federal do Maranhão

Prof. Dr. Allan Kardec Duailibe Barros Filho

Examinador Interno

Universidade Federal do Maranhão

Prof. Dr. Stelmo Magalhães Barros Netto

Examinador Interno

Universidade Federal do Maranhão

Prof. Dr. Leandro Augusto Frata Fernandes

Examinador Externo

Universidade Federal Fluminense

Prof. Dr. Romuere Rodrigues Veloso e Silva

Examinador Externo

Universidade Federal do Piauí

Aos meus pais, irmãos, amigos e família

Agradecimentos

Em primeiro lugar, Deus, sem ele nada seria. Grato pela vida e pelas pessoas que colocou em minha jornada até aqui, todas foram fundamentais para essa conquista, não tenho palavras para expressar meu agradecimento ao Pai por tudo de bom que conquistei.

Em segundo lugar, minha família. Minha mãe e meu pai, com todo seu conhecimento e experiência, me proporcionaram a melhor educação e os maiores valores que eu possa ter. Meus irmãos, Nathalia e Pedro, sempre ao meu lado, na alegria e na tristeza. Cada um com seu jeito de ser, contribuíram para minhas conquistas.

Em terceiro lugar, meu mestre e orientador Ari. Que além de orientar, tornou-se um amigo e colega de profissão. Responsável por muitas das minhas conquistas acadêmicas, sempre se colocando à disposição e sempre me auxiliando, ajudou na formação de grande parte do meu caráter profissional. Todos seus ensinamentos foram valiosos, todas as vezes que me dizia “tu sabe o que faz: a vida é tua” de alguma forma era um ótimo conselho... Até em nossas discussões, sempre com respeito, traziam aprendizado e acredito que era um ensinamento mútuo. Enfim, obrigado, professor!

Não poderia esquecer meu coorientador. Anselmo teve o prazer (ou desprazer) de me conhecer na época de início de graduação. Jovem e inexperiente, mas nunca irresponsável, acabei causando uma má impressão a ele, que se surpreendeu quando me viu pelo NCA. Porém, com minha dupla de orientadores, acredito que consegui trilhar um caminho de sucesso, me apaixonando pela pesquisa científica, com a ambição de publicações, contribuindo muito com o NCA e com o desejo de ingressar na academia, o que aconteceu em maio de 2019 quando consegui entrar no IFMA. Anselmo, assim como Ari, são dois caras que posso chamar de amigos!

Aos meus amigos de universidade, em especial Caio, Giovanni e Jefferson, que me acompanham desde a graduação nessa mesma jornada de mestrado e doutorado. Com eles vivi muitos aperreios em disciplinas, na pesquisa, na vida e nos jogos (principalmente quando Anselmo nos flagrava jogando (rs)).

Aos meus amigos e colegas de NCA, durante minha passagem por lá de quase 10 anos (espero continuar e conseguir contribuir a altura) tive a oportunidade de conhecer várias mentes brilhantes que, sem dúvidas, também contribuíram para minha formação pessoal e profissional.

Aos meus amigos de infância e de bairro, os quais por muitas vezes não me entendiam e não acreditavam nos finais de semanas em frente ao computador sem poder sair para diversão, porém toda diversão que tive e terei é fundamental para construção do meu caráter e ter tido a sabedoria de chegar até o fim do doutorado.

Ao IFMA, instituição que hoje visto a camisa, obrigado pelos momentos que precisei me ausentar para atender minhas demandas de doutorado. Fazer parte de uma instituição de ensino sempre foi minha meta. Hoje consigo desempenhar minhas funções com maestria devido a todos ensinamentos que tive durante minha carreira de estudante e espero contribuir ainda mais com o IFMA.

A UFMA, instituição a qual pertenci e pertenco a mais de 10 anos, aqui construí amigos, me diverti, me preocupei, me compliquei, mas tudo gerou aprendizado e me trouxe até aqui.

A todo programa de pós-graduação, em nome de Alcides, agradeço por tudo, pela competência em resolver os problemas e pelo suporte dado durante esses anos.

Por fim, a todos que estiveram juntos nesta caminhada, contribuindo direta e indiretamente para minha formação acadêmica e pessoal, meu muito obrigado!

*“Saber interpor-se constantemente entre si próprio e as coisas é o mais alto grau da
sabedoria e prudência.”
(Fernando Pessoa)*

Resumo

A radioterapia (RT) é uma das opções mais utilizadas para o tratamento do câncer. Para o planejamento de tratamentos com RT, exames de tomografia computadorizada (TC) são comumente utilizados para delimitar os órgãos de risco (do inglês, organs at risk - OARs) e analisar o impacto da radiação nesses órgãos. Delimitar esses OARs requer muito tempo dos médicos especialistas e envolve uma grande equipe de profissionais. Além disso, essa tarefa executada fatia a fatia torna-se exaustiva e, conseqüentemente, sujeita a erros, especialmente em órgãos como a medula espinhal e o esôfago, que se estendem por várias fatias da TC e exigem uma segmentação precisa. Assim, são propostos, neste trabalho, métodos computacionais capazes de segmentar a medula espinhal e o esôfago em TC de planejamento. Para a segmentação da medula espinhal, são propostos dois métodos, sendo o primeiro compreendido por técnicas como *template matching* adaptativo para segmentação inicial, técnica de superpixel conhecida como *intrinsic manifold simple linear iterative clustering* (IMSLIC) para segmentação de candidatos à medula espinhal e redes neurais convolucionais (convolutional neural network - CNN) para classificação desses candidatos. O segundo método para segmentação da medula espinhal é composto por técnicas como *template matching* adaptativo para segmentação inicial de uma região de interesse e redes neurais convolucionais completamente conectadas (Fully Convolutional Network - FCN) para segmentação final da medula espinhal. Na segmentação do esôfago, é proposto um método composto por técnicas de registro do volume dos pacientes, atlas para definir o volume de interesse onde encontra-se o esôfago, pré-processamento para realçar as paredes do esôfago, FCN para segmentação do esôfago, e um pós-processamento para refinamento da segmentação. Os métodos foram aplicados em 36 imagens de TC de planejamento fornecidas pelo The Cancer Imaging Archive. O primeiro método para segmentação de medula espinhal obteve acurácia de 92,55%, especificidade de 92,87%, sensibilidade de 89,23% e 78,20% de Dice. O segundo método para segmentação de medula espinhal obteve acurácia de 99,24%, especificidade de 99,38%, sensibilidade de 91,52% e 81,69% de Dice. O método para segmentação de esôfago obteve acurácia de 99,69%, especificidade de 99,76%, sensibilidade de 90,61% e 82,15% de Dice. Assim, demonstra-se a viabilidade da análise de TC de planejamento utilizando técnicas de processamento de imagem e reconhecimento de padrões.

Palavras-chaves: Redes neurais convolucionais, TC de planejamento, Imagens médicas, Medula espinhal, Esôfago.

Abstract

Radiotherapy (RT) is one of the most used options for the treatment of cancer. For planning treatments with RT, computed tomography (CT) exams are commonly used to define organs at risk (OARs) and analyze the impact of radiation on these. Delimiting these OARs takes a lot of time from medical specialists, and it involves a large team of professionals. Besides, this task performed slice by slice becomes exhausting and, consequently, subject to errors, especially in organs such as the spinal cord and esophagus, which span several CT slices and require precise segmentation. Thus, in this work, computational methods capable of segmenting the spinal cord and esophagus in planning CT are proposed. For spinal cord segmentation, two methods are proposed, the first being comprised of techniques such as adaptive *template matching* for initial segmentation, a superpixel technique known as *intrinsic manifold simple linear iterative clustering* (IMSLIC) for segmentation of spinal cord candidates and convolutional neural networks (CNN) to classify these candidates. The second method for spinal cord segmentation consists of techniques, such as adaptive *template matching* for initial segmentation of a region of interest and fully convolutional neural networks (FCN) for final spinal cord segmentation. In the esophagus segmentation, it is proposed a method composed of techniques to registration the volume of patients, atlas to define the volume of interest where the esophagus is located, pre-processing to highlight the walls of the esophagus, CNN for esophagus segmentation, and post-processing to refine segmentation. We applied the method to 36 planning CT images provided by The Cancer Imaging Archive. The first method for spinal cord segmentation obtained an accuracy of 92.55%, a specificity of 92.87%, a sensitivity of 89.23%, and 78.20% of Dice. The second method for spinal cord segmentation obtained an accuracy of 99.24%, a specificity of 99.38%, a sensitivity of 91.52%, and 81.69% of Dice. The esophagus segmentation method obtained an accuracy of 99.69%, a specificity of 99.76%, a sensitivity of 90.61%, and 82.15% of Dice. Thus, we demonstrated the feasibility of analyzing CT planning using image processing and pattern recognition techniques.

Keywords: Convolutional Neural Network, Planning CT, Medical images, Spinal cord, Esophagus.

Lista de figuras

Figura 1 – Ilustração da localização e composição da Medula Espinhal.	3
Figura 2 – Esôfago: (a) Ilustração da localização e composição do Esôfago e (b) Localização do Esôfago em TC na visão axial.	4
Figura 3 – Tomografia de planejamento: (a) tomógrafo; (b) imobilização do paciente; (c) tomografia computadorizada; e (d) órgãos de risco delineados pelo dosimetrista e médico radioterapeuta.	20
Figura 4 – Exemplo de órgãos de risco delineados por um especialista.	21
Figura 5 – Medula espinhal: (a) corpo da vértebra; (b) canal vertebral (onde encontra-se a medula espinhal); e (c) osso da vértebra	23
Figura 6 – Exemplo de uma CNN.	30
Figura 7 – Ilustração de uma operação de convolução.	30
Figura 8 – Ilustração da camada de subamostragem.	31
Figura 9 – Representação da Arquitetura U-Net.	33
Figura 10 – Esôfago: (a) fatia de uma tomografia computadorizada (em azul uma boundingbox com a localização do esôfago); (b) corte na região do esôfago; e (c) marcação do especialista	35
Figura 11 – Exemplo de um atlas com a marcação do esôfago.	41
Figura 12 – Matriz de confusão.	41
Figura 13 – Aquisição de imagem: pulmões (roxo), coração (verde), esôfago (amarelo superior) e medula espinhal (amarelo inferior).	43
Figura 14 – Aquisição de imagem: (a) fatia do instituto 1; (b) fatia do instituto 2; (c) fatia do instituto 3; (d) volume do instituto 1; (e) volume do instituto 2; e (f) volume do instituto 3.	44
Figura 15 – Fluxograma do primeiro método do medula espinhal.	45
Figura 16 – Geração do <i>template</i> padrão.	47
Figura 17 – Geração o <i>template</i> inicial.	47
Figura 18 – Adaptando o <i>template</i> para cada fatia: (a) <i>template</i> inicial; (b) fatias de um volume (o <i>bounding box</i> vermelho mostra onde o <i>matching</i> ocorreu); (c) resultado do <i>matching</i> que gera o novo <i>template</i> ; e (d) fatia resultado.	48
Figura 19 – Segmentação de candidatos a medula espinhal: (a) fatia de saída da segmentação inicial (b) aplicação do algoritmo IMSLIC na fatia.	49

Figura 20 – Geração de imagens de entrada: as estruturas ao redor da medula espinhal.	51
Figura 21 – Geração de imagens de entrada: (a) segmentação de candidatos de uma fatia por IMSLIC; b) Amostra de <i>bounding boxes</i> da medula espinhal; c) Amostra de <i>bounding boxes</i> não medula espinhal.	52
Figura 22 – Treino e teste da CNN: (a) representa uma fatia sem marcação; (b) representa a mesma fatia com marcação da medula espinhal e os <i>superpixels</i> gerados pela aplicação do IMSLIC; (c) <i>superpixels</i> tendo pelo menos 20% de seus pixels presentes em qualquer uma das marcações de uma medula espinhal de (b) são pintados na cor vermelha.	53
Figura 23 – Ilustração da arquitetura adotada.	55
Figura 24 – Resultado do passo de segmentação inicial: (a) fatia onde não há medula espinhal (cabeça); (b) resultado da segmentação inicial em (a); (c) fatia onde há medula espinhal; (d) resultado da segmentação inicial em (c). .	57
Figura 25 – Fluxograma do segundo método do medula espinhal.	62
Figura 26 – Adaptando o <i>template</i> : as linhas em vermelho mostram onde ocorreu o <i>matching</i> , as linhas em verde mostram onde o <i>template</i> será executado. .	64
Figura 27 – Imagens de entrada: (a) VOI do paciente com medula espinhal e estruturas ao seu redor; e (b) VOI da marcação especializada da medula espinhal.	64
Figura 28 – Abordagens U-Net: Representação da arquitetura da U-Net em que os blocos residuais são na cor verde e as ligações densas em azul. Figura de Kolařík et al. (2019).	67
Figura 29 – Segmentação inicial: resultado da aplicação do <i>template matching</i> adaptativo.	69
Figura 30 – Estudo de caso - Instituto 1: (a) volume predito pelo método IMSLIC + CNN, (b) predito pelo método ResU-Net e (c) marcação do especialista. .	74
Figura 31 – Estudo de caso - Instituto 2: (a) volume predito pelo método IMSLIC + CNN, (b) predito pelo método ResU-Net e (c) marcação do especialista. .	75
Figura 32 – Estudo de caso - Instituto 3: (a) volume predito pelo método IMSLIC + CNN, (b) predito pelo método ResU-Net e (c) marcação do especialista. .	76
Figura 33 – Fluxograma do método do esôfago.	82
Figura 34 – Registro de volumes: (a) volume fixo com 134 fatias, (b) volume móvel com 279 fatias e (c) volume móvel registrado com 134 fatias.	84

Figura 35 – Geração de atlas: (a) exemplo de volumes de marcações e (b) resultado da geração do atlas.	85
Figura 36 – Corte na região do esôfago: (a) atlas, em amarelo região correspondente ao volume do maior objeto, (b) volumes originais do paciente e da marcação (c) segmentação da VOI do paciente e da marcação.	86
Figura 37 – Pré-processamento: (a) fatia da VOI sem pré-processamento, (b) fatia da VOI com pré-processamento. A seta indica o nítido realce das bordas do esôfago.	87
Figura 38 – Segmentação Final: (a) exemplos de marcações especializadas em verde e resultado de ResU-Net em azul, (b) marcações especializadas em verde e resultado de obtenção da maior região em azul, e (c) marcações especializadas em verde e resultado de erosão em azul.	89
Figura 39 – Estudo de caso - Instituto 1. Em verde a marcação do especialista e em vermelho a segmentação do esôfago predita pelo método.	95
Figura 40 – Estudo de caso - Instituto 1. (a) volume predito pelo método e (b) volume da marcação do especialista.	96
Figura 41 – Estudo de caso - Instituto 2. Em verde a marcação do especialista e em vermelho a segmentação do esôfago predita pelo método.	97
Figura 42 – Estudo de caso - Instituto 2. (a) volume predito pelo método e (b) volume da marcação do especialista.	98
Figura 43 – Estudo de caso - Instituto 3. Em verde a marcação do especialista e em vermelho a segmentação do esôfago predita pelo método.	99
Figura 44 – Estudo de caso - Instituto 3. (a) volume predito pelo método e (b) volume da marcação do especialista.	100

Lista de tabelas

Tabela 1 – Relação entre conjuntos de dados de treinamento e teste.	58
Tabela 2 – Métricas de validação da fase de teste.	59
Tabela 3 – Resultado da etapa de segmentação final usando entropia cruzada como função de perda.	70
Tabela 4 – Resultado da etapa de segmentação final usando a função de <i>Diceloss</i>	71
Tabela 5 – Comparação entre os resultados dos métodos de medula espinhal.	77
Tabela 6 – Comparação com os trabalhos relacionados.	81
Tabela 7 – Resultados das etapas da segmentação do esôfago.	91
Tabela 8 – Comparação com os trabalhos relacionados.	102
Tabela 9 – Produções científicas como autor relacionadas ao método proposto.	107

Sumário

1	Introdução	1
1.1	Objetivos	7
1.1.1	Objetivo geral	7
1.1.2	Objetivos específicos	7
1.2	Contribuições científicas	8
1.3	Organização do trabalho	9
2	Trabalhos relacionados	10
2.1	Trabalhos relacionados à medula espinhal	10
2.2	Trabalhos relacionados a esôfago	13
3	Fundamentação teórica	18
3.1	Radioterapia	18
3.1.1	Tomografia computadoriza de planejamento	19
3.1.2	Delineamento manual dos Órgão de Risco pelo especialista	20
3.2	Segmentação de medula espinhal	22
3.2.1	Medula espinhal	22
3.2.2	<i>Template matching</i>	23
3.2.3	<i>Intrinsic manifold simple linear iterative clustering -</i>	
	IMSLIC	24
3.2.4	Redes neurais artificiais - RNA	25
3.2.4.1	Neurônio artificial	25
3.2.4.2	<i>Multilayer Perceptron</i> - MLP	26
3.2.4.3	<i>Backpropagation</i>	27
3.2.5	Aprendizado profundo	28
3.2.5.1	Redes neurais convolucionais	29
3.2.5.1.1	Camada de convolução	29
3.2.5.1.2	Camada de subamostragem	31
3.2.5.1.3	Camada completamente conectada	32
3.2.5.2	U-Net	32
3.3	Segmentação de esôfago	33

3.3.1	Esôfago	34
3.3.2	Pré-processamento de imagens	35
3.3.2.1	Filtragem em imagens	35
3.3.2.1.1	Equalização do histograma	35
3.3.2.1.2	Filtro Bilateral	36
3.3.2.2	Morfologia matemática	36
3.3.2.3	Registro de imagens	38
3.3.2.3.1	Registro rígido	39
3.3.3	Atlas	40
3.4	Métricas de desempenho	40
4	Segmentação de medula espinhal	43
4.1	Materiais	43
4.2	Segmentação de medula espinhal utilizando <i>template matching</i> adaptativo IMSLIC e CNN	45
4.2.1	Segmentação inicial	45
4.2.1.1	Gerando o <i>template</i> padrão	46
4.2.1.2	Gerando o <i>template</i> inicial	46
4.2.1.3	Adaptando o <i>template</i> para cada fatia	48
4.2.2	Segmentação dos candidatos à medula espinhal	48
4.2.3	Classificação dos candidatos a medula espinhal	50
4.2.3.1	Gerando as imagens de entrada	50
4.2.3.2	Treino e teste da CNN	52
4.2.4	Resultados	55
4.2.4.1	Ambiente de treino/teste	55
4.2.4.2	Aquisição de imagens	56
4.2.4.3	Segmentação Inicial	56
4.2.4.4	Segmentação de candidatos à medula espinhal	57
4.2.4.5	Classificação de candidatos à medula espinhal	58
4.2.4.5.1	Fase de treinamento	58
4.2.4.5.2	Fase de teste	58
4.2.4.6	Considerações	59

4.3	Segmentação de medula espinhal utilizando <i>template matching</i> adaptativo e abordagens U-Net	62
4.3.1	Segmentação inicial	62
4.3.1.1	<i>Template matching</i> adaptativo	63
4.3.2	Segmentação final	63
4.3.2.1	Gerando imagens de entrada	64
4.3.2.2	Abordagens U-Net	65
4.3.2.3	U-Net	65
4.3.2.4	ResU-Net e DenseU-Net	66
4.3.2.5	Treino e teste das abordagens U-Net	66
4.3.3	Resultados	68
4.3.3.1	Ambiente de treino/teste	68
4.3.3.2	Aquisição de imagens	68
4.3.3.3	Segmentação inicial	69
4.3.3.4	Segmentação final	69
4.3.3.5	Considerações	72
4.4	Estudo de caso dos métodos de medula espinhal	74
4.4.1	Estudo de caso 1 - Instituto 1	74
4.4.2	Estudo de caso 2 - Instituto 2	75
4.4.3	Estudo de caso 3 - Instituto 3	76
4.5	Comparação entre os resultados dos métodos de medula espinhal	77
4.6	Comparação com os trabalhos relacionados a segmentação de medula espinhal	79
5	Segmentação de esôfago utilizando U-Net com blocos residuais	82
5.1	Segmentação de VOI	83
5.1.1	Registro de volumes	83
5.1.2	Geração de atlas	84
5.1.3	Corte na região do esôfago	84
5.2	Pré-processamento	85
5.3	Segmentação final	87

5.3.1	Gerando imagens de entrada	87
5.3.2	Treino e teste da U-Net com blocos residuais	88
5.3.3	Refinamento da segmentação	88
5.4	Resultado	89
5.4.1	Ambiente de treino/teste	90
5.4.2	Aquisição de imagens	90
5.4.3	Segmentação de VOI e Pré-processamento	90
5.4.4	Segmentação final	91
5.4.5	Considerações	92
5.5	Estudo de caso da segmentação de esôfago	95
5.5.1	Estudo do caso - Instituto 1	95
5.5.2	Estudo do caso - Instituto 2	96
5.5.3	Estudo do caso - Instituto 3	98
5.6	Comparação com os trabalhos relacionados a segmentação de esôfago	99
6	Conclusão	104
6.1	Segmentação de medula espinhal	104
6.2	Segmentação do esôfago	105
6.3	Comprovação das hipóteses	106
6.4	Produções científicas	107
6.5	Reconhecimento científico	107
	Referências	108

1 Introdução

A radioterapia (RT) é uma opção eficaz e comumente usada no tratamento do câncer, especialmente quando a cirurgia e a quimioterapia apresentam um risco muito grande ao paciente (EVANS; STAFFURTH, 2018). Ela desempenha um papel fundamental no tratamento de mais de 50% dos casos de câncer e 40% dos pacientes curados de seu câncer têm radioterapia como parte de seu tratamento (EVANS; STAFFURTH, 2018).

O objetivo da radiação é erradicar o tumor, preservando a estrutura e a função dos tecidos normais circundantes. A utilidade da radiação no tratamento de pacientes com câncer varia devido às diferenças na resposta de radiação de tumores e tecidos normais. Em geral, as células tumorais são menos capazes do que as células normais de reparar rupturas nos filamentos de ácido desoxirribonucleico (DNA) que são produzidas pela radiação. Essas quebras de DNA causam uma cascata de eventos moleculares que podem levar à morte celular. As diferenças na resposta molecular das células tumorais e dos tecidos normais à radiação resultam em uma razão terapêutica favorável, onde os efeitos benéficos por tratamento de radiação cuidadosamente planejado e entregue, superam o potencial de toxicidade grave (MILOSEVIC et al., 2007).

Devido às suas vantagens, a radioterapia pode ser utilizada em vários locais do tumor e pode ser utilizada antes ou após a cirurgia definitiva, local ou loco regional e no cenário paliativo para o controle dos sintomas. Neste último caso, alivia a dor de metástases ósseas em pelo menos 60% dos pacientes e pode ser usado para aliviar os sintomas da compressão da medula espinhal, metástases cerebrais e sangramento não controlado (EVANS; STAFFURTH, 2018). Pode ser usado como a modalidade definitiva de tratamento do câncer em vários locais do tumor. Isso inclui cabeça e pescoço (CAUDELL et al., 2017), anal (CASTANEDA; ROMAK, 2017), próstata (GÓMEZ-MILLÁN et al., 2015), bexiga (MILOSEVIC et al., 2007), esôfago (ADEBAHR et al., 2016), e câncer do colo do útero (LAAN et al., 2017).

O processo de tratamento radioterápico é multidisciplinar, envolvendo oncologistas clínicos, físicos, médicos e radiologistas terapêuticos. Cada etapa do tratamento tira proveito de uma nova tecnologia que permite uma definição mais precisa do tumor e da radiação, com o objetivo de melhorar os resultados do tratamento e reduzir a toxicidade tecidual normal (EVANS; STAFFURTH, 2018).

No entanto, Ishiyama et al. (2018) destaca que, apesar de suas vantagens, o uso da radioterapia requer o cumprimento de vários protocolos de tratamento, o planejamento por assistência de software e uma análise de risco refinada e específica para cada situação. Por ser um processo complexo e de alto risco, podem ocorrer falhas humanas, levando a danos irreversíveis ao paciente (ISHIYAMA et al., 2018).

A toxicidade da radioterapia pode ser dividida em efeitos agudos/precoces e tardios do tratamento. Por isso, pesquisadores realizaram estudos com acompanhamento temporal de pacientes submetidos à radioterapia para avaliar seus efeitos e riscos. Por exemplo, um estudo realizado por Laan et al. (2017) examinaram o tratamento do câncer do colo do útero em 515 mulheres em 20 anos, com o objetivo de investigar a frequência e os fatores de risco para a toxicidade do intestino tardio após a radioterapia primária. Eles concluíram que a toxicidade intestinal tardia grave é uma complicação frequente da radioterapia definitiva para o câncer do colo do útero.

Na tentativa de contornar os efeitos da toxicidade da radioterapia, várias técnicas e estudos também foram levantados. Técnicas modernas de tratamento de radioterapia oferecem o potencial para melhorar as taxas de cura e reduzir os efeitos adversos. Todos os pacientes nos quais a condição é diagnosticada recentemente são avaliados em um ambiente multidisciplinar, onde os méritos relativos de cirurgia, radioterapia e quimioterapia são considerados individualmente, com o objetivo de otimizar os resultados gerais (MILOSEVIC et al., 2007).

O uso de radioterapia de intensidade modulada volumétrica permite a entrega individualizada de dose altamente direcionada poupando ao mesmo tempo o tecido normal circundante. O uso de tais técnicas avançadas em RT, permite a redução da dose em órgãos em risco (do inglês, *Organ at Risk* - OARs). Esta técnica avançada não só permite uma distribuição mais direcionada de radiação para o alvo, mas também uma diminuição acentuada do gradiente de dose para minimizar as doses nos OARs (ZELEFSKY et al., 2000; TSANG; HOSKIN, 2017). Adebahr et al. (2016) também destacaram o controle local mais preciso das doses de radiação, através de técnicas de radioterapia mais avançadas, como o planejamento 3D, como fator de redução da toxicidade em pacientes.

Na radioterapia, um dos principais OARs para ser protegido é a medula espinhal. Este órgão é uma estrutura segmentar com pares de raízes nervosas dorsais e ventrais, que se estende da junção crânio-cervical até a extremidade da coluna lombar (Figura 1). Os revestimentos da medula espinhal incluem várias partes do sistema nervoso central,

incluindo o líquido cefalorraquidiano, solução salina muito pura, pobre em proteínas e células, que serve como uma almofada para o córtex cerebral e medula espinhal (KHAN, 2010; DOWLATI, 2017; BHATTACHARYYA, 2018).

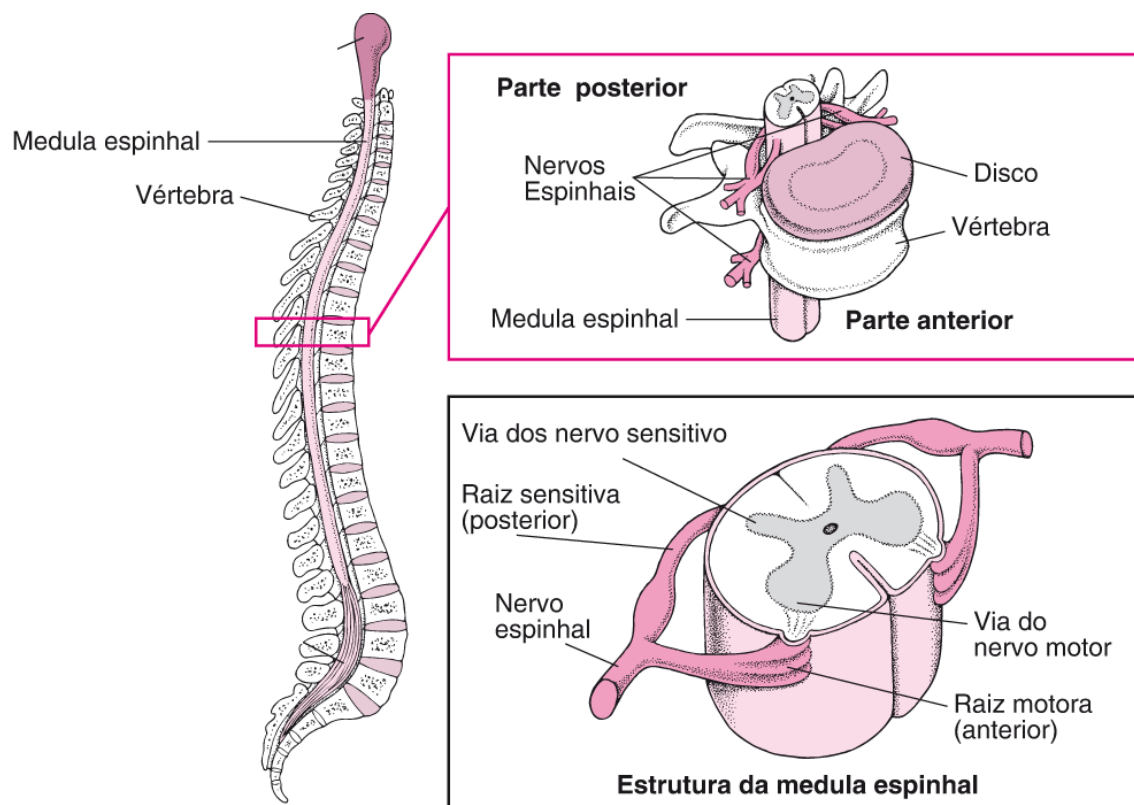


Figura 1 – Ilustração da localização e composição da Medula Espinhal.

Fonte: Adaptado de <https://www.msmanuals.com/pt/casa/distúrbios-cerebrais/>

Outro órgão extremamente importante é o esôfago. O esôfago é um tubo muscular oco que liga a faringe ao estômago (Figura 2(a)). De acordo com Fechter et al. (2017) o esôfago é um órgão que traz muita dificuldade para ser demarcado, em algumas fatias do volume, até mesmo os especialistas têm dificuldade em definir seus limites de maneira confiável e por conta desta dificuldade, o esôfago pode ser facilmente confundido com outras estruturas (Figura 2(b)). Isso leva a uma interpretação cansativa nas imagens de Tomografia Computadorizada (TC), o que torna o processo demorado e altamente propenso à variabilidade inter-observadores. No entanto, com sua mucosa sensível à radioterapia, o esôfago é um dos OAR mais críticos na RT, e um contorno confiável é indispensável. Assim, existe uma necessidade crítica na RT de abordagens automáticas, reproduzíveis e confiáveis para segmentar o esôfago.

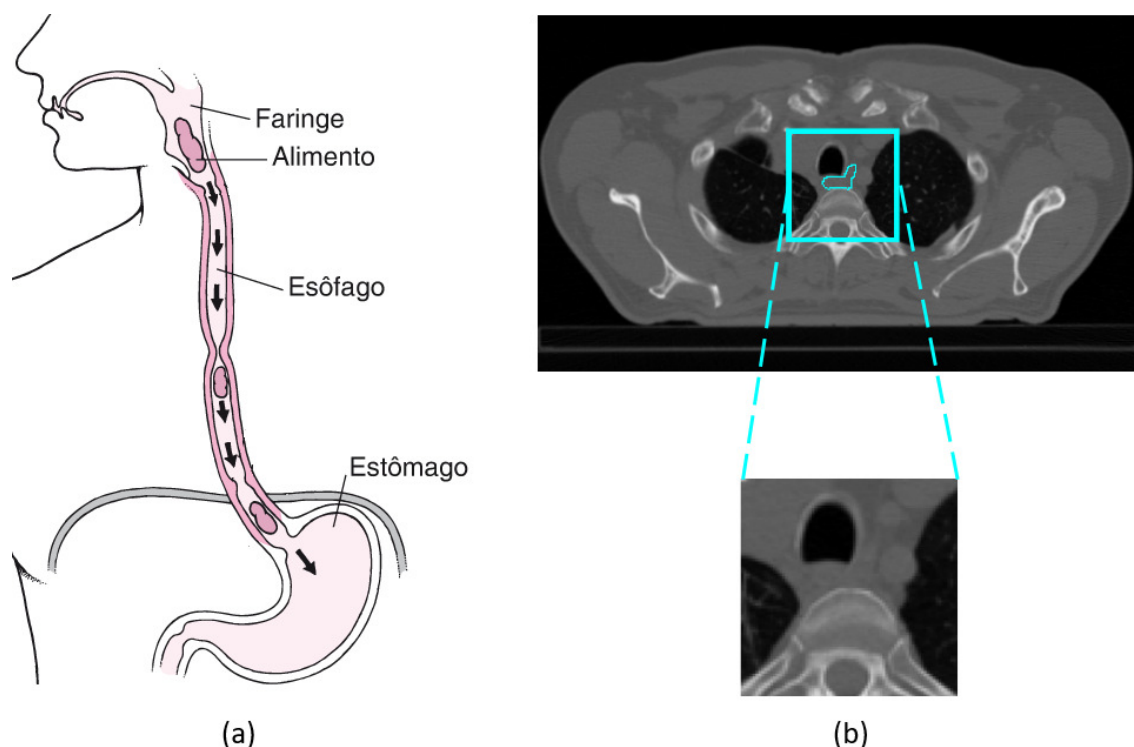


Figura 2 – Esôfago: (a) Ilustração da localização e composição do Esôfago e (b) Localização do Esôfago em TC na visão axial.

Fonte: Adaptado de <https://www.msdmanuals.com/pt/casa/distúrbios-digestivos/>

Segundo estudos de Evans e Staffurth (2018), o processo de radioterapia consiste em três etapas distintas: (a) imobilização, definição da imagem e volume alvo, (b) planejamento do tratamento e (c) entrega do tratamento e verificação da configuração. Na primeira etapa, o paciente é submetido a uma tomografia computadorizada de planejamento, na qual o paciente é colocado em um posicionamento ótimo de conforto e reprodutibilidade, adaptado a cada local do tumor. Este passo ajuda a delineação do volume alvo, identificando com precisão o tumor (BUTTON; STAFFURTH, 2010; GWYNNE et al., 2012). A segunda parte do processo é definir o arranjo de feixe de radioterapia ideal que fornecerá a dose necessária ao volume alvo. Os físicos otimizam o plano para maximizar a liberação da dose no alvo e minimizar a dose nos OARs, usando restrições definidas e através do uso da varredura da imagem (verificando fatia-a-fatia do exame). Depois que um plano ideal é criado, ele é revisado, aprovado pelo médico responsável pelo tratamento e os dados são transferidos para a máquina de tratamento.

Assim, a fim de ajudar os especialistas a contornar os efeitos da toxicidade da radioterapia e reduzir os efeitos adversos no corpo humano, muitas pesquisas

investiram substancialmente no desenvolvimento e uso de sistemas baseados em técnicas de processamento de imagens para delinear o tumor e os OARs com maior precisão (BUTTON; STAFFURTH, 2010; GWYNNE et al., 2012). Essas técnicas foram combinadas para desenvolver a detecção assistida por computador (Computer-Aided Detection - CAD). Especificamente, algumas dessas estão focados em fornecer imagens tridimensionais de órgãos internos para fornecer suporte ao planejamento e pré-tratamento da radioterapia (CHEN et al., 2013; LEENER; COHEN-ADAD; KADOORY, 2015).

Atualmente, o sistema mais comumente usado para delinear tumores e OARs é o *Varian EclipseTM Treatment Planning System*¹, onde o delineamento é feito por um módulo de segmentação que usa um modelo de solução para definir os OARs, mas não precisamente, o especialista precisa olhar fatia por fatia e ajustar o delineamento manualmente. Essa tarefa leva muito tempo dos especialistas médicos, além do fato de que envolve uma grande equipe de profissionais. Além disso, essa tarefa feita, fatia a fatia torna-se exaustiva e, conseqüentemente, sujeita a erros, especialmente em órgãos como a medula espinhal e o esôfago, que se estendem por várias fatias da TC e exigem uma segmentação precisa. Assim, este trabalho complementa esses sistemas, visando realizar a segmentação totalmente automática da medula espinhal e do esôfago, no planejamento de TC para auxiliar os especialistas na calibração dos aparelhos radioterápicos, para que os raios atinjam menos OARs possíveis.

Para complementar esses sistemas na tarefa de segmentação automática da medula espinhal, são propostos dois métodos computacionais, ambos baseados em *template matching* adaptativo e em redes neurais convolucionais (do inglês, *Convolutional Neural Network* - CNN). Na segmentação automática do esôfago, o método consiste em uma série de técnicas, tais como, registro, atlas, realce e CNN, por ser um dos OARs mais complexos de segmentar em TC.

Observa-se a necessidade de criação de métodos automáticos para a segmentação desde dois OARs, principalmente por dois pontos. Primeiro, fazer uma boa definição do limite desses OARs em relação a outros, uma vez que, a dose de radiação para cada um é diferente, desta forma, com uma boa segmentação o especialista consegue calibrar

¹ Este software possui um módulo chamado *Smart Segmentation* que fornece um contorno baseado no conhecimento adquirido ao clínico com as ferramentas para melhorar a eficiência e consistência do contorno. Mais informações sobre este software podem ser encontradas no site: <https://www.varian.com/oncology/products/software/treatment-planning/eclipse-treatment-planning-system>

a radiação específica para cada OAR, evitando erros de entrega do tratamento de RT. Segundo, ambos OARs possuem dificuldades de delineamento em CT, a medula espinhal se estende por praticamente toda a CT do paciente e a checagem manual de cada fatia é algo extremamente cansativo e, conseqüentemente, sujeito a erros. Por sua vez, o esôfago é um OAR que traz muita dificuldade para ser demarcado em algumas fatias do volume, até mesmo os especialistas têm dificuldade em definir seus limites de maneira confiável e por conta desta dificuldade, o esôfago pode ser facilmente confundido com outras estruturas (FECHTER et al., 2017).

Desta forma, métodos automáticos robustos podem ser cruciais para contornar estas dificuldades. Pretende-se, portanto, alguns questionamentos devem ser alcançados na elaboração de métodos robustos, tais como:

- É possível desenvolver métodos capazes de detectar a medula espinhal e o esôfago em TC?
- Quais características esses métodos devem ter para se adaptarem a TCs de diferentes centros médicos, protocolos de aquisição, posicionamentos do paciente?
- Qual técnica de processamento de imagem seria capaz de se adaptar às mudanças de formato da medula ao longo das diversas fatias da TC?
- Qual técnica de processamento de imagem seria capaz de diferenciar o contorno do esôfago, em contraste com as outras estruturas que o circundam?
- Qual técnica de reconhecimento de padrões é capaz de classificar regiões entre medula espinhal e o restante das estruturas da TC?
- Qual técnica de reconhecimento de padrões é capaz de classificar regiões entre esôfago e o restante das estruturas da TC?
- Finalmente, como seria possível contornar todos esses problemas, no desenvolvimento de métodos robustos, capazes de detectar a medula espinhal e o esôfago em TC de planejamento a radioterapia?

A partir das questões de pesquisa, este trabalho considera as seguintes hipóteses:

- A redução da região de interesse das TC permite uma diminuição na carga computacional e resulta em melhor desempenho na segmentação final da medula espinhal e do esôfago;

- Técnicas de processamento de imagens tradicionais são suficientes para a segmentação inicial da medula espinhal;
- A padronização das imagens e a utilização desta informação padronizada no formato de um atlas probabilístico é útil para a segmentação inicial do esôfago;
- Técnicas de filtragem de imagem podem ser utilizadas para o realce da TC de modo a facilitar a delineação do esôfago;
- Técnicas de aprendizado de máquina supervisionado são eficientes para a segmentação da medula espinhal e do esôfago;
- Técnicas de processamento de imagens conseguem melhorar o resultado encontrado pelos modelos de aprendizado de máquina.

Assim, no desenvolvimento dos métodos, as hipóteses vão sendo confirmadas pelas técnicas empregadas e pelos resultados alcançados pelos métodos propostos.

1.1 Objetivos

Esta seção destaca os objetivos geral e específicos alcançados durante a elaboração dos métodos. As próximas seções destacam tais objetivos.

1.1.1 Objetivo geral

O objetivo geral do trabalho proposto é desenvolver métodos computacionais, baseados em técnicas de processamento de imagens e reconhecimento de padrões, para a tarefa de segmentação automática da medula espinhal e do esôfago em tomografia de planejamento para radioterapia.

1.1.2 Objetivos específicos

Com o intuito de alcançar o objetivo geral exposto, faz-se necessário passar por alguns objetivos específicos, os quais destacam-se:

- Aprimorar o conhecimento sobre as tomografias computadorizadas no tratamento radioterápico;

- Desenvolver avanços na técnica de *template matching* a ser utilizada em TC, tornando-o adaptável as fatias da TC;
- Entender os tipos de técnicas de superpixel;
- Utilizar a técnica de superpixel IMSLIC no contexto de imagem médica;
- Identificar os filtros mais efetivos para realçar o esôfago;
- Compreender as redes neurais convolucionais e a viabilidade de aplicá-las na tarefa segmentação de medula espinhal e do esôfago;
- Validar os métodos através de métricas comumente usadas em trabalhos de imagens médicas;
- Comparar os métodos propostos com métodos já consolidados na literatura.

1.2 Contribuições científicas

Os métodos propostos englobam uma série de contribuições, das quais pode-se destacar:

- Avanços na técnica de *template matching*, tornando-o adaptável a cada fatia de uma imagem de TC;
- Uso da técnica de superpixel *intrinsic manifold simple linear iterative clustering* (IMSLIC) pela primeira vez em imagens médicas e com resultados promissores;
- Proposição de uma arquitetura de CNN capaz de classificar esses superpixels na medula espinhal;
- Realce do esôfago a partir de técnicas de filtragem de processamento de imagens, uma vez que seus limites não são bem definidos em TC;
- Proposição de uma arquitetura de FCN capaz de segmentar de forma eficaz o esôfago;
- Dois métodos robustos capazes de detectar automaticamente a medula espinhal em TC de planejamento;
- Um método robusto capaz de realçar e detectar automaticamente o esôfago em TC de planejamento.

Assim, apresenta-se métodos eficientes para serem utilizados no planejamento de sistemas de radioterapia baseados em TC capazes de detectar automaticamente a medula espinhal e o esôfago e, assim, ajudar a poupar tempo e auxiliar aos especialistas nesta tarefa.

1.3 Organização do trabalho

Além deste capítulo de introdução, o trabalho ainda está organizado em mais cinco capítulos, que completam esta tese.

No Capítulo 2 serão apresentados estados da arte e como nosso método preenche lacunas nos trabalhos já existentes.

O Capítulo 3 apresentará conceitos fundamentais pertinentes para o entendimento dos métodos propostos.

No Capítulo 4 serão apresentados os métodos usados para a segmentação da medula espinhal em imagens de TC e os resultados alcançados.

No Capítulo 5 será apresentados o método usado para a segmentação do esôfago em imagens de TC e os resultados alcançados.

Por fim, no Capítulo 6 serão apresentadas as conclusões do trabalho, onde é feito uma avaliação dos métodos propostos e planos futuros para melhoramento dos mesmos.

2 Trabalhos relacionados

Neste capítulo, serão apresentados os trabalhos relacionados tanto a medula espinhal quanto a esôfago. Vale ressaltar que métodos que fazem a detecção/segmentação destes órgãos com OARs em TC de planejamento ainda são escassos. Assim, alguns trabalhos que fazem segmentação destes órgãos tanto em TC convencional quanto em ressonância magnética (RMI) serão apresentados, com o a finalidade de mostrar a importância das suas segmentações.

2.1 Trabalhos relacionados à medula espinhal

A literatura indica que a eficiência dos sistemas CAD está intimamente relacionada às técnicas que os compõem. Além disso, estudos que enfocam a segmentação da medula espinhal podem ser vistos na literatura, seja usando a abordagem de imagens por TC ou mesmo para a segmentação da medula espinhal em RMI. Trabalhos atuais considerados relevantes no campo científico serão apresentados.

Segundo Archip et al. (2002), o planejamento preciso da radioterapia implica a definição de volumes de tratamento e uma delimitação clara do tecido normal, cuja exposição desnecessária deve ser evitada. Eles propuseram uma abordagem *knowledge-based* para identificar a medula espinhal em imagens de tomografia computadorizada do tórax. A abordagem baseia-se em uma base de conhecimento que consiste em um chamado mapa de estruturas anatômicas (do inglês, *anatomical structures map* - ASM) e uma arquitetura orientada a tarefas chamada de *plan-solver*. O ASM contém uma representação do conhecimento em forma de moldura da macro-anatomia no tórax humano. Experimentos realizados no material de imagem de 23 pacientes mostram que a abordagem resulta em um reconhecimento confiável da medula espinhal, com 92% de acurácia, e do canal medular, 85% de acurácia. O trabalho apresenta métricas relevantes de acurácia, contudo, o método é dependente de um conhecimento prévio baseado na estrutura anatômica dos pacientes da base de dados.

Rangayyan, Deglint e Boag (2006) propuseram um método para a segmentação automática 3D do canal medular (onde encontra-se a medula espinhal) em imagens de TC. O método utiliza informações radiológicas e anatômicas a priori, morfologia matemática e técnicas de crescimento de regiões. A transformada de Hough (GONZALEZ; WOODS, 2002)

é aplicada para a detecção de círculos. Em seguida, os círculos detectados são usados para o algoritmo de conectividade difusa para segmentar o canal medular. Em um estudo preliminar, o método segmentou com sucesso o canal medular em oito volumes de TC de quatro pacientes. Este método apresenta várias etapas dependente da estrutura anatômica da base de dados, ainda, não há uma técnica de aprendizado de máquina para determinar com precisão as regiões da medula espinhal.

Banik, Rangayyan e Boag (2010) também propuseram um método que utiliza um conhecimento prévio para detectar a medula espinhal. Eles propuseram um método para realizar a identificação automática da estrutura da costela, da coluna vertebral e do canal medular (onde encontra-se a medula espinhal) em imagens de TC de pacientes pediátricos. A transformada de Hough é utilizada para a detecção de círculos a um mapa de borda cortada que corresponde ao canal medular. Os centros dos círculos detectados são usados para obter as informações necessárias para o algoritmo *fuzzy connectivity* usado para segmentar o canal medular. Os métodos foram testados em 39 exames de TC de 13 pacientes. Os resultados da segmentação da coluna vertebral e do canal medular foram avaliados quantitativa e qualitativamente, comparando com a segmentação realizada por um radiologista sem uso do método proposto. A métrica utilizada neste trabalho foi o Dice, que alcançou uma média de 60%.

Macchia et al. (2012) comparou três softwares comerciais que realizam segmentação de órgãos com base em atlas: VelocityAI, MIM 5.1.1 e ABAS 2.0. Os autores avaliaram a segmentação do software em 15 pacientes. Para comparar os resultados, primeiramente os três softwares executam sua segmentação automática. Em seguida, é feita uma correção manual por um especialista em cada segmentação. Por fim, o Dice é calculado entre as segmentações do software e as correções manuais. Os índices de Dice obtidos por ABAS, MIM e VelocityAI são 70%, 81% e 78%, respectivamente. Contudo, assim como outros trabalhos, utilizam de atlas para segmentar a medula espinhal.

Leener, Cohen-Adad e Kadoury (2015) propuseram uma estrutura automática para a segmentação da medula espinhal e do canal medular, parcelada pelos níveis vertebrais cervical e torácico. O método é adaptado a dois contrastes diferentes na ressonância magnética, com imagens de T_1 e T_2 ponderadas. O método é baseado na transformada de Hough para encontrar elipses em cada corte axial e usar um conhecimento anatômico para preencher as regiões detectadas. Os resultados também são encontrados em comparação com a segmentação manual e eles destacam um Dice de 91%. Apesar da métrica de

Dice ser bem promissora, o trabalho é aplicado a imagens de RMI, onde os limites da medula espinhal estão mais bem definidos. Ainda, o trabalho também utiliza conhecimento anatômico para detecção da medula espinhal.

Embora estudos utilizando abordagens baseadas em conhecimento apresentem resultados satisfatórios, o advento das placas gráficas e o uso crescente da CNNs mostraram-se mais eficientes para a criação de métodos totalmente automáticos. O trabalho proposto por Verhaart et al. (2014) apresenta um método automático para segmentar vários OARs. Para a segmentação da medula espinhal, os autores usam marcações de especialistas como dados de entrada para CNN e, em seguida, pós-processamento para refinar a segmentação usando campos aleatórios de Markov. O método utilizado em 27 TCs e obteve um Dice 87% para segmentação da medula espinhal. O método se mostra muito promissor, alcançando métricas relevantes. Contudo, as métricas calculadas nesse trabalho, levam em consideração apenas fatias que existam medula espinhal, desconsiderando o restante da CT.

Em Ibragimov e Xing (2017), um método é proposto para a segmentação da medula espinhal em TC de cabeça e pescoço usando um CNN. Em primeiro lugar, o método extrai regiões de interesse (regions of interest - ROI) de volumes com base em marcações de especialistas e os usa para treinar uma CNN. A CNN é usada para classificar todos os voxels da ROI, e então uma técnica baseada em campos aleatórios de Markov melhora a segmentação. O método foi aplicado em 50 tomografias e obteve um índice de Dice de 87%. Contudo, este método é aplicado apenas em TCs de cabeça e pescoço.

Em Dong et al. (2019), é proposto um método para segmentação de órgãos de risco em tomografia de planejamento. Os autores propõem o uso de uma rede adversarial juntamente com uma rede convolucional completamente conectada para alcançar seus resultados. Especificamente para a medula espinhal, os autores definem uma ROI baseado na localização dos pulmões classificados por uma primeira rede. Os autores aplicam seu método em uma base com 35 pacientes e alcançam resultados de 90% de Dice.

Em Liu et al. (2020), é proposto um método para segmentação de OARs. São segmentados 8 OARs, são eles bexiga, medula óssea, cabeça femoral esquerda, cabeça femoral direita, reto, intestino delgado e medula espinhal. O método é aplicado a uma base de dados de privada com 105 pacientes chineses. As marcações são feitas por um radiologias e confirmadas por oito oncologistas em um comitê. O método utiliza como técnica de segmentação uma variação da U-Net (RONNEBERGER; FISCHER; BROX, 2015), apresentando como resultado para a segmentação da medula espinhal 82,7% de Dice. O

trabalho apresenta métricas relevantes e mostra a importância de redes FCN para a tarefa de segmentação da medula espinhal.

Pode-se ver nesses trabalhos, que há um grande destaque na segmentação da medula espinhal e a importância dessa segmentação para auxiliar os especialistas. Também é notado que a estes trabalhos usam um conhecimento prévio (como atlas ou informações anatômicas dos pacientes) para elaborar os métodos. Também é mencionado que a literatura carece de estudos que fazem a segmentação da medula espinhal em imagens de TC de planejamento. Outro ponto relevante é que alguns artigos utilizam técnicas de reconhecimento de padrões para a segmentação da medula espinhal, são bastante utilizadas técnicas baseadas em Hough juntamente com conhecimento anatômico e estrutural para fazer a segmentação.

Assim, com base nas lacunas da literatura, propõe-se dois métodos de segmentação totalmente automáticos da medula espinhal em tomografia de planejamento. Sendo o primeiro utilizando *superpixel*, que, por ser técnica de clusterização, ajusta o formato da segmentação aos limites da medula, diferentemente da transformada de Hough que gera círculos. No segundo método apresentado, propõe-se o uso de arquiteturas de FCNs e por ser uma classificação pixel a pixel, garante uma maior precisão na segmentação deste OAR.

2.2 Trabalhos relacionados a esôfago

Estudos de segmentação de esôfago, assim como de medula espinhal, podem ser visto na literatura em várias modalidades de imagens. Também, existem várias abordagens que propõe segmentação do esôfago de forma semi-automática. A maioria dos autores relatam a dificuldade de distinção do esôfago em TC, até mesmo quando se trata de especialista. Vale destacar que, por não haver um contraste bem definido do esôfago em TC, a segmentação se torna uma tarefa árdua e, conseqüentemente, um processo exaustivo quando executada manualmente. Assim, serão apresentados os principais trabalhos que fazem a segmentação do esôfago em tomografias computadorizadas.

Em Feulner et al. (2011), é proposto um método probabilístico para a segmentação do esôfago. De acordo com os autores, o método é constituído por multi-etapas. Primeiro, considera-se que pelo menos uma fatia da TC possui uma abertura de ar do esôfago, então os autores detectam este fenômeno juntamente com a traqueia então uma ROI é delimitada e extraído apenas a ROI do exame. Na próxima etapa, utiliza-se de detecção de elipse para ter uma região prévia do esôfago. Assim, essas informações são inseridas a

um modelo de cadeia de Markov. Em seguida é realizada uma abordagem de “detectar e conectar” para obter a estimativa máxima a posteriori da forma aproximada do esôfago, a partir de hipóteses sobre o contorno do esôfago em fatias axiais de imagem. Finalmente, a superfície dessa aproximação é deformada por um registro a não rígido para melhor se ajustar aos limites do órgão. Os experimentos dos autores são executados em uma base particular com 144 CT e apresentam um resultado de 74% de Dice. Percebe-se que este método é altamente dependente dos modelos descritivos e do conhecimento prévio, ainda, é necessário um registro não rígido, causando deformidades na estrutura inicial do esôfago.

O trabalho proposto por Grosgeorge et al. (2013), também destaca a dificuldade na segmentação do esôfago, uma vez que a parede não é muito bem definida e apresenta baixo contrastes em relação as estruturas que a circundam em TC. De acordo com os autores, é proposto um método original para a segmentação do esôfago. O método alcançou seus resultados usando um modelo em forma de esqueleto para orientar a segmentação. Ainda, o método é composto de duas etapas: uma segmentação 3D por uma técnica de *graph cut* e técnicas de esqueletização, seguido de uma propagação 2D. O trabalho é aplicado a uma base de dados com 6 pacientes e apresenta Dice igual a 61%. Apesar de ser dito um método automático, o mesmo depende da criação de um modelo de esqueleto a partir da marcação de especialista.

O trabalho proposto por Larsson, Zhang e Kahl (2017) propõe um método para segmentação automática de vários órgãos abdominais em TC em uma base de imagens pública, intitulada de MICCAI 2015 (LANDMAN et al., 2015). A primeira etapa do método é a localização dos órgãos, utilizando um modelo de atlas. Para definir o atlas de cada órgão, os autores utilizam registro afim para definirem a localidade. Em seguida, são classificados voxel a voxel por duas CNN. Os autores descartam estruturas baseadas em um limiar e mantêm apenas regiões conectadas. Os autores aplicaram seus experimentos em uma base de 30 pacientes, dentre os vários órgãos segmentados, o esôfago apresenta um Dice de 66,2%. Vale ressaltar, que o trabalho utiliza o atlas como um segmentador inicial, definindo a probabilidade de um voxel ser ou não um determinado órgão se este estiver “casado” com o atlas. Em seguida, a CNN se encarrega de dizer se esses voxels pertencem ou não ao determinado órgão. Neste tipo de abordagem, além de não classificar o volume inteiro, a classificação dos voxels é completamente dependente do atlas.

Em Trullo et al. (2017a), os autores destacam que em uma recente revisão literária, a literatura não apresenta recentes contribuições para a segmentação do esôfago (LITJENS

et al., 2017). De acordo com estes autores, o trabalho se apresenta de forma completamente automática, o método é constituído de apenas duas etapas. A primeira etapa, uma CNN é utilizada para estimar a localização inicial do esôfago, neste primeiro trabalho, os autores utilizaram esta rede e apresentou Dice de 66% para o esôfago (TRULLO et al., 2017b). A partir desta primeira segmentação, é feito um corte nessa estimacão (retirado a *bounding box*), e a mesma rede foi treinada e classificou apenas essa pequena região. O método foi testado em 30 CT e apresentou Dice de 72%. Assim como a maioria dos trabalhos, este destaca a necessidade de um corte na localização do esôfago, todavia, como os próprios autores destacam, a primeira rede não é capaz de localizar corretamente o esôfago, assim o corte é necessário e em seguida, a segunda rede classifica o esôfago. Todavia, ao observar o fato de que a primeira rede possa ignorar a localização do esôfago, ou marcá-lo em região errada, a segunda rede também não será capaz de segmentá-la.

Um dos trabalhos mais recentes que trata a segmentação do esôfago como OAR é o trabalho proposto em Fechter et al. (2017). O método proposto pelos autores utiliza uma CNN para elaboração de um modelo probabilístico que em seguida é utilizado como entrada para um modelo de contorno ativo (do inglês, *active contour model* - ACM). Em seguida, tanto o modelo da CNN quanto o gerado pelo ACM são apresentados a um algoritmo de *random walker* (HORNE; PARKER, 1967) baseado nas escalas de Hounsfield (HU) (BROOKS, 1977) dos volumes. O método é aplicado em uma base de imagens com 50 pacientes, alcançando resultado de 76% de Dice. Apesar do método utilizar CNN como parte de suas etapas, a segmentação final depende da escala de hounsfield da imagem, isso torna o modelo suscetível a erros quando há uma mudança brusca de HU em casos que o processo de aquisição das TCs não seja bem definido.

Em Dong et al. (2019), é proposto um método para segmentação de órgãos de risco em tomografia de planejamento. Os autores propõem o uso de uma rede adversária juntamente com uma FCN para alcançar seus resultados. Especificamente para o esôfago, por compreender apenas uma pequena região do exame de TC, o método faz um corte no volume baseado na localização dos pulmões feita primeiro por sua rede, tendo como premissa que o esôfago se encontra no centroide da fatia com a maior área do que fora classificado como pulmão pela rede. Os autores apresentam seus experimentos em uma base de dados de 36 pacientes, contudo, apenas 35 foram utilizados para elaboração do método. Os resultados são mensurados utilizando a técnica de *leave-one-out*, onde o treinamento é sempre feito com 34 e testado em 1, por fim é calculado a média das 35 execuções.

Dentre os demais OARs, o esôfago apresenta 75% de Dice, 73% de sensibilidade, e 99% de especificidade. O trabalho se mostra bastante promissor, contudo, a segmentação precisa do pulmão é necessária para a extração do volume de interesse onde possa estar o esôfago.

O trabalho proposto por Chen et al. (2019) realizou uma segmentação semântica do esôfago usando um U-Net. Seu banco de dados contém 15 TCs, totalizando mais de duas mil fatias. Embora a segmentação seja 2D, as máscaras resultantes são empilhadas para criar uma 3D. O resultado encontrado é um valor médio de dados de 79%. Apesar de ser um trabalho recente e de utilizar uma rede bastante conhecida, só se aplica a 15 exames.

Por sua vez, Feng et al. (2019) apresenta um método de segmentação usando abordagens de CNN para vários órgãos. O método usa uma corte na região de interesse para cada órgão com base em sua localização conhecida. No final, as segmentações de cada órgão são unidas para gerar a segmentação final de múltiplos órgãos. O método é treinado e testado em um banco de dados composto por 36 TCs. Para o esôfago, o método alcança um valor Dice de 72%. Neste trabalho, observa-se a importância dos cortes nas regiões de interesse do volume para que as segmentações sejam mais precisas. No entanto, o método tem baixo valor de Dice em comparação com outros estudos da literatura.

Lou et al. (2020) também propõe um método de segmentação de esôfago em TC. O método faz um estudo do uso de U-Net e suas variações para esta tarefa. O método usa técnicas de aumento de dados para contornar o problema de pouca amostragem da base de dados. Após o uso das U-Net, é feito um refinamento da segmentação através de processos semi-automáticos para criar a marcação manual. Por fim, é comparado a segmentação da rede com a marcação semi automática e extraídas as métricas de validação, o resultado do Dice foi de 86.80% em 17 pacientes. Apesar da métrica relevante, o trabalho compara a segmentação do método com seu próprio resultado ajustado pelo especialista.

Como pode-se observar na literatura, a maioria dos trabalhos destacam três fatos extremamente importante quanto trata-se de segmentação de esôfago. Primeiro, é um órgão de extrema dificuldade para segmentação, até mesmo para especialistas, pois não há uma distinção clara do contraste da parede do esôfago com os demais órgãos na maioria das fatias. Segundo, muitas pesquisas mostram a necessidade de segmentação deste OAR como uma etapa dificultosa no tratamento por RT e há necessidade urgente de métodos automáticos. Por fim, os trabalhos sempre fazem alguma localização do esôfago baseado seja em modelos probabilísticos, em atlas, ou em técnica de redução de região de interesse (corte).

Assim, baseado na importância da segmentação deste órgão e na necessidade de métodos mais robustos, neste trabalho, propõe-se um método automático de segmentação do esôfago. O método leva em consideração os problemas destacados na literatura, apresentando uma forma eficaz de fazer a localização do esôfago, através do registro das imagens, seguido de uma diminuição do volume de interesse utilizando atlas. Em seguida, ao observar a necessidade de destacar o esôfago das demais estruturas da TC, é proposto uma técnica de pré-processamento para realce das estruturas. Assim, essas imagens são apresentadas para uma arquitetura de FCN, apresentando resultados promissores. Por fim, é utilizado uma técnica para manter na imagem apenas o maior objeto conectado dentro do volume.

3 Fundamentação teórica

Este capítulo trata da fundamentação teórica que expõe conceitos importantes para compreensão dos métodos propostos. Será abordado neste capítulo conceitos como: radioterapia, *template matching*, *superpixel*, rede neural convolucional e métricas de desempenho.

3.1 Radioterapia

A radioterapia (RT) é uma forma comum de tratamento do câncer. De acordo com a *American Society of Radiation Oncology* (ASTRO), dois em cada três pacientes com câncer são tratados com radioterapia, isoladamente ou em combinação com outros métodos terapêuticos, como cirurgia e quimioterapia. Cada caso é único e o médico é a melhor pessoa para decidir sobre a adequação da radioterapia para o seu tratamento (EVANS; STAFFURTH, 2018).

A radioterapia é o uso das radiações ionizantes para destruir ou inibir o crescimento das células anormais que formam um tumor. Existem vários tipos de radiação, porém, as mais utilizadas são as eletromagnéticas (Raios X ou Raios gama) e os elétrons (disponíveis em aceleradores lineares de alta energia). Embora as células normais também possam ser danificadas pela radioterapia, geralmente elas podem se reparar, o que não acontece com as células cancerígenas (EVANS; STAFFURTH, 2018).

O tratamento com radioterapia deve ser sempre cuidadosamente planejado de modo a preservar o tecido saudável, tanto quanto possível. No entanto, sempre haverá tecido saudável que será afetado pelo tratamento, embora estes tecidos possam se regenerar, se expostos a um alto grau de radiação, pode causar possíveis efeitos colaterais. A RT é administrada por uma equipe multidisciplinar, composta por diversos especialistas altamente qualificados, cada um responsável por diferentes cuidados e demandas de cada paciente. Alguns dos membros típicos dessa equipe são: oncologista, radio-oncologista, médico físico, dosimetrista, técnico de radioterapia, enfermeiro oncológico, radiologista, entre outros (EVANS; STAFFURTH, 2018).

Ao planejar o tratamento, o radioterapeuta usará todas as informações obtidas durante o diagnóstico. Ele também pode solicitar alguns exames adicionais para obter

informações sobre o tamanho e localização do tumor e ter uma compreensão mais clara da região do corpo a ser tratada.

Para planejar o tratamento o radioterapeuta precisa inicialmente determinar o tamanho e a localização exata do tumor. Isso geralmente é feito através da criação de uma imagem tridimensional detalhada como, por exemplo, uma tomografia computadorizada.

3.1.1 Tomografia computadoriza de planejamento

Para determinar o tamanho e a localização exata do tumor que se deseja tratar, uma imagem tridimensional do paciente é criada, utilizando um tomógrafo. A tomografia computadorizada gerada, irá guiar o radioterapeuta para produzir o melhor tratamento para cada paciente e cada tipo de tumor.

No tratamento de radioterapia, a TC é uma poderosa ferramenta de auxílio. A tomografia fornece excelente contraste de tecidos moles, permitindo uma grande melhora na localização e definição do tumor. Se o paciente é examinado com tomografia na posição desejada para o tratamento antes da simulação, os limites do campo de tratamento e os parâmetros de colimação podem ser dispostos com respeito à posição do alvo, como determinados pelos cortes de tomografia. O paciente é posicionado de acordo com o seu tratamento, onde é imobilizado com suportes e fitas de fixação, posicionamento dos marcadores radiopacos (bips), onde será definido o isocentro, com referência dos lasers (BUTTON; STAFFURTH, 2010; GWYNNE et al., 2012).

Após o posicionamento, imobilização e definição do isocentro são realizadas radiografias digitais (topograma), onde será programado as sequências de imagens e espessura de corte. Em seguida, será realizado um corte axial para conferir o posicionamento e alinhamento do paciente tomando como referência o isocentro obtido pelos bits. O dosimetrista e o médico radioterapeuta são os responsáveis por delinear as estruturas alvo e os órgãos de risco, de acordo com as imagens da tomografia. Esse delineamento é importante para o físico médico poder planejar a dose de prescrição adequada no volume de tratamento de órgãos e tecidos sadios, dentro das tolerâncias pré-estabelecidas por protocolos internacionais (BUTTON; STAFFURTH, 2010; GWYNNE et al., 2012; EVANS; STAFFURTH, 2018). A Figura 3 ilustra etapas importantes do processo da tomografia de planejamento.

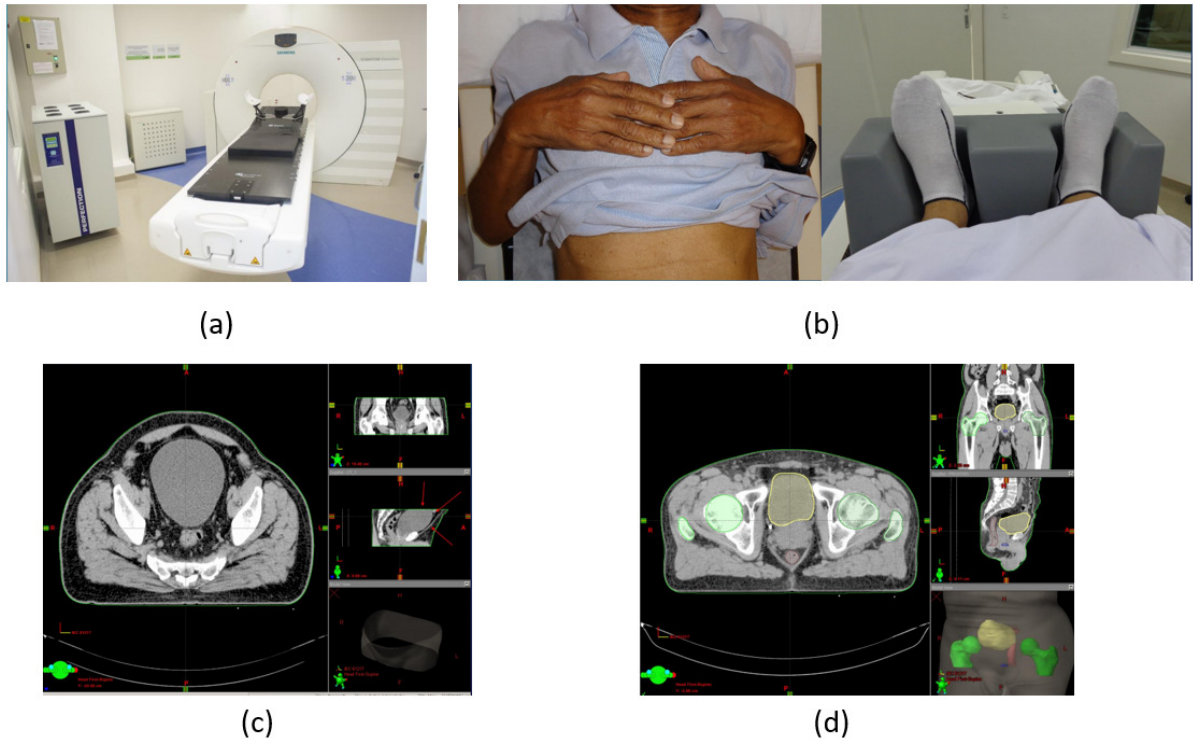


Figura 3 – Tomografia de planejamento: (a) tomógrafo; (b) imobilização do paciente; (c) tomografia computadorizada; e (d) órgãos de risco delineados pelo dosimetrista e médico radioterapeuta.

Os órgãos de risco compreendem órgãos e tecidos saudáveis, que devem ser protegidos dos raios emitidos na radioterapia. Dependendo da localização do tumor, delineia-se alguns órgãos para evitar o excesso de radiação.

3.1.2 Delineamento manual dos Órgão de Risco pelo especialista

Após a aquisição da tomografia computadorizada, o próximo passo é isolar os órgãos de risco. O médico físico, otimiza os feixes da radioterapia em um plano ideal, fornecendo a dose necessária para o volume alvo. Eles otimizam o plano para maximizar a liberação da dose alvo e minimizar a dose nos OARs, usando restrições definidas através do uso da varredura da imagem. Depois que um plano ideal é criado, ele é revisado e aprovado pelo médico responsável pelo tratamento e os dados são transferidos para a máquina de tratamento (EVANS; STAFFURTH, 2018).

Percebe-se que além da necessidade de calibrar os mecanismos da RT para fornecer a quantidade adequada ao volume alvo, existem uma necessidade extrema de proteção de órgãos que venham a envolver o volume alvo. Algumas estruturas do corpo humano

são muito sensíveis à radiação, como o cristalino que tolera níveis baixos de radiação (500 cGy¹). Ultrapassando este nível de dose o paciente poderá desenvolver catarata. Outro exemplo de estrutura sensível é o ovário. Dependendo da idade da paciente, doses ao redor de 800 cGy são suficientes para levar à menopausa precoce.

Ao irradiar o abdome dos pacientes, deve-se estar atentos para alguns órgãos como os rins e o fígado. As doses de tolerância para esses órgãos ficam ao redor de 2000 cGy e 2500 cGy, respectivamente. Doses acima desses valores podem levar à insuficiência renal e hepática. A Figura 4 mostra alguns exemplos de órgão de risco delineados por um especialista.

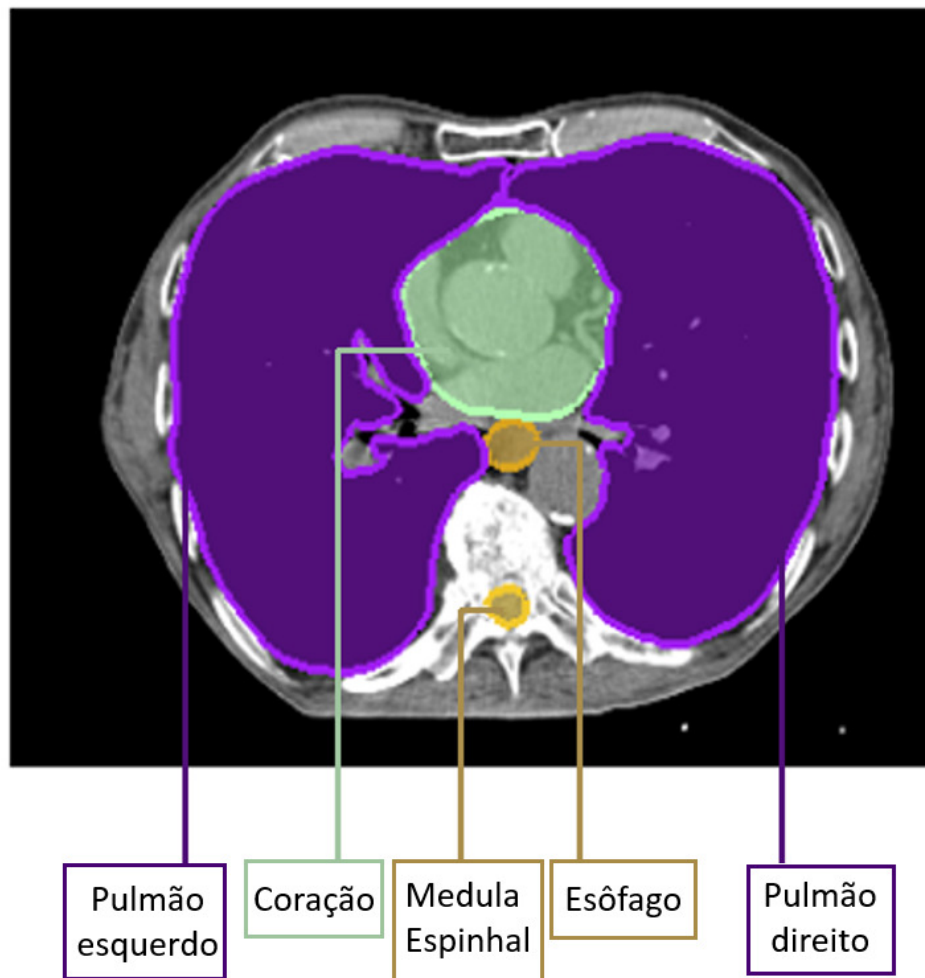


Figura 4 – Exemplo de órgãos de risco delineados por um especialista.

A medula espinhal tolera doses ao redor de 4500 a 5000 cGy, dependendo da extensão da medula que está sendo irradiada. Doses elevadas podem levar à paraplegia ou tetraplegia, dependendo da extensão que foi irradiada. Por sua vez, o esôfago é órgão mais

¹ Unidade de medida de absorção das radiações ionizantes

crítico, a dose recomenda na literatura gira em torno de 50 a 60 cGy (MINSKY et al., 2002; RUPPERT et al., 2010).

Neste trabalho, são propostos dois métodos para segmentação de medula espinhal e um método para segmentação de esôfago. Assim, para cada método, um conjunto de técnicas são aplicadas. Para melhor entendimento, essas técnicas foram divididas em duas seções, a primeira trata dos métodos de segmentação de medula espinhal (Seção 3.2) e o segundo de esôfago (Seção 3.3).

3.2 Segmentação de medula espinhal

Esta seção destaca os conceitos referentes aos métodos de segmentação medula espinhal que serão apresentados no Capítulo 4.

3.2.1 Medula espinhal

Um dos principais OARs para lidar com a radioterapia é a medula espinhal. Este órgão é uma estrutura segmentar com pares de raízes nervosas dorsais e ventrais, entrando e saindo em cada nível, que se estende da junção crânio-cervical até a extremidade da coluna lombar. Os revestimentos da medula espinhal incluem várias partes do sistema nervoso central, incluindo o líquido cefalorraquidiano, solução salina muito pura, pobre em proteínas e células, que serve como uma almofada para o córtex cerebral e medula espinhal. Sua estrutura interna é organizada em duas áreas distintas: massa cinzenta interna com corpos celulares e uma substância branca externa, contendo fibras mielinizadas e não mielinizadas (KHAN, 2010; DOWLATI, 2017; BHATTACHARYYA, 2018).

Mielite transversa induzida por radiação, parestesias, paralisia total, perda neurológica parcial, síndrome de Brown Sequard são algumas doenças causadas por radiação excessiva (MCCUNNIFF; LIANG, 1989). Diante disso, é indubitavelmente a proteção desse órgão em tratamentos radioterápicos.

A literatura aponta diversos trabalhos envolvendo metodologias computacionais para segmentação de outros OARs, auxiliando a radioterapia (ADEBAHR et al., 2016; GÓMEZ-MILLÁN et al., 2015; MCVICAR; POPESCU; HEATH, 2016). Contudo, poucos trabalhos publicados sugerem metodologias computacionais para segmentação de medula espinhal em tomografia computadorizada de planejamento.

A medula espinhal encontra-se no canal vertebral, quando delineada pelos especialistas na tomografia computadorizada, a marcação compreende a base dos limites ósseos do canal vertebral. A Figura 5 mostra uma ilustração das estruturas que envolvem a medula espinhal.

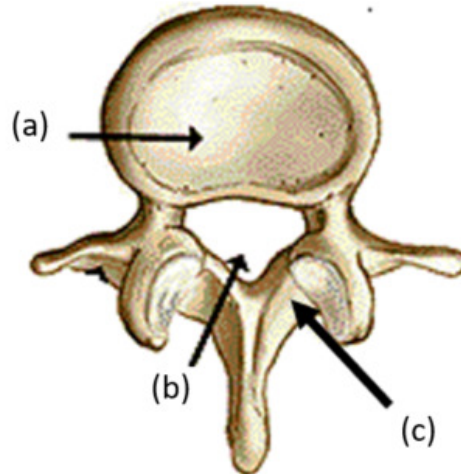


Figura 5 – Medula espinhal: (a) corpo da vértebra; (b) canal vertebral (onde encontra-se a medula espinhal); e (c) osso da vértebra

Fonte (Adaptado): <https://pt.dreamstime.com/ilustração-stock-espinha-vértebra-em-segundo-lombar>

Assim, os métodos propostos para medula espinhal neste trabalho, têm o intuito de prover uma segmentação da medula espinhal, auxiliando o especialista na tarefa de isolar este OAR no tratamento da radioterapia, evitando efeitos colaterais que possam ser irreversíveis em pacientes que estão sendo submetidos a este tipo de tratamento.

3.2.2 *Template matching*

O *template matching* é uma técnica de reconhecimento de padrões baseada na comparação entre imagens. Com base em uma imagem, o *template matching* pesquisará nessa imagem a área que mais se aproxima de um padrão, definida por outra imagem denominada *template* (BRUNELLI, 2009).

O *template matching* tem vários métodos de implementação. Cada um se concentra em um cálculo de métrica de similaridade. Neste trabalho, foi utilizado o método de combinação de diferenças quadradas (BRUNELLI, 2009).

O método de combinação de diferença quadrada é definido na Equação 1.

$$R_{sq_diff}(x, y) = \sum_{x', y'} [T(x', y') - I(x + x', y + y')]^2 \quad (1)$$

onde I é a imagem de entrada, T é o *template* e R é o valor resultante, x e y são as coordenadas de largura e altura, respectivamente.

Analisando a Equação 1, há uma combinação da diferença dos quadrados nas coordenadas x e y na soma. Essas coordenadas em T são equivalente na sub-imagem de I . O resultado obtido em sua aplicação é um valor $R \in [0, \infty]$. Quando há uma correspondência perfeita entre T e a sub-imagem em I , o resultado em R encontrado tem valor 0, e quanto menor a semelhança, maior o valor de R .

A técnica de *template matching* foi utilizada neste trabalho nos dois métodos de segmentação de medula espinhal para diminuir a região de interesse do volume, tornando mais fácil a segmentação da medula espinhal.

3.2.3 *Intrinsic manifold simple linear iterative clustering* - IMSLIC

Superpixels são regiões uniformes perceptualmente significativas que podem efetivamente capturar características da imagem. Entre vários métodos para calcular superpixels, o algoritmo *simple linear iterative clustering* (SLIC) é popular devido à sua simplicidade e alto desempenho (ACHANTA et al., 2012). Um recente trabalho proposto por Liu et al. (2018) apresentou uma extensão do SLIC, chamado *Intrinsic Manifold SLIC*, *Intrinsic MSLIC* ou IMSLIC.

De acordo com Liu et al. (2018), superpixels geralmente têm as seguintes características:

1. Conectividade: cada superpixel é uma região simplesmente conectada e cada pixel na imagem é atribuído a exatamente um superpixel;
2. Compacidade: na região sem característica, os superpixels são regulares em tamanho e forma;
3. Preservação de características: os superpixels devem aderir bem aos limites das imagens;
4. Sensibilidade de conteúdo: a densidade de superpixels é adaptável à co-ocorrência de conteúdos de imagem;

5. Performance: computar os superpixels deve ser rápido, eficientes em termos de memória e escalar em imagens de alta resolução;
6. Fácil de usar: os usuários simplesmente especificam o número desejado de superpixels e não devem ser incomodados pelo ajuste de outros parâmetros

O algoritmo IMSLIC desenvolvido por Liu et al. (2018) é implementado em C++ e disponível na biblioteca do OpenCV 3.3. Este algoritmo precisa apenas de dois parâmetros: número inicial de superpixel e o fator de compacidade do superpixel.

Neste trabalho, o IMSLIC foi utilizado para segmentar candidatos à medula espinhal no primeiro método de segmentação da medula espinhal. Após segmentados, os superpixels são apresentados a um classificador. Neste trabalho, utilizou-se as redes neurais convolucionais como classificador.

3.2.4 Redes neurais artificiais - RNA

As redes neurais artificiais (RNA) são modelos matemáticos que usam uma coleção de unidades computacionais simples, chamadas de neurônios artificiais, interligados em uma rede. Esses modelos são comumente usados na tarefa de reconhecimento de padrões.

O trabalho proposto em McCulloch e Pitts (1943) foi o primeiro trabalho envolvendo as RNAs e suas peculiaridades.

De acordo com Hafemann (2014), a motivação para McCulloch e Pitts (1943) estudarem redes neurais, foi o fato de que o cérebro humano era superior a um computador em muitas tarefas, uma afirmação que se mantém ainda hoje em dia para tarefas como o reconhecimento de objetos e rostos, apesar dos enormes avanços na velocidade de processamento em modernos computadores.

3.2.4.1 Neurônio artificial

Neurônio artificial é a unidade básica de uma RNA e é utilizado para construção de modelos mais poderosos. Um único neurônio implementa uma função matemática dada suas entradas, para fornecer uma saída.

Um neurônio artificial pode ser expresso em termos matemáticos pela Equação 2:

$$f(x) = \sigma\left(\sum_{i=1}^n x_i w_i + b\right) \quad (2)$$

onde x_i é a entrada i , w_i é o peso associado a entrada i , b é o termo bias e σ é a função não-linear.

Alguns exemplos de funções não-lineares são a função sigmoide e a função tangente hiperbólica. Porém, recentemente, uma nova e simples função não-linear foi proposta em Krizhevsky, Sutskever e Hinton (2012a), chamada de *Rectified Linear Units* (ReLU) e, de acordo, com os autores essa função converge até 6 vezes mais rápida que RNAs que utilizam a função tangente hiperbólica. Essa função é apresentada na Equação 3.

$$\sigma(x) = \max(0, x) \quad (3)$$

3.2.4.2 *Multilayer Perceptron* - MLP

Uma *Multilayer perceptron* (MLP) é um tipo de rede neural artificial composta por múltiplas camadas de nós, com cada camada totalmente conectada à próxima. Cada nó (exceto os nós de entrada) é um neurônio artificial com uma função de ativação não-linear. De acordo com Hornik, Stinchcombe e White (1989), as MLPs são aproximadores universais, ou seja, são capazes de aproximar qualquer função mensurável para qualquer grau de precisão desejado.

Basicamente, as MLPs são compostas por uma camada inicial de entrada, seguida por uma ou mais camadas ocultas/escondidas e a última camada seria a saída da rede. Cada camada é totalmente conectada com sua camada adjacente e produz um vetor de saída de acordo com o vetor da camada anterior. A saída de cada camada é calculada aplicando a função de ativação de cada neurônio em todos os neurônios da camada e a Equação 4 descreve a saída de uma camada.

$$y^l = \sigma(W^l y^{l-1} + b^l) \quad (4)$$

onde y^l é o vetor de saída, W^l é a matriz dos pesos atribuídos a cada par de neurônio da camada l e $l - 1$, e b^l é o vetor dos termo bias de cada neurônio da camada l .

As MLPs utilizam uma técnica de aprendizagem supervisionada chamada *backpropagation* para treinar a rede.

3.2.4.3 *Backpropagation*

Durante o treinamento de uma MLP, uma função de erro é definida. Esta função calcula o erro das previsões do modelo em relação a um conjunto de dados. O objetivo do treinamento é minimizar a soma das funções de erro aplicadas a todos os exemplos de um conjunto de dados.

O algoritmo de *backpropagation* é baseado na regra de aprendizagem por correção de erros. De acordo com Silva (2004), o algoritmo utiliza pares de entradas e saídas desejadas e, por meio de um mecanismo para correção dos erros, ajusta os pesos da rede, assim, para a minimização do erro obtido pela rede e o ajuste dos pesos, o algoritmo utiliza a regra de delta generalizada, com aplicação do gradiente.

Ainda segundo Silva (2004), no treinamento com o algoritmo de *backpropagation*, a rede opera em uma sequência de dois passos: primeiro, um padrão é apresentado à camada de entrada da rede, a atividade resultante flui através da rede, camada por camada até que a resposta seja produzida pela camada de saída; segundo, a saída obtida é comparada à saída desejada para esse padrão, se esta não estiver correta, o erro é calculado. O erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado.

Resumidamente, os passos do algoritmo de aprendizagem por retropropagação (algoritmo de *backpropagation*) são:

1. Ajustar os pesos dos elementos de processamento com pequenos valores aleatórios;
2. Apresentar as entradas, um vetor $x_0, x_1, \dots, x_N$ de medidas, e especificar um vetor de saída desejado;
3. Calcular as saídas da rede $y_1, y_2, \dots, y_N$ definidas pela equação dos neurônios artificiais (Equação 2);
4. Reajustar os pesos. Usar um algoritmo recursivo começando pelos neurônio da saída, retropropagando da última à primeira camada. Os pesos são ajustados através da equação $w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j x_i$; onde w_{ij} é o peso do neurônio oculto j em um

dado tempo t , x_i pode ser tanto um neurônio de saída quanto um de entrada, η é a taxa de aprendizagem e δ_j é um termo de erro para o neurônio j . Se j for um neurônio de saída, então $\eta = y_j(1 - y_j)(d_j - y_j)$; onde d_j denota a saída desejada e y_k é a saída real da rede; se j for um neurônio oculto, então $\delta_j = x_j(1 - x_j) \sum_k \delta_k w_{jk}$, onde k denota todos os neurônios acima do neurônio j ;

5. Repetir retornando para o passo 2.

A taxa de aprendizagem influencia a magnitude das mudanças dos pesos, desempenhando papel fundamental no aprendizado. Uma taxa de aprendizado pequena implica em pequenas variações, tornando o treinamento lento e aumentando as chances de paradas em mínimo locais; altas taxas de aprendizado, no entanto, podem levar a MLP à saturação ou mesmo à oscilação, comprometendo todo o processo de aprendizado. Esta taxa de aprendizagem é introduzida na rede com o objetivo de permitir maior rapidez na convergência ao erro desejado, enquanto o erro estiver diminuindo, e ao mesmo tempo evita que a rede venha oscilar, diminuindo a taxa de aprendizagem quando o erro tende a aumentar (SILVA, 2004).

3.2.5 Aprendizado profundo

A profundidade de uma arquitetura de rede neural refere-se ao número de operações não-lineares que compõe esta rede. Enquanto muitas aplicações bem sucedidas tenham usado redes “rasas” (até 3 camadas), a organização do cérebro de um mamífero é uma arquitetura de profundidade (BENGIO, 2009).

De acordo com Hafemann (2014), as redes neurais profundas têm sido investigada por décadas, porém produziram resultados pobres, até muito recentemente. Foi observado em muitas experiências que as redes profundas são mais difíceis de treinar do que as redes “rasas” e que as redes profundas muitas vezes ficam presas em mínimos locais aparentes, quando se inicia com uma aleatoriedade nos parâmetros da rede. Assim, foi descoberto que os melhores resultados podem ser alcançados através de um pré-treino de cada camada com um algoritmo de aprendizagem não-supervisionada.

Muitos trabalhos com pré-treino não supervisionado são demonstrados com comparações estatísticas em Bengio et al. (2007) e Larochelle et al. (2007), até que, recentemente, uma rede neural profunda que foi treinada apenas com aprendizado

supervisionado começou a registrar resultados semelhantes em algumas tarefas. Cireşan et al. (2011) demonstram que as redes neurais profundas devidamente treinadas são capazes de alcançar melhores resultados em muitas tarefas.

De acordo com Hafemann (2014), para classificação de imagens, os melhores resultados publicados usam um tipo de arquitetura chamada de rede neural convolucional.

3.2.5.1 Redes neurais convolucionais

As redes neurais convolucionais (*Convolutional Neural Network* - CNN), de acordo com LeCun et al. (1998a), combinam a ideia de 3 arquiteturas distintas: pesos compartilhados, subamostragem espacial/temporal e campos receptores locais. Elas são compostas de, basicamente, três camadas: convolução, subamostragem e completamente conectadas.

As características são extraídas de uma camada para outra da rede, onde um neurônio na camada atual conecta-se a um campo receptivo local na camada anterior. Dessa forma, as características são combinadas sequencialmente para, assim, serem extraídas características de mais alto nível (FERNANDES, 2013). Vale lembrar, que as camadas convolucionais da CNN se encarregam de extrair característica de baixo nível, em suas camadas iniciais, como aquelas de textura simples e de bordas. A medida que mais camadas de convolução são adicionadas, características de alto nível são extraídas. Desta forma, a CNN consegue reproduzir a maioria das características simples de textura e geometria (LECUN; BENGIO; HINTON, 2015a). Um exemplo de uma CNN é demonstrado na Figura 6 e em seguida são explicados detalhes sobre cada camada.

3.2.5.1.1 Camada de convolução

As camadas de convolução são compostas por filtros treináveis que são aplicados em toda a imagem de entrada. Para cada filtro, pode existir um neurônio que é ligado com a saída de um subconjunto de neurônios da camada anterior. Por exemplo, dado uma imagem como entrada, os filtros definem uma pequena área (3×3 , 5×5 , 7×7 *pixels*) e cada neurônio é ligado somente aos neurônios nas proximidades da camada anterior. Os pesos são compartilhados entre os neurônios, levando os filtros a aprender os padrões

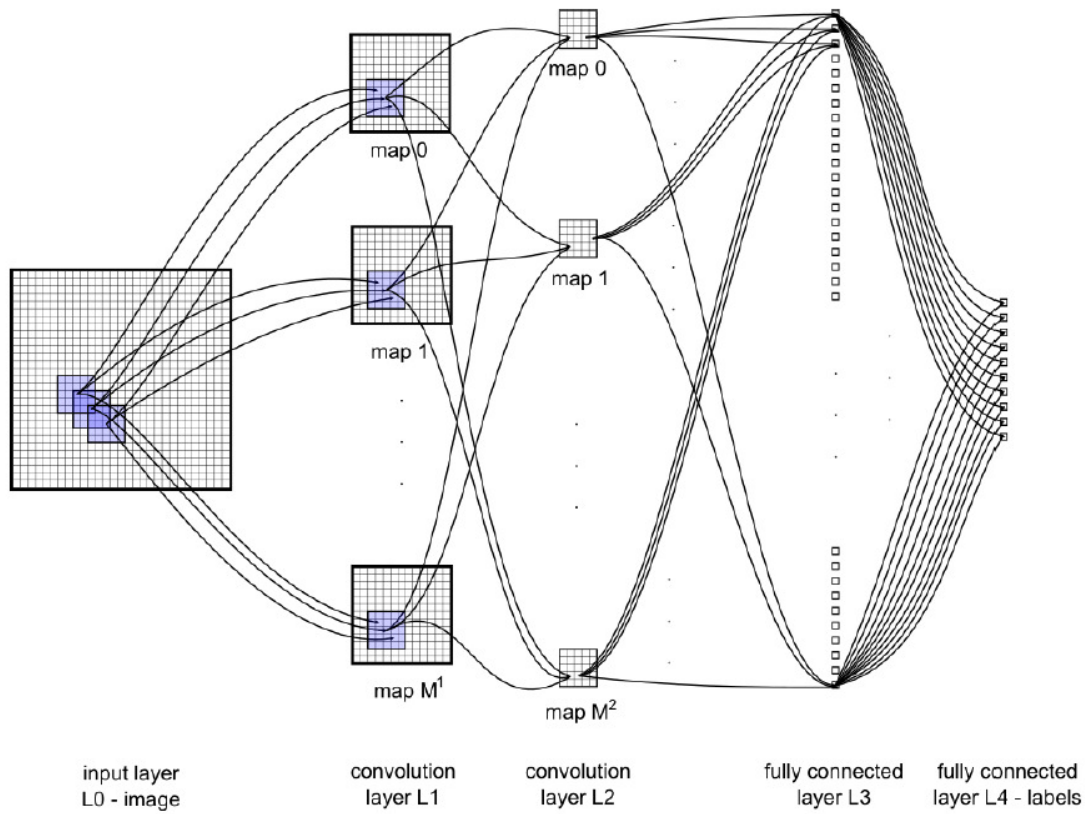


Figura 6 – Exemplo de uma CNN.

Adaptado de (HAFEMANN, 2014)

frequentes que ocorrem em qualquer parte da imagem. A Figura 7 ilustra como ocorre a convolução em uma imagem.

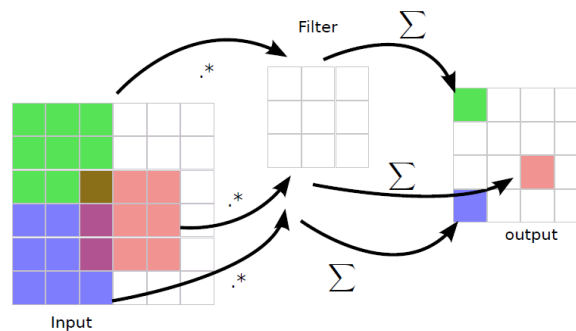


Figura 7 – Ilustração de uma operação de convolução.

Adaptado de (HAFEMANN, 2014)

Cada filtro treinável é aplicado individualmente em cada imagem de entrada gerando vários mapas de características, ou seja, cada filtro é responsável por detectar um tipo de característica da imagem.

3.2.5.1.2 Camada de subamostragem

De acordo com Hafemann (2014), as camadas de subamostragem implementam uma função de redução da solução não-linear, a fim de reduzir a dimensionalidade e capturar pequenas invariâncias da imagem.

É na camada de subamostragem que há a redução da resolução espacial dos mapas de características e onde serão selecionadas as características invariantes à deslocamentos e distorções (LECUN et al., 1998a). Um tipo muito utilizado de subamostragem é mostrado na Figura 8.

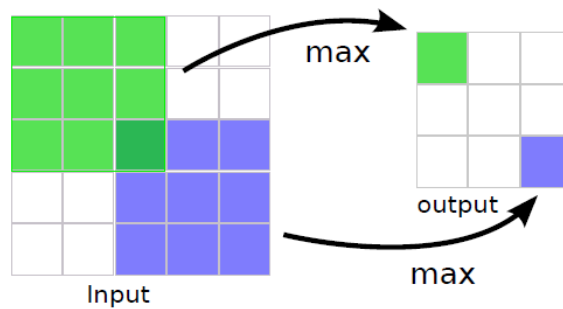


Figura 8 – Ilustração da camada de subamostragem.

Adaptado de (HAFEMANN, 2014)

O tipo de subamostragem mostrado na Figura 8 é conhecido como máxima ativação (*max pooling*), em que apenas o *pixel* de maior intensidade do campo receptivo anterior é mantido, isto é, em uma máscara 2×2 com 4 valores, apenas o maior valor de *pixel* dentre eles ficará na imagem atual.

Ainda, existem as operações de *up conv*, que geralmente são operações de *up sampling*. Basicamente funcionam como o inverso das camadas de *max pooling*, ou seja, dobram o tamanho do mapa de característica da camada anterior por operações de interpolação (LONG; SHELHAMER; DARRELL, 2015). Em rede com etapas de contração e expansão (semelhante a que será apresentada a Seção 3.2.5.2), são utilizadas as informações concatenadas do caminho de contração para esta interpolação.

3.2.5.1.3 Camada completamente conectada

Ao término das sequências de camadas de convolução e subamostragem que têm a tarefa de extrair características da imagem, os *pixels* de todos os mapas de características da camada anterior são dados como entrada às camadas completamente conectadas, que por sua vez, são responsáveis pela classificação dos padrões de entrada, semelhante à MLP.

Neste trabalho, a CNN foi utilizada para classificar as regiões segmentadas pelo IMSLIC como regiões de medula e não medula. As arquiteturas baseadas no modelo mostrado na Figura 6 estabelecem o estado da arte para o problema de classificação de imagens.

3.2.5.2 U-Net

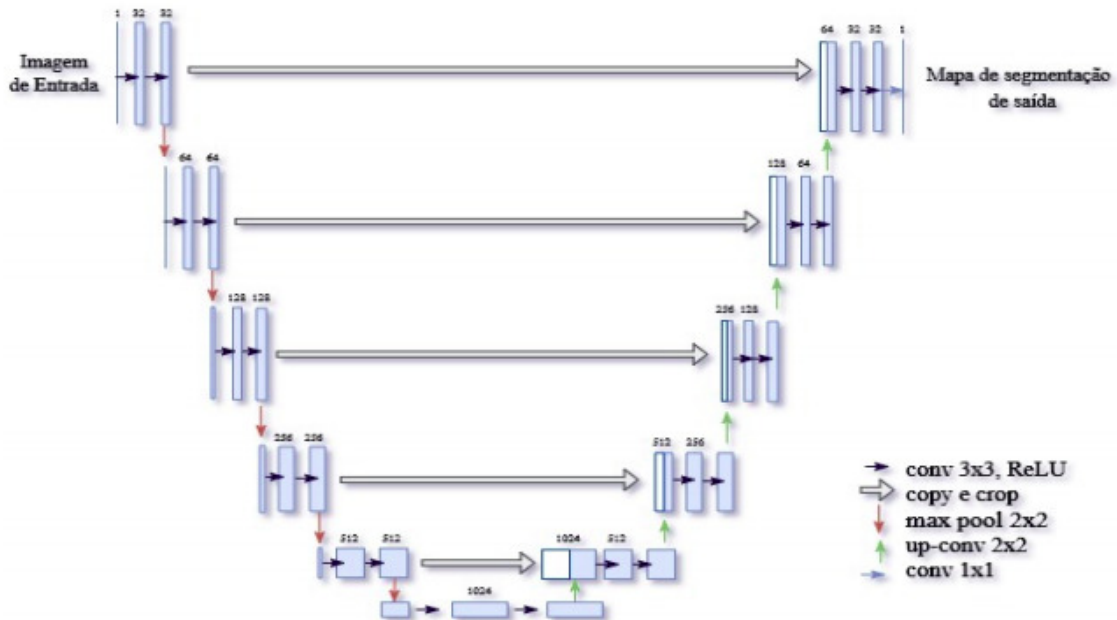
Ronneberger, Fischer e Brox (2015) apresentaram a U-Net, que é uma RNA completamente convolucional proposta para segmentação de imagens biomédicas. A U-Net é uma rede totalmente convolucional, cuja arquitetura consiste em um caminho de contração para capturar o contexto, e um caminho de expansão simétrico que permite uma segmentação precisa. A principal estratégia que diferencia a U-Net das outras arquiteturas FCN é a combinação entre os mapas de características do estágio de contração e seus correspondentes simétricos no estágio de expansão, permitindo a propagação de informações de contexto para os mapas de características de alta resolução. Esse modelo vem superando os métodos existentes em diversos desafios biomédicos de segmentação de imagens (RONNEBERGER; FISCHER; BROX, 2015). A Figura 9 ilustra a arquitetura e as camadas que constituem a rede.

A arquitetura consiste em um caminho de contração (lado esquerdo) para capturar o contexto e um caminho de expansão (lado direito) simétrico que permite uma localização precisa. O caminho de contração é uma arquitetura tradicional de uma CNN, várias aplicações repetidas de duas convoluções 3x3, cada uma acompanhada por uma função de ativação ReLU e uma operação de subamostragem com ativação máxima de 2x2, reduzindo a largura e altura da imagem, com *stride* 2 para subamostragem. Em cada etapa de subamostragem, duplica o número de canais de características.

O caminho expansivo consiste em um levantamento do mapa de características, seguido de um sobre-amostragem 2x2 que faz a metade do número de canais de

Figura 9 – Representação da Arquitetura U-Net.

Fonte: (RONNEBERGER; FISCHER; BROX, 2015)



características, que é uma junção com o mapa de características correspondentemente do caminho de contração e duas convoluções 3x3, cada uma seguida por uma função de ativação ReLU. Essa etapa do caminho expansivo é importante devido à perda de *pixels* de borda em cada convolução no caminho de contração. Na última camada é utilizada uma convolução 1x1 para mapear cada vetor de características para o número desejado de classes.

Em resumo, essa rede pode ser treinada de ponta a ponta a partir de poucas imagens, onde a U-Net simplesmente concatena os mapas de características do codificador para mapear mapas de características do decodificador, em todas as etapas, para formar uma estrutura como escada. Essa arquitetura por suas conexões de concatenação permite que o decodificador em cada etapa aprenda as características relevantes que são perdidas quando agrupadas no codificador. Neste trabalho, a U-Net foi utilizada no segundo método de segmentação da medula espinhal e no método de segmentação de esôfago.

3.3 Segmentação de esôfago

Nesta seção, destacam-se os conceitos necessário para compreensão do método de segmentação em esôfago que será descrito no Capítulo 5.

3.3.1 Esôfago

Outro órgão de risco que apresenta inúmeras dificuldades de segmentação não só nas tomografias de planejamento, mas em tomografias convencionais, é o esôfago. O esôfago é um órgão oco, de aproximadamente 25 cm de comprimento e 3 de largura. É constituído de tecido muscular e conjuntivo, sendo rico em vasos sanguíneos e glândulas mucosas.

A esofagite é um dos problemas mais comuns relacionado a excesso de radiação no esôfago foi relatado por Werner-Wasik et al. (2010). Apesar de comum, a esofagite aguda (ocorrendo < 90 dias após o início do tratamento) em pacientes submetidos à radioterapia (TR) para tumores torácicos, esta pode exigir hospitalização, testes de diagnóstico invasivos (por exemplo, endoscopia), intervenção cirúrgica (por exemplo, tubo de gastrostomia endoscópica percutânea) ou quebras de RT que podem diminuir o controle local do tumor (WERNER-WASIK et al., 2010). Lesões tardias no esôfago são relatadas com menos frequência, talvez porque os pacientes possam não viver o tempo suficiente para manifestar toxicidade (por exemplo, a sobrevida específica da doença é relativamente curta para muitos cânceres torácicos) (WERNER-WASIK et al., 2010).

O esôfago é um órgão que traz muita dificuldade para ser demarcado, em algumas fatias do volume, até mesmo os especialistas têm dificuldade em definir seus limites de maneira confiável, logo, o esôfago pode ser facilmente confundido com outras estruturas. Isso leva a uma interpretação entediante das imagens de TC, o que torna o processo demorado e altamente propenso à variabilidade interobservadores. No entanto, com sua mucosa sensível à rádio, o esôfago é um dos OAR mais críticos na RT, e um contorno confiável é indispensável (FECHTER et al., 2017; GROSGEORGE et al., 2013). Um exemplo da dificuldade de localização do esôfago é apresentado por Larsson, Zhang e Kahl (2017) e exemplificado na Figura 10.

Diante disto, o presente trabalho apresenta um método para segmentação automática do esôfago em imagens de tomografia computadorizada de planejamento a radioterapia, desta forma, auxiliando o especialista e, conseqüentemente, diminuindo os impactos negativos neste OAR.

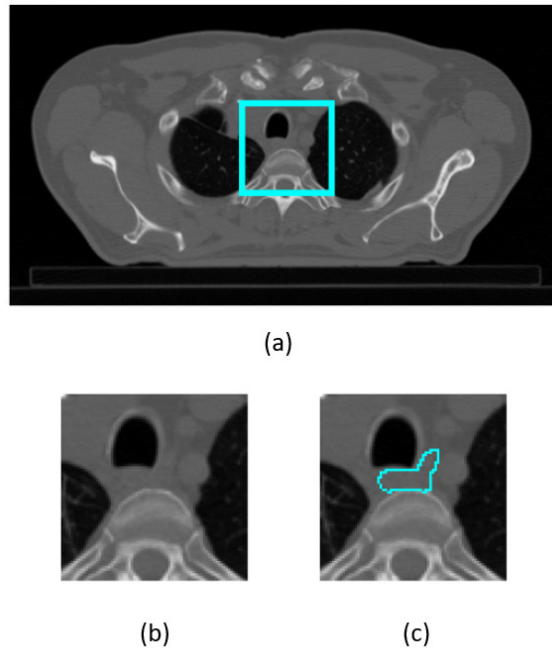


Figura 10 – Esôfago: (a) fatia de uma tomografia computadorizada (em azul uma boundingbox com a localização do esôfago); (b) corte na região do esôfago; e (c) marcação do especialista

3.3.2 Pré-processamento de imagens

Essa seção apresenta algumas técnicas de pré-processamento de imagens utilizadas ao longo do método proposto.

3.3.2.1 Filtragem em imagens

O processo de filtragem na imagem, de acordo com Gonzalez e Woods (2002), são técnicas baseadas em convolução de *templates* (janelas, matrizes, máscaras) ou tuplas (conjunto de pixels). Neste trabalho, foram utilizados duas técnicas: filtro bilateral e equalização do histograma.

3.3.2.1.1 Equalização do histograma

De acordo com Gonzalez e Woods (2002), equalização do histograma nada mais é que criar uma imagem com níveis de cinza igualmente distribuídos ao longo da escala de cinza. A sua principal função é melhorar a qualidade da imagem.

Os níveis de cinza de uma imagem podem ser considerados como variáveis aleatórias no intervalo $[0, 1]$. Se os níveis de cinza forem variáveis contínuas, os valores originais e transformados podem ser caracterizados por suas funções densidade de probabilidade $p_f(f)$ e $p_g(g)$, respectivamente. Dada uma imagem contendo $n = M * n$ pixels, assumindo valores discretos para os níveis de cinza $k = 0, 1, \dots, L - 1$, uma forma de equalização do histograma é utilizar a função acumulada de probabilidade, a qual pode ser expressa pela Equação 5.

$$g_k = T(f_k) = \sum_{i=0}^k p_f(f_i) = \sum_{i=0}^k \frac{n_i}{n} \quad (5)$$

onde n_i é o número de ocorrências do nível de cinza i , e $p_f(f_i)$ é a probabilidade do i – *esimo* nível de cinza.

Para que a função de distribuição acumulada de probabilidade possa ser utilizada para equalizar uma imagem, deve-se normalizar os níveis de cinza no intervalo de $[0, 1]$.

3.3.2.1.2 Filtro Bilateral

O filtro bilateral foi proposto em 1998 por Tomasi e Manduchi (1998). Resumidamente, o filtro bilateral suaviza as imagens enquanto preserva as bordas, por meio de uma combinação não linear de valores de imagem próximos. Combina os níveis de cinza com base na proximidade geométrica e na semelhança da imagem e, prefere valores próximos a valores distantes no domínio e no intervalo. O resultado é uma imagem suavizada em regiões homogêneas, mas com bordas preservadas. Ainda de acordo com Tomasi e Manduchi (1998), o filtro bilateral é capaz de reduzir o ruído em uma imagem em uma ordem de magnitude, mantendo as bordas.

O filtro bilateral foi utilizado em conjunto com uma variação de equalização do histograma para realçar a região do esôfago de forma a deixar mais perceptível as bordas deste OAR.

3.3.2.2 Morfologia matemática

A morfologia matemática (ou somente morfologia) é o ramo do processamento de imagens que concentra-se nas estruturas geométricas da imagem. De acordo com Gonzalez e Woods (2010), a morfologia é usada predominantemente em: pré-processamento de

imagens (filtragem, ruído, simplificação de formas), segmentação de objetos, detecção da estrutura dos objetos e descrição quantitativa de objetos (área e perímetro).

Ainda segundo Gonzalez e Woods (2010), para compreender a morfologia é necessária a definição de dois operadores: reflexão e translação.

A reflexão do conjunto B , denotada por $\widehat{B}$ é definida como:

$$\widehat{B} = \{w | w = -b, b \in B\} \quad (6)$$

A translação do conjunto A pelo ponto $z = (z_1, z_2)$, denotada por $(A)_z$, é definida como:

$$(A)_z = \{c | c = a + z \in A\} \quad (7)$$

A base da morfologia matemática consiste em extrair as informações relativas à geometria e a topologia de um conjunto desconhecido (no caso uma imagem) pela transformação com outro conjunto bem definido, chamado elemento estruturante. Existem dois principais operadores morfológicos: erosão e dilatação.

Para os conjuntos A e B em Z^2 , a erosão de A por B é definida como:

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (8)$$

ou seja, a Equação 8 indica que a erosão de A por B é o conjunto de todos os pontos z tal que B , transladado por z , está contido em A . Onde A é a imagem original e B o elemento estruturante.

Para os conjuntos A e B em Z^2 a dilatação de A por B é definida como:

$$A \oplus B = \left\{ z | [(\widehat{B}_z \cap A)] \subseteq A \right\} \quad (9)$$

isto é, a Equação 9 indica que a dilatação de A por B é o conjunto de todos os deslocamentos de z , tal que a reflexão de B intercedida por A é sobreposta por pelo menos um elemento. Novamente A é a imagem original e B o elemento estruturante.

Neste trabalho, foi utilizado a erosão na etapa de refinamento da segmentação do esôfago.

3.3.2.3 Registro de imagens

O registro de imagens tem sido de grande importância em diversas áreas, em especial na área médica. Ele tem sido empregado em segmentação, diagnóstico e também na análise temporal de doenças.

A literatura emprega diversas definições ao registro de imagens. Brown (1992) define o registro de imagens como o processo de transformação de diferentes conjuntos de dados para um mesmo sistema de coordenadas. Os dados podem ser várias fotografias, dados de diferentes sensores, de diferentes épocas, ou a partir de diferentes pontos de vista.

Crum, Hartkens e Hill (2014) por sua vez, explica o registro de imagens como um processo para a determinação da correspondência entre as características de imagens recolhidas em momentos diferentes ou utilizando diferentes métodos de imagem. As correspondências podem ser usadas para alterar a aparência - por rotação, translação, alongamento etc. - de uma imagem, de forma que mais se assemelhe a outra a fim de que o par possa ser diretamente comparado, combinado ou analisado. O uso mais intuitivo de registro é para corrigir posições diferentes de pacientes entre as varreduras.

O registro de imagens é composto pelos seguintes elementos: duas imagens, sendo uma fixa e uma móvel, uma medida de similaridade, uma função de transformação, de um otimizador e um interpolador.

A medida de similaridade é uma medida numérica responsável por dizer o quanto a imagem móvel está similar a fixa. A função de transformação é a função que mapeia a imagem móvel na imagem fixa. Essa função de transformação é calculada pelo otimizador, otimizando certa medida de similaridade. Dessa forma, o registro pode ser visto como um problema de otimização, no qual um otimizador estima a função transformação que melhor mapeia a imagem fixa da imagem móvel, de acordo com um valor obtido pela medida de similaridade.

Historicamente, os registros dividem-se em dois grupos: registros rígidos e registros deformáveis (ou não rígidos) (CRUM; HARTKENS; HILL, 2014). Os registros rígidos consideram que a imagem móvel é um objeto que precisa ser rotacionado e/ou transladado para que haja uma correspondência satisfatória com a imagem fixa. Ou seja, esse tipo de registro usa transformação global. Os registros deformáveis, por outro lado, obtêm essa correspondência através de algumas deformações localizadas. Esse por sua vez usa transformação local.

3.3.2.3.1 Registro rígido

As aplicações do registro rígido são aplicações simples que se resume em operações de rotação e translação, de forma que a imagem móvel não é deformada, e sim posta no mesmo sistema de coordenadas da imagem fixa.

O registro rígido, por si só, não é o mais adequado para a maioria dos problemas de desalinhamento de imagens. Isso ocorre porque ele não corrige erros locais que apenas seriam resolvidos por meio de deformações. No entanto, ele é comumente usado como pré-processamento de registros deformáveis, possibilitando que os mesmos sejam executados com maior eficiência e menor custo computacional. Isso acontece porque, uma vez que a imagem móvel está no mesmo sistema de coordenadas da fixa, as deformações passam a ser mais locais e menos frequentes.

Geralmente, a transformação rígida é feita utilizando versores (JOHNSON et al., 2007) e matrizes (NETTO et al., 2016). Neste trabalho adotou-se a formulação por matrizes por ser utilizada na maioria dos trabalhos que usam processamento de imagens.

Mais formalmente, a transformação de coordenadas de imagem móvel, $M(x)$, nas coordenadas da imagem fixa $F(x)$ é dada pela Equação 10:

$$x' = T(x) = R_x(\Theta_x)R_y(\Theta_y)R_z(\Theta_z)x + t \quad (10)$$

onde R_x , R_y e R_z são as matrizes de rotação em torno do eixo x , y e z , respectivamente, considerando um ângulo Θ e t é o vetor translação. As matrizes de rotação são representadas pelas Equações 11, 12 e 13.

$$R_x(\Theta_x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Theta_x & -\sin\Theta_x \\ 0 & \sin\Theta_x & \cos\Theta_x \end{pmatrix} \quad (11)$$

$$R_y(\Theta_y) = \begin{pmatrix} \cos\Theta_y & 0 & \sin\Theta_y \\ 0 & 1 & 0 \\ -\sin\Theta_y & 0 & \cos\Theta_y \end{pmatrix} \quad (12)$$

$$R_z(\Theta_z) = \begin{pmatrix} \cos\Theta_z & -\sin\Theta_z & 0 \\ \sin\Theta_z & \cos\Theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (13)$$

O registro rígido foi utilizado neste trabalho na etapa de padronização dos volumes na segmentação do esôfago.

3.3.3 Atlas

O atlas corresponde a distribuição espacial dos valores de *voxels* pertencentes a algum objeto de interesse. Por exemplo, existe uma base de tomografias computadorizadas de vários pacientes, para cada paciente o médico especialista faz uma marcação específica considerando a estrutura anatômica de cada paciente. Ao gerar o atlas da marcação de um especialista, cada elemento representa a probabilidade amostral do *voxel* pertencer ao objeto. Para a construção do atlas denotado por A , considera-se $F_i(x, y, z)$ o *voxel* da posição (x, y, z) do i -ésimo volume que pertence ao objeto, definindo $P(x, y, z)$ como a probabilidade amostral do *voxel* da posição (x, y, z) fazer parte do objeto, que é dada pela Equação 14

$$P(x, y, z) = \frac{\sum_{i=0}^n F_i(x, y, z)}{n} \quad (14)$$

onde n é a quantidade de volumes.

A Figura 11 demonstra a criação do atlas baseado em marcações do esôfago de vários pacientes. Quanto mais branco, maior incidência da marcação naquele local.

Neste trabalho, o atlas é utilizado apenas para verificar a maior região volumétrica que compreenderia todo um esôfago, desta forma, pode-se extrair uma região 3D, chamada de volume de interesse (VOI), onde encontra-se localizado espacialmente o esôfago em vários pacientes.

3.4 Métricas de desempenho

Para validar os resultados obtidos por um método proposto, comumente adota-se métricas de validação de resultados. Essa avaliação é um processo de comparação das medidas obtidas durante a etapa de classificação do método.

Geralmente, as métricas utilizadas para validação de um modelo derivam do resultado de uma matriz de confusão (Figura 12). Apesar de existirem métricas que medem a distância entre superfícies (distância de Hausdorff, distância média, mediana e padrão da superfície), estas são expressas em milímetros e calculadas considerando espaçamento



Figura 11 – Exemplo de um atlas com a marcação do esôfago.

entre fatias e outras medidas, o que muitas vezes sofre pelo processo de aquisição de imagens e de difícil interpretação, principalmente pela assimetria e erros de subestimação em diferentes regiões durante o processo de avaliação (WANG; WANG; ZHU, 2020). Assim, em nosso trabalho adotamos as métricas de Dice, sensibilidade, especificidade e acurácia. Vale destacar que o Dice é a mais utilizada para verificar a efetividade de métodos de segmentação e é a métrica mais comparada para este fim (TAHA; HANBURY, 2015).

As medidas mencionadas são as mais utilizadas na área de processamento de imagens médicas.

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	VP Verdadeiro Positivo	FP Falso Positivo
	negativos	FN Falso Negativo	VN Verdadeiro Negativo

Figura 12 – Matriz de confusão.

- Dice é uma estatística usada para comparar a similaridade de duas amostras. No nosso caso, para comparar a marcação de especialista e a marcação do método (Equação 15). Esta métrica também é conhecida como F-score, F1-sScore e F-measure.

$$Dice = \frac{2VP}{2VP + FP + FN} \quad (15)$$

- Sensibilidade é a proporção de verdadeiros positivos: a capacidade do sistema em prever corretamente a condição para casos que realmente a tem. O cálculo da sensibilidade é feito de acordo com a Equação 16.

$$SEN = \frac{VP}{VP + FN} \quad (16)$$

- Especificidade é a proporção de verdadeiros negativos: a capacidade do sistema em prever corretamente a ausência da condição para casos que realmente não a tem. O cálculo da especificidade é feito com na Equação 17.

$$SPE = \frac{VN}{VN + FP} \quad (17)$$

- Acurácia é a proporção de predições sem levar em consideração o que é positivo e o que é negativo. O cálculo da acurácia é feito com a Equação 18.

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (18)$$

4 Segmentação de medula espinhal

Para a tarefa de segmentação de medula espinhal, são propostos dois métodos. Ambos os métodos fizeram uso de uma base de dados pública proveniente de um desafio, que está descrita na Seção 4.1. O primeiro método é baseado em *superpixel* e redes neurais convolucionais (Seção 4.2) e segundo baseado em arquiteturas U-Net (Seção 4.3).

4.1 Materiais

O base de imagens de TC de planejamento usado neste trabalho é uma base de imagens de um desafio chamado *AAPM Thoracic Auto-segmentation* (YANG et al., 2018). O principal objetivo deste desafio é incentivar os participantes a criar algoritmos para a segmentação de órgãos de risco (OARs), em TC de planejamento de pacientes, que serão submetidos a tratamento radioterápico. Um exemplo de uma fatia de um volume está apresentado na Figura 13.

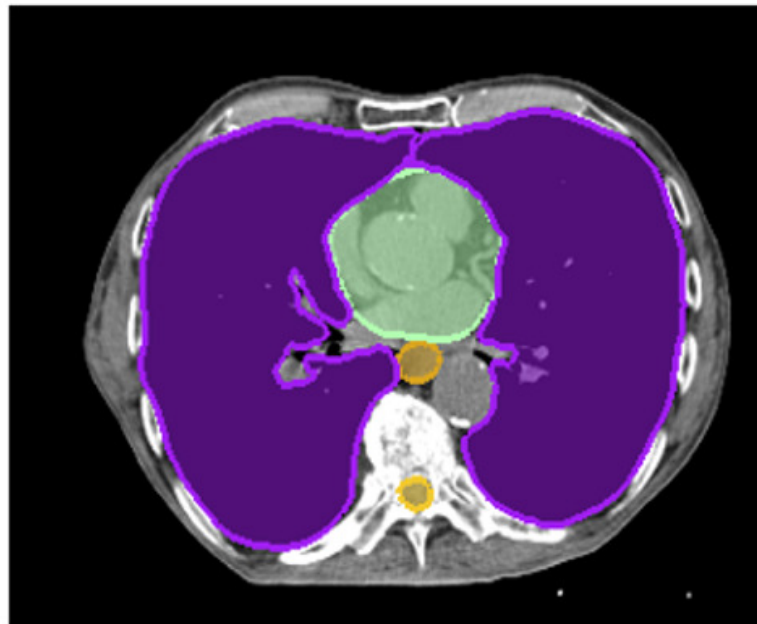


Figura 13 – Aquisição de imagem: pulmões (roxo), coração (verde), esôfago (amarelo superior) e medula espinhal (amarelo inferior).

O desafio é composto por 36 pacientes de três institutos diferentes (12 volumes de pacientes para cada instituto). A base de imagens fornece a marcação dos OARs em um arquivo denominado *Radiotherapy Structure Set* (RT-STRUCT) para cada volume da

base de imagens. Cada arquivo tem a marcação de 5 órgãos (pulmão esquerdo, pulmão direito, coração, esôfago e medula espinhal).

Uma das variedades dessa base de imagens é o número de fatias que podem ir de 134 a 288, algumas com fatias da cabeça e outras não. Além disso, alguns volumes têm a medula espinhal em todas as fatias, enquanto outros não. Essa informação mostra o quanto diversificado é a base de imagens. Exemplos de pacientes dos três institutos podem ser vistos na Figura 14.

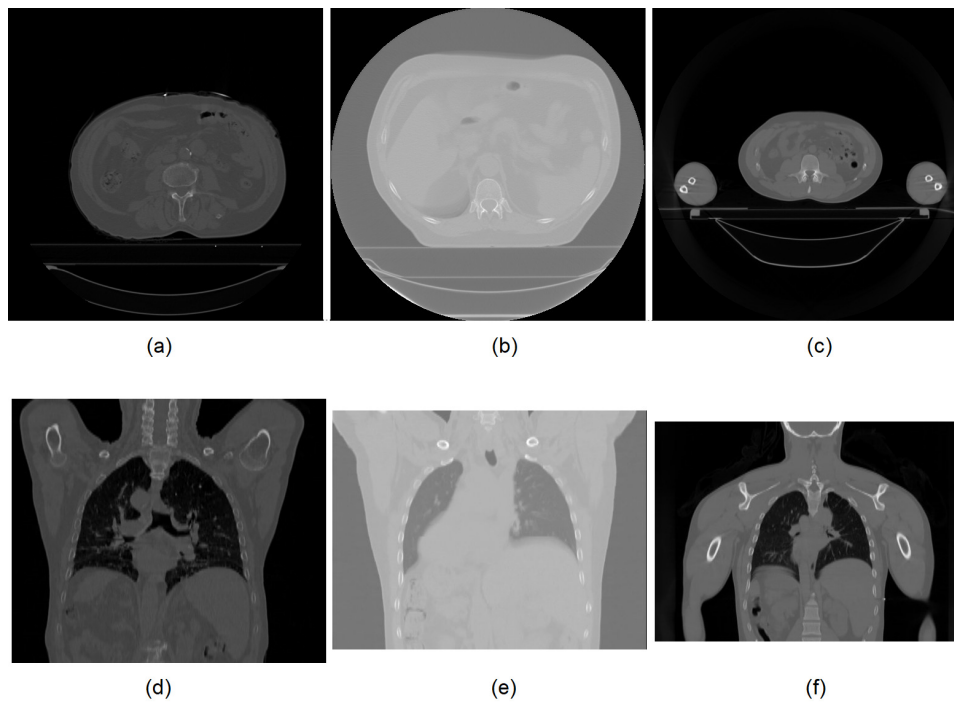


Figura 14 – Aquisição de imagem: (a) fatia do instituto 1; (b) fatia do instituto 2; (c) fatia do instituto 3; (d) volume do instituto 1; (e) volume do instituto 2; e (f) volume do instituto 3.

A marcação da medula espinhal na base de imagens é contornada com base nos limites ósseos do canal vertebral. Acredita-se que as marcações foram feitas desta forma, pelo fato da dificuldade de observar a medula espinhal nos exames de TC. Além disso, o fato dessa segmentação se basear no planejamento da radioterapia, o protocolo para delimitar a medula espinhal é até o limite do canal vertebral, para que o especialista saiba onde calibrar a radioterapia para não atacar a medula espinhal.

4.2 Segmentação de medula espinhal utilizando *template matching* adaptativo IMSLIC e CNN

O primeiro método proposto para a segmentação da medula espinhal em TC de planejamento é basicamente dividido em quatro etapas. Essas etapas são descritas na Figura 15. Resumidamente, o primeiro passo é a aquisição de imagens que se refere a base de dados descrita no Seção 4.1. O segundo passo, é a segmentação inicial, essa etapa é feita para reduzir a região de interesse dos volumes isolando a região da medula espinhal, para isso, utiliza-se uma técnica de *template matching*. A terceira etapa, segmentação de candidatos, regiões são agrupadas usando a técnica de *superpixel* IMSLIC. A quarta etapa, classificação dos candidatos, regiões agrupadas pelo IMSLIC serão utilizadas para treinamento e classificação em uma rede neural convolucional (CNN), sendo, finalmente, extraídas as métricas de validação para quantificar o método proposto. Nas próximas seções, serão explicados mais detalhes de cada etapa.

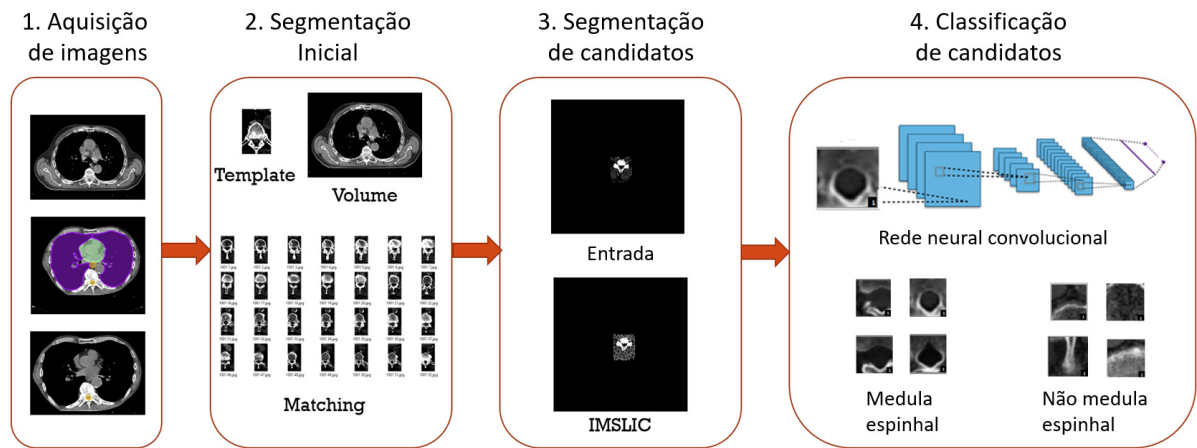


Figura 15 – Fluxograma do primeiro método do medula espinhal.

4.2.1 Segmentação inicial

Esta etapa visa delimitar a região inicial de interesse, que é a medula espinhal, juntamente com todas as estruturas ao seu redor. Isto é importante, uma vez que, trabalhar com todas as estruturas do volume, tornaria o método mais custoso e mais suscetível a erros, ao isolar a região de interesse (ROI), menos regiões da TC são exploradas e, conseqüentemente, há uma diminuição no custo computacional.

Para isso, uma técnica muito comum no contexto de processamento de imagem e reconhecimento de padrões é usada, essa técnica é chamada de *template matching*. Além disso, foram propostas melhorias nessa técnica para que ela se tornasse robusta na tarefa de segmentar a ROI composta pela medula espinhal, chamou-se esta melhoria de *template matching* adaptativo. Conhecendo as características do *template matching* apresentada na Seção 3.2.2, nesta seção será explicado como ele é usado neste método.

4.2.1.1 Gerando o *template* padrão

Define-se a imagem de *template* padrão que foi gerada a partir de um volume de um paciente da base de imagens. Essa imagem é criada apenas uma vez, de forma que seja genérica para todos os pacientes. Basicamente, um volume é selecionado entre todos os volumes da base de imagens, aleatoriamente. Então uma fatia deste volume é escolhida e um *template* padrão é construído da seguinte forma: primeiro, a medula espinhal é encontrada utilizando uma marcação disponível uma vez que, anatomicamente, todo paciente possui medula espinhal. Após encontrada a medula espinhal, o centro de massa desta marcação é encontrado, agora uma região correspondente a duas vezes o tamanho da marcação é cortada em todos os lados desta fatia, definiu-se este tamanho como margem de segurança para não haver perda da medula entre as fatias. Este modelo será usado como o ponto inicial para nosso algoritmo de *template matching* adaptativos. A Figura 16 ilustra a definição do *template* padrão gerado em um volume.

Depois de definir o *template* padrão, a próxima etapa é encontrar o *template* inicial para cada volume.

4.2.1.2 Gerando o *template* inicial

Com o *template* padrão definido, um *template* inicial será gerado para cada volume. Este processo acontece da seguinte forma: primeiro, o algoritmo de *template matching* executa em cada fatia do volume, onde o *template* é o *template* padrão (definido na Seção 4.2.1.1); segundo, a similaridade de cada *matching* em cada fatia é calculado (combinação de diferença quadrada, conforme apresentado na Seção 3.2.2); terceiro, o *template* inicial é definido como o *matching* de maior similaridade em relação ao *template* padrão, também o número da fatia que gerou o *template* inicial é salva. Ao fim desta etapa, tem-se o *template*

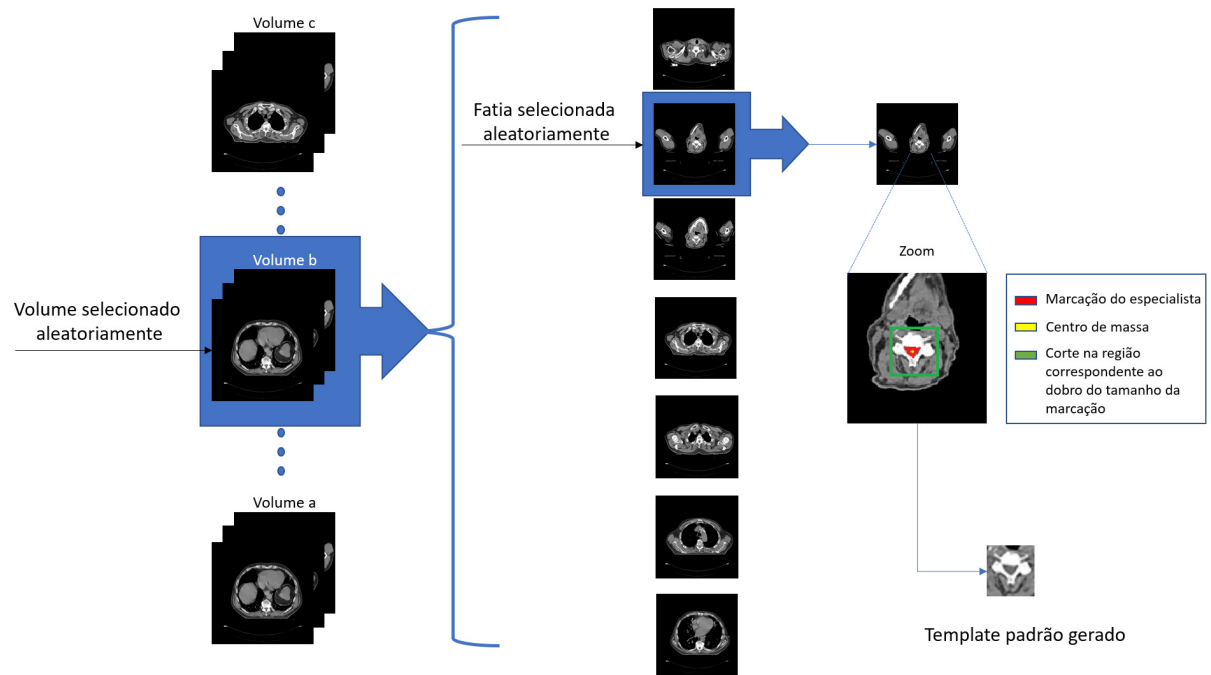


Figura 16 – Geração do *template* padrão.

inicial, que será único para cada volume. A Figura 17 ilustra como o *template* inicial é gerado.

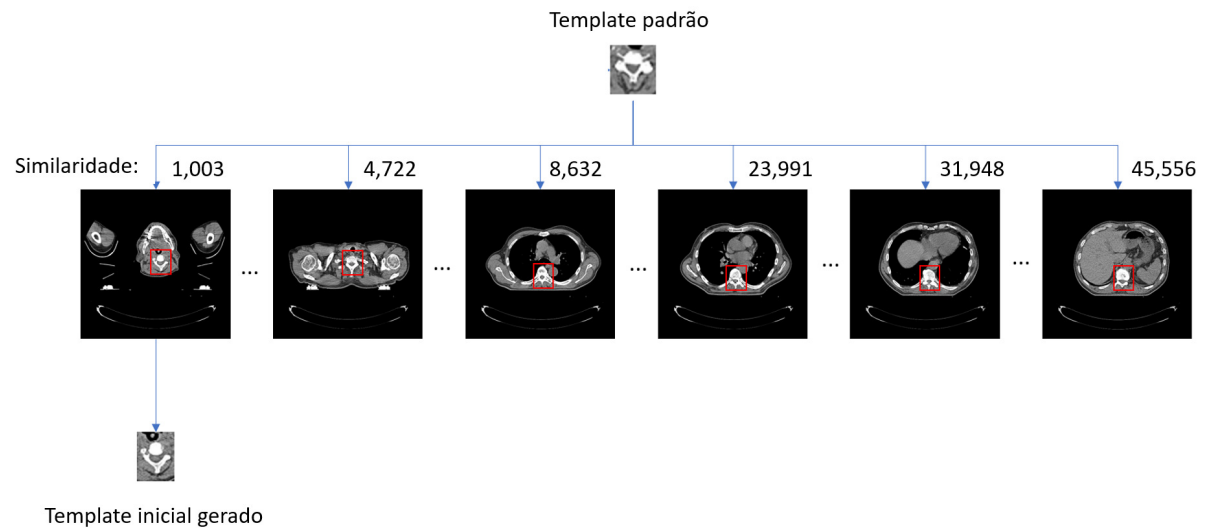


Figura 17 – Geração o *template* inicial.

4.2.1.3 Adaptando o *template* para cada fatia

Com o *template* inicial (Figura 18a) e o número da fatia onde houve o melhor *matching*, a próxima etapa é executar este *template* no respectivo volume. O *template matching* começará a ser executado na fatia em que ocorreu o *matching* (Figura 18b), resultando em dois artefatos: o novo *template* (que será adaptável para cada fatia de cada volume, Figura 18c) e a imagem resultante (Figura 18d). O novo *template* executará na próxima fatia, novamente, o resultado será um novo *template* e uma imagem resultante. Deve-se observar que a fatia em que o *template* inicial ocorre pode ser no centro do volume, então o mesmo processo acontece para as fatias anteriores. A Figure 18 ilustra este processo.

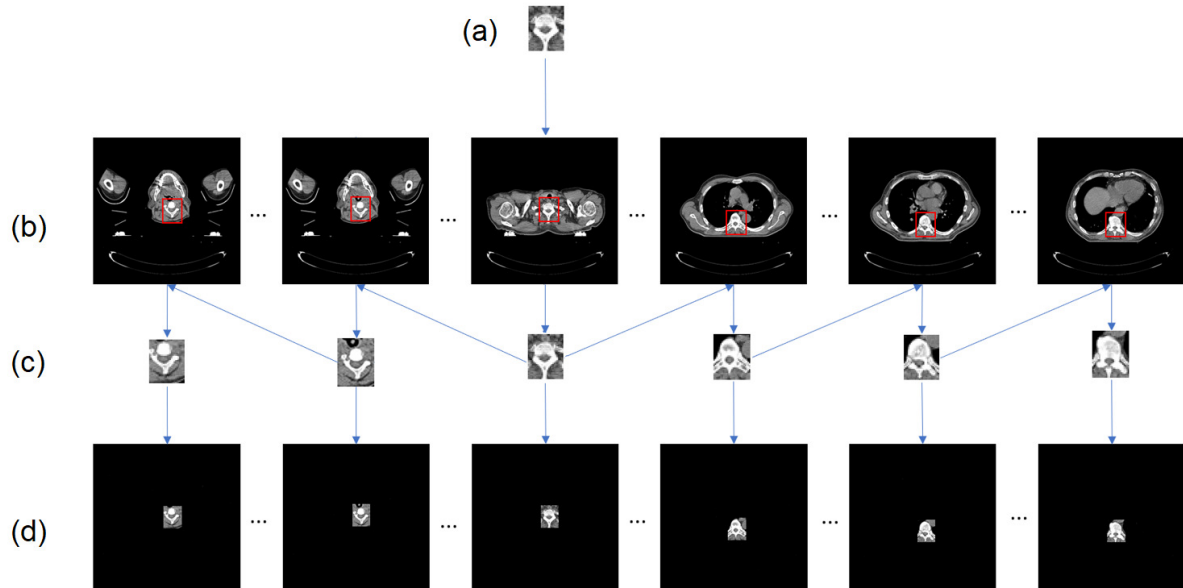


Figura 18 – Adaptando o *template* para cada fatia: (a) *template* inicial; (b) fatias de um volume (o *bounding box* vermelho mostra onde o *matching* ocorreu); (c) resultado do *matching* que gera o novo *template*; e (d) fatia resultado.

Depois do *template matching* adaptativo executar em todas as fatias de um volume, o resultado será um volume onde terá somente a região em volta da medula espinhal.

4.2.2 Segmentação dos candidatos à medula espinhal

Nesta etapa, a medula espinhal é segmentada. Para isso, um algoritmo de *clustering* baseado em *superpixel* chamado *Intrinsic manifold simple linear iterative clustering*

(IMSLIC) é utilizado (Seção 3.2.3). Este algoritmo foi utilizado pela sua capacidade de agrupar pixels vizinhos com base na intensidade, isso é de extrema vantagem quando se sabe que os valores de intensidade da medula espinhal diferem das outras regiões ao seu redor. De acordo com Liu et al. (2018), dentre as técnicas de *superpixel*, o IMSLIC é a técnica que utiliza o menor número de parâmetros, além de ser uma técnica recente e considerada a melhor técnica quando comparada a outras variações do SLIC. Assim, este algoritmo agrupará os pixels vizinhos com base na intensidade e na distância espacial. Esses grupos separados podem corresponder à medula espinhal ou às regiões restantes da TC, e esses grupos são conhecidos como *superpixel*.

Como falado anteriormente, o algoritmo foi usado para criar grupos que serão candidatos a serem classificados na próxima etapa. Basicamente, os grupos de *superpixels* são divididos em dois: aqueles correspondentes à medula espinhal (chamada de grupo da medula espinhal) e aqueles que não correspondem à medula espinhal (grupo de não medula espinhal).

Então, a partir da saída das fatias da segmentação inicial, o algoritmo IMSLIC é executado, gerando grupos dentro da região correspondente ao *template matching* adaptativo. Um critério a ser considerado é que os *superpixels* que compõem a borda da segmentação inicial são ignorados e não são levados em conta na classificação dos candidatos (*superpixels* marcados com círculo vermelho na Figura 19), uma vez que estes correspondem apenas as bordas da segmentação inicial e não englobam regiões da medula espinhal. Um exemplo desse passo pode ser visto na Figura 19.

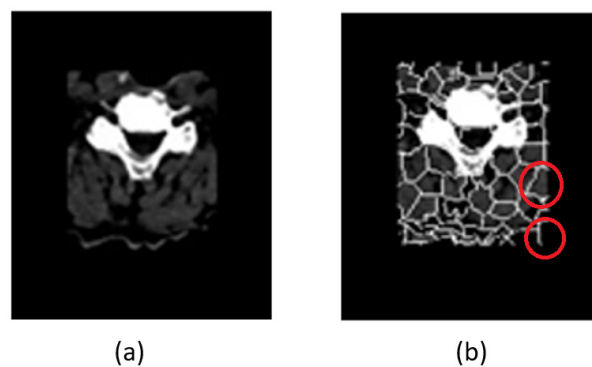


Figura 19 – Segmentação de candidatos a medula espinhal: (a) fatia de saída da segmentação inicial (b) aplicação do algoritmo IMSLIC na fatia.

O próximo passo classificará essas várias regiões em medula espinhal e não medula espinhal.

4.2.3 Classificação dos candidatos a medula espinhal

Após a segmentação dos candidatos, o próximo passo é classificá-los. Os candidatos (superpixels) serão classificados em regiões de medula espinhal ou regiões de não medula espinhal. Para esta tarefa, a técnica de aprendizado de máquina de redes neurais convolucionais (LECUN et al., 1998b) é utilizada (Seção 3.2.5.1).

A rede neural convolucional (CNN) tem sido aplicada em um grande número de aplicações de visão computacional (GUO et al., 2016; TOMPSON et al., 2015; TAIGMAN et al., 2014; WANG et al., 2012) e também foram introduzidos no campo da imagem médica (DINIZ et al., 2018; LI et al., 2014; BROSCHE et al., 2014; ROTH et al., 2014; SILVA et al., 2017). Essa rede é um tipo particular de rede profunda e *feedforward* que consiste em muitas camadas convolucionais, intercaladas com camadas de agrupamento que reduzem a dimensionalidade do sinal de entrada, algumas camadas totalmente conectadas e uma camada de classificação final. Essa abordagem ignora a etapa de projetar e extrair um determinado conjunto de características criados manualmente e, em vez disso, alimenta os dados por imagem diretamente apresentados as redes profundas. (LECUN et al., 1998b; LECUN; BENGIO; HINTON, 2015b).

A seguir, é explicado com mais detalhes como as imagens de entrada da CNN foram criadas e como os procedimentos de treinamento e teste foram projetados e executados.

4.2.3.1 Gerando as imagens de entrada

Muitas técnicas de aprendizado de máquina utilizam como atributos de entrada características extraídas das imagens. Nesses casos, para obter bons resultados, é necessária uma efetiva engenharia de atributos para que eles sejam representativos o suficiente para discriminar a classe da imagem. Uma das vantagens da CNN é que ela não precisa de uma etapa explícita de extração de atributos. Em vez disso, a rede neural convolucional usa a própria imagem como entrada e tenta extrair implicitamente os melhores atributos.

Os candidatos em nosso método são *superpixel*. Portanto, é necessário gerar imagens que os representem e que sejam adequados como entrada para a CNN. Uma possível imagem para representar um *superpixel* seria formada através das *bounding boxes* desse *superpixel*.

Sabe-se que a marcação da medula espinhal na base de imagens é contornada com base nos limites ósseos do canal vertebral que é envolvido por outro conjunto de estruturas, como as vértebras (corpo da vértebra e osso da vértebra), ao formar essa *bounding box*, deve-se observar essa informação, uma vez que, tais estruturas têm um comportamento muito peculiar na TC, pode ser visto na Figura 20, que a medula espinhal (que se encontra dentro do canal vertebral) é cercada por várias estruturas brancas, o que é uma característica muito importante para integrar as *bounding boxes*.



Figura 20 – Geração de imagens de entrada: as estruturas ao redor da medula espinhal.

Assim, ao formar as *bounding boxes*, opta-se por levar em conta não apenas os valores de pixel correspondentes aos superpixels, mas preenche-se toda a *bounding box* com os valores de pixel ao redor desse *superpixel*. Para isso, utiliza-se o centróide do *superpixel* formado juntamente com a largura e altura que deseja criar esta *bounding box*.

Pode-se listar algumas vantagens em criar as *bounding box* desta maneira: (a) a *bounding box* terá informações não só dos superpixels, mas também dos pixels vizinhos que compreendem valores importantes, uma vez que a medula espinhal tem estruturas brancas circundantes; (b) sabe-se que a CNN faz vários redimensionamentos na imagem de entrada até atingir a camada totalmente conectada, utilizando apenas informações de *superpixel*, ao chegar na camada totalmente conectada, os valores de *superpixels* poderiam ser tão pequenos que não haveria informação suficiente para que a camada conseguisse generalizar suas características; também (c) sabe-se que um dos requisitos da CNN é que as imagens de entrada tenham a mesma dimensão, portanto ao criar uma *bounding box* de igual altura e largura, as imagens já estão prontas para serem apresentadas à CNN, após várias sessões de treinamento foi percebido que o melhor tamanho para as *bounding box* era 30×30 .

O tamanho de 30×30 foi escolhido, porque o tamanho do *superpixel* gerado na etapa de segmentação do candidato é sempre menor que esse tamanho. O tamanho das imagens de entrada é originalmente as *bounding box* de *superpixels* com o voxel que o envolve. Como os *superpixels* nem sempre são do mesmo tamanho, os *bounding boxes* dos *superpixels* também seriam diferentes. Portanto, foi necessário preencher as imagens menores com uma borda para que todas as imagens de entrada tenham o mesmo tamanho. Para isso, foi calculada a maior caixa delimitadora de todos os superpixels. Depois de todo esse processo, calculou-se que todas as imagens de entrada deveriam ter o tamanho de 30×30 para a base de imagens usada. Assim, as imagens que serão submetidas à CNN terão informações valiosas para que a rede aprenda um padrão.

Exemplos dessas *bounding boxes* podem ser vistas na Figura 21, onde uma fatia segmentada inicial é dividida em medula espinhal e não medula espinhal após segmentação de candidatos.

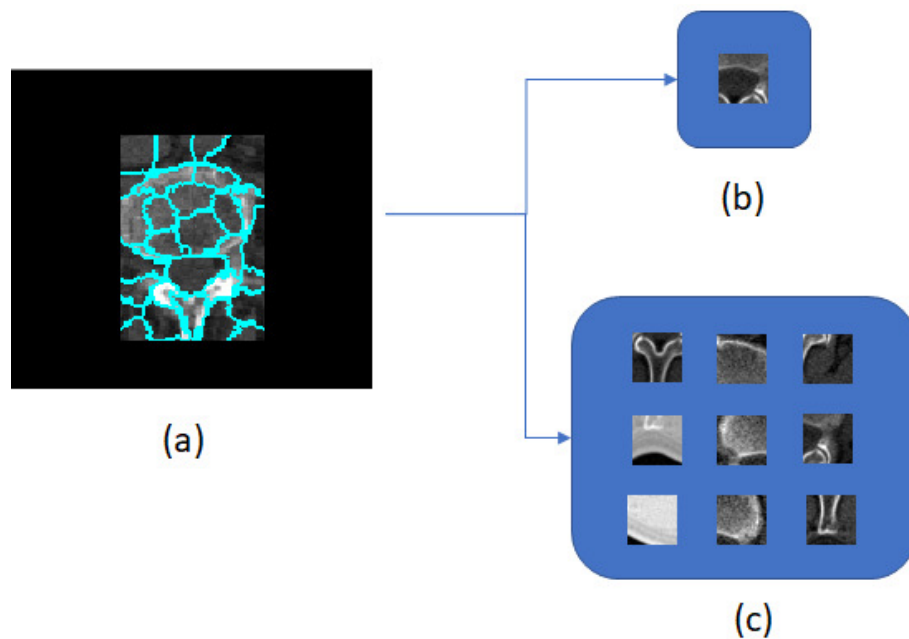


Figura 21 – Geração de imagens de entrada: (a) segmentação de candidatos de uma fatia por IMSLIC; b) Amostra de *bounding boxes* da medula espinhal; c) Amostra de *bounding boxes* não medula espinhal.

4.2.3.2 Treino e teste da CNN

Para o treinamento e o teste, as imagens de entrada precisam ser divididas em dois conjuntos de dados: o primeiro composto por imagens de entrada que representam

superpixels da medula espinhal e o segundo composto por imagens de entrada que representam *superpixels* e não medula espinhal.

Para definir quais *superpixels* são regiões da medula espinhal, são usadas marcações de especialistas. Um *superpixel* é considerado uma região da medula espinhal se tiver pelo menos 20% de seus pixels encontrados em qualquer marcação da medula espinhal, esta proporção foi a que produziu os melhores resultados. A Figura 22 ilustra como esse procedimento é feito.

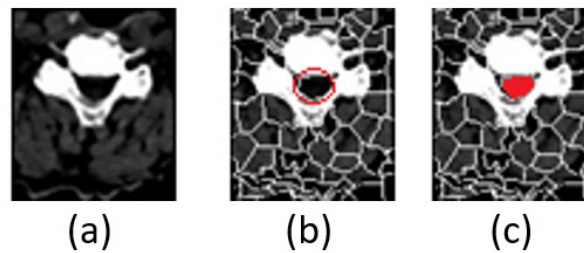


Figura 22 – Treino e teste da CNN: (a) representa uma fatia sem marcação; (b) representa a mesma fatia com marcação da medula espinhal e os *superpixels* gerados pela aplicação do IMSLIC; (c) *superpixels* tendo pelo menos 20% de seus pixels presentes em qualquer uma das marcações de uma medula espinhal de (b) são pintados na cor vermelha.

Na Figura 22, existem *superpixels* que possuem pixels presentes em alguma marcação, mas não são considerados região da medula espinhal. Assim como mostrado por Diniz et al. (2018), observou-se que os *superpixels* com menos de 20% de seus pixels presentes em alguma marcação possuem mais características da classe negativa, assim, considerou-se este critério para definir este *superpixels* como não medula espinhal. Uma vez que se sabe quais *superpixels* são regiões da medula espinhal, todos os outros *superpixels* da TC são considerados regiões de não medula espinhal.

Depois de gerar todas as imagens de entrada de todos os volumes, agora é necessário separar a base de imagens de treinamento e testar as bases de imagens para a etapa de treinamento e teste. Para cada uma dessas bases de imagens, os volumes foram selecionados aleatoriamente e as imagens de entrada de seus *superpixels* foram geradas.

Sabendo da efetividade dessas redes em vários domínios, utilizou-se para treinar e classificar os *superpixels* da medula espinhal em TC de planejamento. Para isso, será necessário definir a arquitetura da CNN que será utilizada neste método.

De um modo geral, CNN consiste basicamente em cinco camadas: (a) camadas convolucionais; (b) camadas de agrupamento; (c) camadas de ativação; (d) camadas de

abandono; e (e) camadas totalmente conectadas. A rede extrai recursos implícitos de padrões visuais apresentados pela entrada e classifica os padrões a partir dos atributos extraídos. A saída de cada camada serve como entrada para a próxima (LECUN et al., 1998b; LECUN; BENGIO; HINTON, 2015b).

Falando brevemente de cada camada, as camadas convolucionais têm filtros treináveis que são aplicados em toda a imagem de entrada. Camadas de agrupamento são camadas de redução de resolução não lineares. Neste trabalho, foi adotada a função *max-pooling* que gera valores máximos em uma vizinhança de mapas de características. As camadas de ativação aplicam ativações não lineares nos neurônios de entrada. Aqui, foi adotada a função de ativação da *Rectified Linear Units* (ReLU), uma vez que, de acordo com Krizhevsky, Sutskever e Hinton (2012b), fornece treinamento várias vezes mais rápido que outras funções de ativação.

As últimas camadas da rede são camadas totalmente conectadas. O número de neurônios de entrada para as camadas totalmente conectadas é definido pelo número de pixels resultantes da camada anterior a ele.

O treinamento supervisionado é realizado usando uma forma de gradiente descendente estocástico (SGD) para minimizar a discrepância entre a saída desejada e a saída atual da rede, com base na função de perda de entropia cruzada (LECUN et al., 1998b). Os pesos foram inicializados usando uma inicialização normalizada e atualizados com o algoritmo padrão de retropropagação (LECUN et al., 2010).

Diante disso, a arquitetura utilizada neste trabalho é a de LeCun et al. (1998a), e funciona usando o seguinte esquema:

1. imagem de entrada;
2. camada de convolução com 20 filtros de tamanho 5×5 com *stride* 1;
3. camada de agrupamento com função de max-pooling de tamanho 2×2 ;
4. camada de convolução com 50 filtros de tamanho 5×5 com *stride* 1;
5. camada de agrupamento com função de max-pooling de tamanho 2×2 ;
6. camada totalmente conectada composta de uma camada de entrada com o número de pixels das imagens anteriores a essa camada (500), seguida por uma camada de ativação com funções ReLU e uma camada de saída com uma função softmax, dando a probabilidade de a entrada pertencer a qualquer classe (medula espinhal ou não medula espinhal).

A Figura 23 ilustra a arquitetura usada em nosso método.

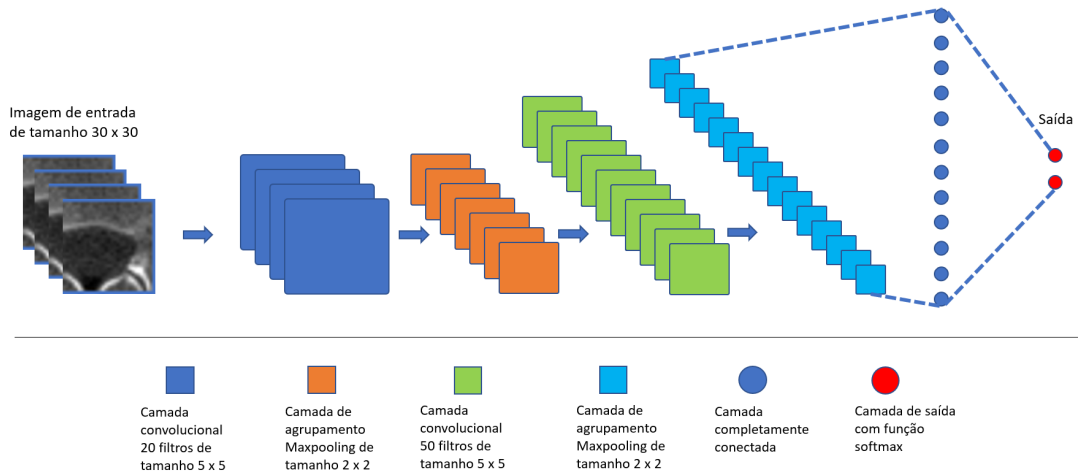


Figura 23 – Ilustração da arquitetura adotada.

Foram utilizadas 100 épocas como critério de parada, valor da taxa de aprendizado de 0,001 e *batch* de entrada de 64 imagens.

Após o treinamento desse modelo, ele foi testado em outros pacientes e as métricas de validação foram calculadas: Dice (LEENER; COHEN-ADAD; KADOURY, 2015), sensibilidade (SEN), especificidade (SPE), e acurácia (ACC), para medir a robustez do método. Os resultados deste método de segmentação de medula espinhal são apresentados na seção que segue.

4.2.4 Resultados

Esta seção contém os resultados obtidos aplicando o método apresentado na Seção 4.2. Além disso, será mostrado e discutido o resultado de cada etapa.

4.2.4.1 Ambiente de treino/teste

No desenvolvimento do método, foram utilizados dois computadores. Para gerar as imagens de entrada para CNN (a etapa de *template matching* adaptativo - Seção 4.2.1 e IMSLIC - Seção 4.2.2), utilizou-se um computador com Intel Core i5-7300HQ 2.50GHz, 8 GB de memória DDR3, embora a maioria das operações não tenha sido otimizada para o processamento em vários núcleos. A geração de patches (*bounding boxes* dos *superpixels*) por volume dura em torno de 1 minuto. Para construir o modelo, utilizou-se Keras (CHOLLET

et al., 2018) com tensorflow-gpu como back-end, em uma máquina com duas placas gráficas NVIDIA Titan X para possibilitar resultados mais rápidos e mais experimentos para construir a melhor arquitetura. O tempo de treinamento era de aproximadamente 35 segundos por época.

Para testar o modelo alcançado pela CNN, utilizou-se o primeiro computador: Intel Core i5-7300HQ 2.50GHz, 8 GB de memória DDR3. Ainda, a maioria das operações não foram otimizadas para o processamento em vários núcleos.

4.2.4.2 Aquisição de imagens

Como apresentado na Seção 4.1, a base de imagens usada para elaborar e testar o método foi a LCTSC, disponível no *The Cancer Imaging Archive* (YANG et al., 2018).

Esta base de imagens é composta por 36 pacientes de três institutos diferentes, correspondendo a 12 pacientes por instituto. Para elaboração deste método, dividiu-se este conjunto em treinamento e teste. Para o treinamento, utilizou-se 10 pacientes de cada instituto, gerando um total de 30 TC para treinamento. Para o teste, o restante dos volumes foi utilizado, com um total de 6 TC. Essa divisão foi feita de forma totalmente aleatória, apenas garantiu-se que o teste houvesse 2 pacientes de cada instituto, este critério foi adotado para que os diferentes tipos de aquisições fossem testados.

Durante o treinamento, há a necessidade de um grupo maior de imagens, de forma a conter o maior número de exemplos possíveis para o aprendizado da rede. Assim, foram reservados um maior número de imagens para o treinamento (30 exames). No teste, apesar de ter 6 exames, se for considerado o fato de que cada exame tem cerca de 200 fatias e que é garantido que 2 exames de cada instituto fossem colocados para essa base de imagens.

4.2.4.3 Segmentação Inicial

Nesta etapa, os dois conjuntos de dados (treinamento e teste) passaram pelo algoritmo de *template matching* adaptativo. Em todos os volumes dos dois conjuntos de imagens, não houve perda da medula espinhal nas fatias em que existiu.

Vale a pena notar que algumas tomografias incluíram todo o corpo (incluindo a cabeça), onde não há medula espinhal nem marcação, o *template matching* adaptativo considerou a região mais semelhante ao *template* da fatia anterior como medula espinhal.

No entanto, nas próximas etapas, todas essas regiões serão consideradas como não medula espinhal, já que não há marcação para validar. Esta é destacada como uma desvantagem do *template matching*, uma vez que não haja marcação, considera-se que aquela fatia não tenha medula espinhal, todavia, a segmentação inicial pelo *template matching* sempre achará uma subimagem em nossa abordagem. Um exemplo desses casos pode ser visto na Figura 24, onde (a) não tinha marcação do especialista e mesmo assim a segmentação inicial resultou em uma subimagem (b). Porém, em casos corretos, onde há medula espinhal como em (c), é possível encontrar corretamente a subimagem (d).

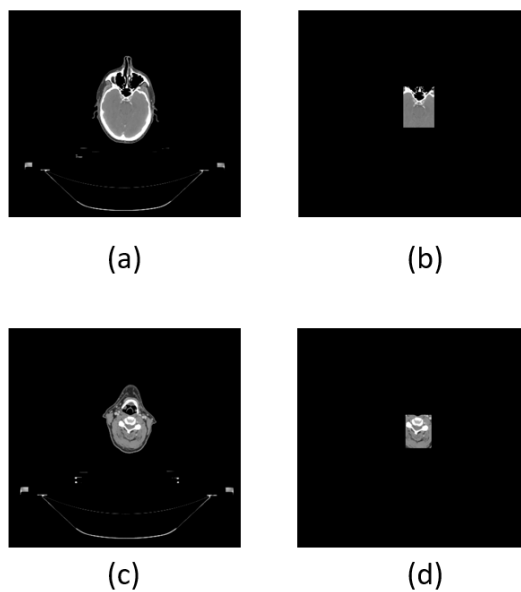


Figura 24 – Resultado do passo de segmentação inicial: (a) fatia onde não há medula espinhal (cabeça); (b) resultado da segmentação inicial em (a); (c) fatia onde há medula espinhal; (d) resultado da segmentação inicial em (c).

4.2.4.4 Segmentação de candidatos à medula espinhal

Nesta etapa, as estruturas segmentadas pela etapa de segmentação inicial passam pelo algoritmo de *superpixel* IMSLIC. O algoritmo é executado em todos os conjuntos de dados (treinamento e teste). Após uma série de testes para encontrar o melhor tamanho de *superpixel* inicial e o fator de compacidade, os que apresentaram maior eficácia foram os valores de 20 e 135, respectivamente.

Assim, 56.698 regiões foram agrupadas, onde 4.647 regiões foram da medula espinhal e 52.051 regiões de não medula espinhal. A Tabela 1 mostra a relação dos volumes de treinamento e teste com o número de *superpixels*.

Tabela 1 – Relação entre conjuntos de dados de treinamento e teste.

Imagens	Pacientes	Região de medula espinhal	Região de não medula espinhal
Treino	30	3.737	42.567
Teste	6	910	9.484
Total	36	4.647	52.051

Ressalta-se que foram utilizadas as marcações fornecidas pela base de imagens para saber quais *superpixels* eram medula espinhal e não medula espinhal, como explicado na Seção 4.2.2. Com essas regiões conhecidas, o próximo passo é criar um modelo e classificar essas regiões.

4.2.4.5 Classificação de candidatos à medula espinhal

Depois de ter nossas regiões de *superpixels* separadas entre treinamento e teste, o próximo passo é gerar um modelo a partir das imagens de treinamento da medula espinhal e não medula espinhal. Para fazer isso, utilizou-se uma CNN, como explicado na Seção 4.2.3. Agora, será descrito a fase de treinamento e a fase de teste para melhor compreensão.

4.2.4.5.1 Fase de treinamento

Sabendo que existem 30 volumes para treinamento, considera-se todos os *superpixels* da medula espinhal e de não medula espinhal para geração do modelo. Para validação, 10% dessa amostra foi escolhida, aleatoriamente. As imagens de entrada utilizadas foram descritas na Seção 4.1.

Imagens de treinamento e validação foram usadas para procurar a melhor configuração CNN. O processo adotado para esta arquitetura está descrito na Seção 4.2.3.2. Após o ajuste dos parâmetros, o próximo passo é aplicar este modelo gerado nas imagens de teste.

4.2.4.5.2 Fase de teste

Uma vez que as melhores configurações de CNN foram definidas, elas podem ser usadas para classificar as imagens de entrada de teste.

Para avaliar o método proposto, aplicou-se o modelo à base de imagens ainda desconhecido que foi separado de todos os dados de treinamento (conjunto de dados de teste). Como explicado na Seção 4.2.4.2, 2 pacientes de cada instituição foram escolhidos para validar a robustez do modelo.

A Tabela 2 mostra os resultados obtidos aplicando o modelo às imagens de entrada do conjunto de teste.

Tabela 2 – Métricas de validação da fase de teste.

Sensibilidade	89,23%
Especificidade	92,87%
Acurácia	92,55%
Dice	78,20%

De todas as imagens de entrada da base de imagens de teste, 910 eram imagens de regiões da medula espinhal e 9.484 de não medula espinhal, que foram geradas a partir de 6 exames. O modelo classificou corretamente 89,23% de regiões de medula espinhal e 92,87% de regiões de não medula espinhal.

Os resultados mostram a robustez do método proposto, mesmo em uma base de dados heterogênea, composta por exames de diferentes dispositivos de TC de três institutos diferentes. Os valores dos parâmetros utilizados são essenciais para alcançar esses resultados. Os resultados visuais da aplicação do método na base de teste serão apresentado nos estudos de casos (Seção 4.4).

4.2.4.6 Considerações

O método proposto mostrou-se robusto na tarefa de segmentação da medula espinhal. Através da avaliação e validação do método, com o uso de 36 pacientes de aproximadamente 200 fatias em cada volume, em uma base de imagens pública, as seguintes vantagens são destacadas:

1. A utilização da técnica de *template matching* adaptativo, foi eficaz na tarefa de reduzir a região de interesse dentro dos volumes;
2. A proposição do uso de um *template* adaptativo mostrou-se bastante eficaz, esta técnica foi testada em todos os 36 volumes de base de imagens, e em nenhum desses exames houve perda da região da medula espinhal;

3. O uso do *template matching* adaptativo garante que outras regiões que não fazem parte desta região de interesse e que envolvam a medula espinhal atrapalhem a CNN, uma vez que essas regiões já foram “descartadas” ao utilizar essa abordagem;
4. A técnica de (IMSLIC) proposta por Liu et al. (2018) foi utilizado pela primeira vez no contexto de imagens médicas, mostrando eficácia na tarefa de agrupar regiões de medula espinhal e não medular;
5. Em nosso trabalho, o IMSLIC se mostrou a eficiente em agrupar conjuntos de pixels correlacionados com base em semelhanças espaciais e de intensidade, o que tornou possível delimitar bem as bordas da medula espinhal;
6. A imagem de entrada para CNN é composta pelo *superpixel* e pixels em torno dele até um tamanho de *patch* de 30×30 . Esta etapa acrescenta um fator importante às imagens de entrada: sabe-se que as bordas da medula espinhal são envolvidas por estruturas com alta intensidade na TC, ao utilizar o IMSLIC na construção dos *patches* de entrada, consegue-se inserir essa característica nos patches da medula espinhal facilitando o treinamento pela CNN;
7. Ao usar a CNN para classificar os *patches* em medula espinhal, elimina-se dois passos importantes nos sistemas CAD, que exigem muito tempo: extração e seleção de características. Assim, as próprias camadas convolucionais são responsáveis por essa tarefa (LECUN; BENGIO; HINTON, 2015b).
8. Ao apresentar o *patch* de entrada construído com *template matching* adaptativo e IMSLIC com uma imagem bem definida da medula espinhal, a CNN consegue treinar e gerar um modelo robusto;
9. A combinação de todas as técnicas permitiu resultados promissores na segmentação da medula espinhal, alcançando uma sensibilidade de 89,23%, uma especificidade de 92,87%, uma acurácia de 92,55%, um Dice de 78,2%;
10. Por fim, a base de imagens utilizada é diversificada, composta por três institutos diferentes, com diferentes padrões de aquisição, com variação no número de fatias e, em alguns casos, com fatias correspondentes à imagem da cabeça. Mesmo com toda essa diversidade, o método mostrou-se robusto e promissor na tarefa de segmentação da medula espinhal.

Por outro lado, o método apresenta algumas limitações e alguns pontos negativos que podem ser destacadas como trabalhos futuros para refinar o resultado obtido.

1. O método apresentado foi treinado e validado em TCs diversificadas, com protocolo de aquisição distinto e despadronizado. Apesar disto, o método apresentou bons resultados em todas as etapas. Contudo, essa base apresenta apenas 36 pacientes, acredita-se que quando aplicado em um conjunto maior de imagens, algumas modificações sejam necessárias para generalizar ainda mais o trabalho.
2. A medula espinhal se estende por quase todas as fatias da CT. Em algumas fatias houve a perda da medula espinhal pela classificação da CNN, contudo, em fatias próximas a que houve perda a medula espinhal é encontrada (melhor descrito na Seção 4.4). Um algoritmo de interpolação entre as fatias poderia resolver esse problema, permitindo o preenchimento das fatias em que o método não encontrou a medula espinhal;
3. Observa-se que as métricas de validação para acurácia, sensibilidade e especificidade apresentaram valores significativos, no entanto, considera-se o Dice ainda baixo. Isto acontece ao calcular a métrica levando em conta a marcação do especialista e a gerada pelo IMSLIC, por utilizar valores correlacionados de pixels agrupados, às vezes a marcação do IMSLIC é maior ou menor que a do especialista. Uma maneira de superar essa desvantagem seria usar algoritmos baseados em level-sets (EVANS; SPRUCK, 1992), refinando a segmentação gerada por nosso método;
4. Por fim, sabe-se que as redes neurais convolucionais requerem um grande número de parâmetros como tamanho de filtros, quantidade de camadas, quantidade de filtros, entre outros, portanto, um estudo de técnicas para automatizar o valor desses parâmetros é dado como uma sugestão.

Todas essas vantagens agregam valor a este método. A combinação das três técnicas permitiu resultados relevantes em comparação com a literatura (Seção 5). Ainda, os estudos que realizam essa tarefa são escassos, principalmente em tomografia computadorizada, onde é mais complexa a localização da medula espinhal. Assim, mesmo em uma base de imagens tão diversificada e complexa, e com algumas limitações demonstradas, o método foi capaz de obter métricas de validação relevantes, logo, acredita-se que o método proposto é inovador, robusto e promissor.

4.3 Segmentação de medula espinhal utilizando *template matching* adaptativo e abordagens U-Net

Para o segundo método de segmentação da medula espinhal em TC de planejamento, é proposto um método dividido em três etapas. Primeira aquisição de imagem, que se refere a base de dados descrita no Seção 4.1. Segundo, segmentação inicial, onde um subvolume é gerado por paciente contendo apenas a região da medula e os objetos circundantes. Finalmente, a etapa de segmentação, a medula espinhal será segmentada usando três abordagens da U-Net e as métricas de validação são calculadas para validar a robustez do método proposto. Estas etapas são descritas na Figura 25.

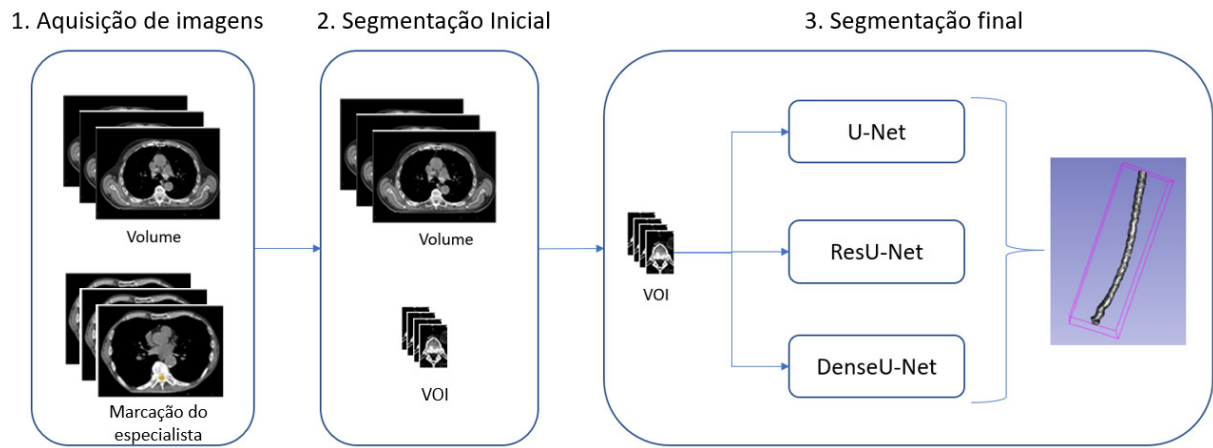


Figura 25 – Fluxograma do segundo método do medula espinhal.

4.3.1 Segmentação inicial

Por ser um volume tridimensional muito grande, com várias fatias e porque a medula espinhal compreende uma região muito pequena do volume, a etapa inicial de segmentação tem a tarefa de gerar um subvolume de interesse no qual o volume original é reduzido para apenas parte que compreende a medula espinhal e suas estruturas circunscritas. Assim, além de definir um VOI, também diminuimos o custo computacional de analisar todo o volume e todas as estruturas por fatia.

Para a segmentação inicial, foi utilizada a mesma técnica proposta na Seção 4.2.1 e publicada em Diniz et al. (2019), que foi melhorada e será explicada nas subseções que seguem.

4.3.1.1 *Template matching* adaptativo

Como mostrado na Seção 4.2.1, para esta técnica, primeiramente é necessária a definição do *template* padrão. Este é definido igualmente é descrito na Seção 4.2.1.1. Em seguida, para cada volume é definido o *template* inicial, este passo é igual ao apresentado na Seção 4.2.1.2.

Por fim, na adaptação do *template* inicial a cada fatia do volume de um paciente, são propostas melhorias a técnica apresentada na Seção 4.2.1.3. Como apresentado na Seção 4.2.1.3, executar o *template* em toda a fatia de todos os volumes da base se torna custoso. Além do fato de não garantir que em uma nova base de dados, com algum artefato na estrutura das fatias (fora do corpo), possa fazer com que o modelo se perca ao encontrar a melhor correspondência na região da medula espinhal. Além disso, pesquisar a fatia inteira para essa região se torna caro, por isso é proposto aprimorar essa técnica, melhorando a complexidade e a eficiência.

Para resolver esse problema, é proposto executar o *template matching* apenas na região correspondente ao dobro da região de *matching* nas fatias seguintes (*bounding box* verdes na Figura 26) e não na fatia inteira (como proposto na Seção 4.2.1.3), dessa maneira, tornou-se o algoritmo mais rápido e também artefatos fora do corpo do paciente não atrapalham o algoritmo. A Figura 18 ilustra o processo de adaptação do modelo.

O resultado da aplicação da etapa inicial de segmentação é um subvolume de interesse (VOI) por paciente que contém a medula espinhal e as estruturas circundantes. Este volume será usado para treinar três U-Net diferentes para avaliar o desempenho dessas redes na tarefa de segmentação da medula espinhal. Em seguida, avalia-se os resultados em uma fase de teste.

4.3.2 Segmentação final

A etapa final de segmentação resultará em um VOI da região considerada como medula espinhal a partir do treinamento de um modelo baseado em redes neurais convolucionais. Primeiro, é necessário entender como as imagens de entrada são geradas e, em seguida, a arquitetura das redes usadas para geração e classificação de modelos de VOIs. A seguir, será explicado como as imagens de entrada nas redes U-Net serão geradas.

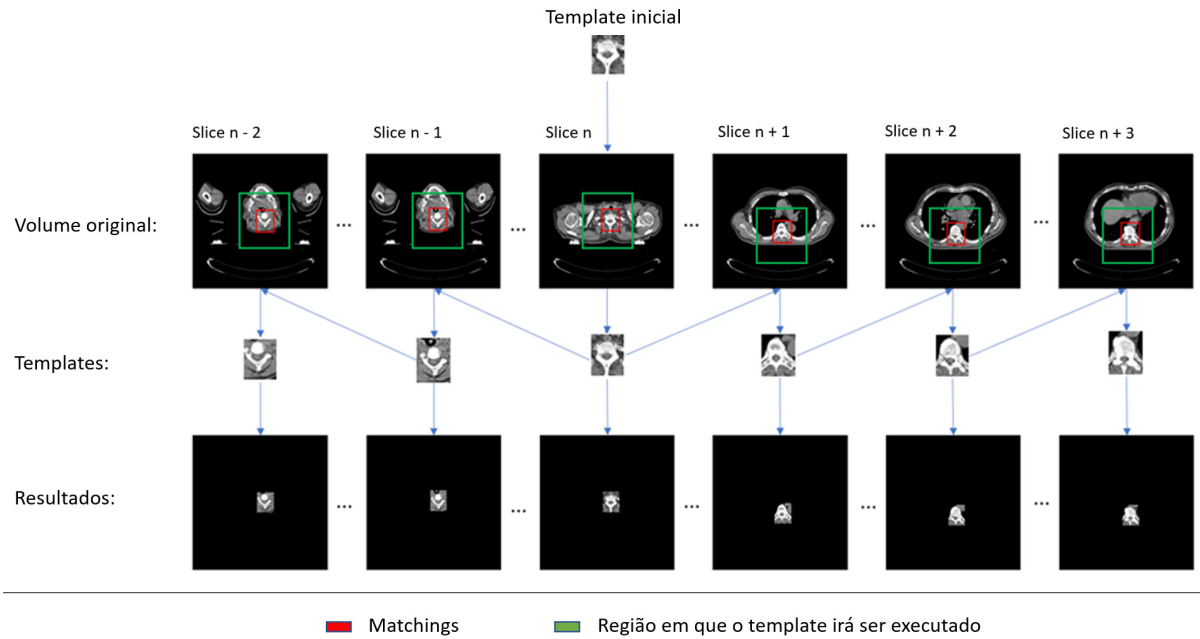


Figura 26 – Adaptando o *template*: as linhas em vermelho mostram onde ocorreu o *matching*, as linhas em verde mostram onde o *template* será executado.

4.3.2.1 Gerando imagens de entrada

Como já mencionado, ao final da segmentação inicial, temos um VOI de cada paciente. Esse VOI consiste na medula espinhal e nas estruturas ao seu redor. Além disso, o que é feito com o paciente, é feito com a marcação do especialista, ou seja, é gerado um VOI com a mesma dimensão do paciente, mas com a marcação do especialista. A Figura 27 ilustra esses VOIs.

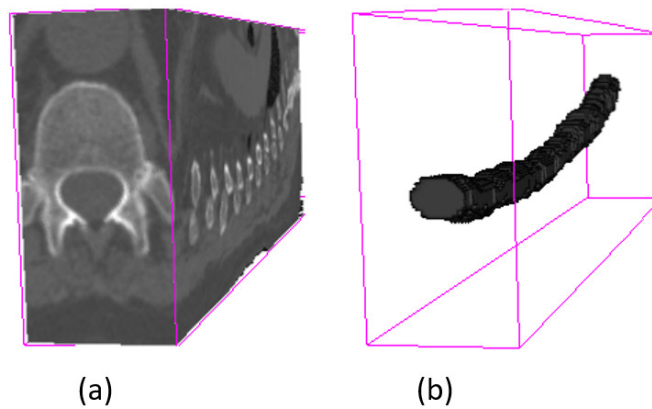


Figura 27 – Imagens de entrada: (a) VOI do paciente com medula espinhal e estruturas ao seu redor; e (b) VOI da marcação especializada da medula espinhal.

Para a segmentação da medula espinhal, foi utilizada uma abordagem 2D, ou seja, o VOI de cada paciente foi passado fatia por fatia na rede. No entanto, ao final da classificação das redes, é construído um volume 3D das fatias classificadas corretamente pela rede.

Optou-se por usar as fatias em vez do volume, porque se for considerado o número de volumes existentes, existem apenas 36, o que significaria que a rede teria pouca diversidade de amostra na fase de treinamento. À medida que as fatias são usadas, a diversidade de conjuntos de dados de treinamento/validação e teste é aumentado, uma vez que, cada paciente possui aproximadamente 200 fatias. Ainda, dado o tamanho dos volumes, aumenta o grau de complexidade para execução do treinamento da rede com os dados 3D. Assim, mais uma vez a estratégia adotada 2D mostrou-se válida.

O próximo passo é o treinamento da U-Net para criar um modelo robusto capaz de classificar novos pacientes. Assim, é necessário entender alguns conceitos da rede U-Net e algumas variações de arquitetura. As próximas seções apresentam o conceito de U-Net e a arquitetura de três redes: U-Net (RONNEBERGER; FISCHER; BROX, 2015), ResU-Net (KOLARŽÍK et al., 2019) e DenseU-Net (KOLARŽÍK et al., 2019).

4.3.2.2 Abordagens U-Net

Como explicado na Seção 3.2.5.2, a U-Net é uma rede neural convolucional, portanto, usa os conceitos e as camadas de uma CNN, como camadas convolucionais, camadas de subamostragem, camadas de ativação e camadas de *dropout*.

Em nosso trabalho, é proposto o uso de três variações da arquitetura U-Net para compor o segundo método de segmentação da medula espinhal como um órgão de risco. A seguir, define-se as três arquiteturas.

4.3.2.3 U-Net

A arquitetura U-Net foi a mesma arquitetura proposta por Ronneberger, Fischer e Brox (2015), que difere apenas no fato de que a fatia de entrada tem um tamanho diferente daquele usado na arquitetura proposta por eles. O que é passado para a rede são fatias do VOI do paciente junto com a fatia da marcação do especialista. A saída da rede será uma

fatia classificada como medula espinhal. A rede é treinada com a função de perda *Diceloss* representada pela Equação 19.

$$diceloss = 1 - \frac{2TP}{2TP + FP + FN} \quad (19)$$

Optamos por usar a *Diceloss*, pois pretende-se otimizar os resultados da segmentação com base na marcação do especialista. Assim, o uso desta função como perda ajuda a rede a melhorar esta métrica durante o treinamento.

4.3.2.4 ResU-Net e DenseU-Net

Duas novas técnicas foram propostas para essas redes que produziram melhores resultados em vários domínios de imagem, sendo elas: blocos residuais (HE et al., 2016) e camadas densas (TONG et al., 2017).

Com a ideia de blocos residuais e camadas densas, Kolařík et al. (2019) propuseram uma nova arquitetura U-Net, que usou essas duas técnicas. De acordo com Kolařík et al. (2019), os resultados alcançados usando essas técnicas para melhorar a U-Net produziram resultados muito melhores do que usar a U-Net. A arquitetura proposta por Kolařík et al. (2019), pode ser vista na Figura 28.

Para testar a abordagem U-Net com blocos residuais, utilizou-se a mesma rede proposta por Kolařík et al. (2019) na Figura 28, excluindo conexões densas. Essa abordagem é chamada de ResU-Net.

Para testar a abordagem conjunta de blocos residuais e camadas densas, utilizou-se a mesma rede proposta e apresentada na Figura 28. Essa abordagem intitulou-se de DenseU-Net.

Ainda, todas as redes (U-Net, o ResU-Net e o DenseU-Net) são treinados com a função de perda *Diceloss*, representada pela Equação 19.

Com as arquiteturas definidas, o próximo passo é a definição da fase de treinamento e teste. A próxima seção explica como esse processo foi realizado.

4.3.2.5 Treino e teste das abordagens U-Net

Nesta seção, a base de dados é dividida em dois grupos: o primeiro com as fatias de entrada de treinamento e o segundo com as fatias de entrada de teste. A base de

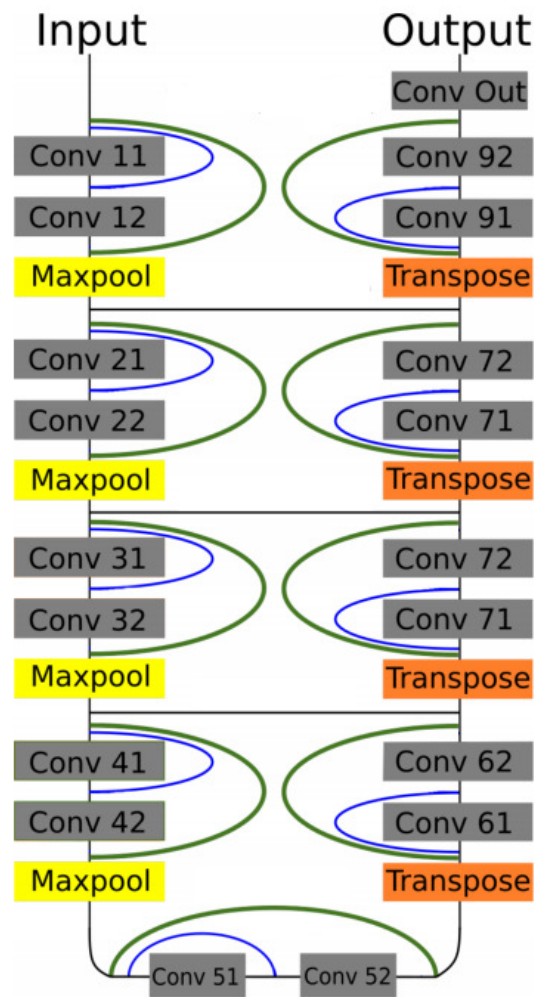


Figura 28 – Abordagens U-Net: Representação da arquitetura da U-Net em que os blocos residuais são na cor verde e as ligações densas em azul. Figura de Kolařík et al. (2019).

treinamento é composta por fatias de pacientes e fatias correspondentes de marcações especializadas.

Na fase de treinamento, uma parte dos dados é separada para validação; é a partir da validação que são feitas simulações para encontrar os melhores parâmetros da rede. No final do treinamento e validação, o modelo é gerado. Este modelo é então aplicado nas imagens separadas para teste.

A fase de teste consiste em usar o modelo para classificar novas imagens, desconhecidas para o modelo. Para validar os resultados do teste, a imagem de marcação do especialista é obtida e as métricas de validação são calculadas, assim, é avaliado se o modelo foi bom o suficiente na segmentação da medula espinhal como um OAR. As métricas de validação utilizadas neste trabalho foram: Dice, sensibilidade (SEN), especificidade (SPE),

acurácia (ACC). Os resultados para o segundo método de segmentação de medula espinhal são apresentados na seção que segue.

4.3.3 Resultados

Os resultados alcançados na tarefa de segmentação da medula espinhal, usando o segundo método proposto no planejamento da radioterapia, serão descritos nesta seção. Também será destacado o ambiente de treinamento e discutiremos alguns resultados.

4.3.3.1 Ambiente de treino/teste

O ambiente de treinamento para executar a etapa inicial de segmentação foi um computador com Intel Core i5-7300HQ a 2,50GHz, 8 GB de memória DDR3. Ao melhorar o *template matching* adaptativo (Seção 4.3.1), que durava aproximadamente 1 minuto por volume, neste segundo método, gasta-se menos de 20 segundos para gerar VOI por paciente que serão apresentados às U-Nets. Na construção do modelo de rede e das arquiteturas U-Nets, foi usada uma máquina com duas placas gráficas NVIDIA Titan X, na qual foi utilizado a biblioteca Keras com o TensorFlow-GPU como back-end.

4.3.3.2 Aquisição de imagens

Como já mencionado na Seção 4.1, o banco de dados é composto por três institutos, com 12 pacientes por instituto, resultando em 36 volumes de pacientes. Para treinar e validar no segundo método, foi dividido a base de imagens em 2 conjuntos de dados. O primeiro para treinamento, composto por 30 pacientes (10 por instituto) e 6 por teste (2 por instituto). Essa divisão foi feita aleatoriamente (igualmente o primeiro método), garantindo apenas a proporção do instituto nos dois conjuntos de dados.

Além disso, após a geração do modelo usando os 30 volumes e o teste em 6 volumes, as métricas de validação são calculadas.

4.3.3.3 Segmentação inicial

Na etapa de segmentação inicial, descrita na Seção 4.3.1, os dois conjuntos de dados passaram para o algoritmo proposto usando a correspondência de modelos. Como no método proposto descrito na Seção 4.2, não houve perda da medula espinhal em nenhuma das fatias de nenhum paciente. Para validar a etapa aleatória de selecionar qualquer volume e fatia, foram feitas várias execuções (exclusivas) da seleção *template*, e a medula espinhal não foi perdida.

O mesmo problema destacado na Seção 4.2, onde alguns pacientes não foram marcados em todas as fatias e, em alguns casos, houve fatia da cabeça, o que também foi observado nos testes. No entanto, mesmo com esse problema, o método mostrou-se eficiente. A Figura 29 demonstra o resultado inicial da segmentação em um paciente do banco de dados¹.

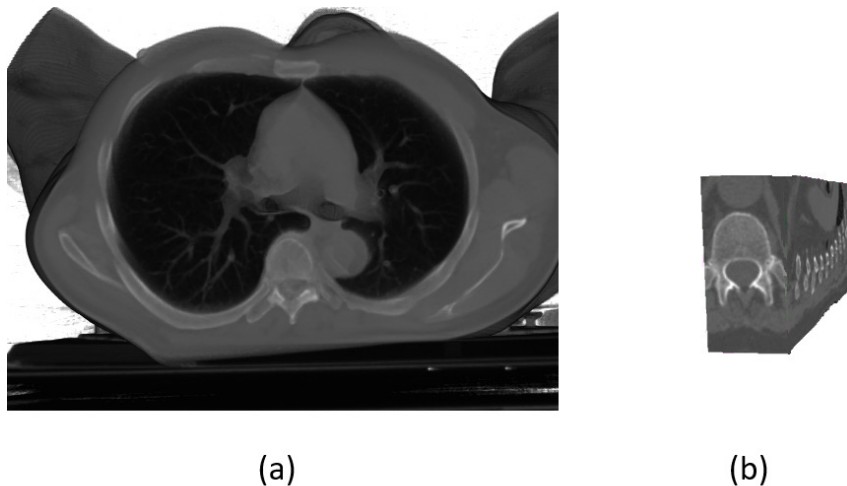


Figura 29 – Segmentação inicial: resultado da aplicação do *template matching* adaptativo.

4.3.3.4 Segmentação final

Nesta etapa, o resultado encontrado pela segmentação inicial será apresentado a três arquiteturas U-Net distintas. Primeiro, as fatias dos pacientes são passadas para as redes. Como destacado na Seção 4.3.2, essa estratégia foi usada para a rede ter mais imagens na amostra para aprender os padrões em cada fatia. No entanto, na fase de teste,

¹ Esta visualização foi gerada pelo software 3D Slicer, a inclinação nas imagens têm o intuito de mostrar a resultado do *template matching* adaptativo ao longo do volume.

após a classificação fatia por fatia, é possível criar uma marcação de volume 3D predita pelo modelo gerado.

As imagens usadas para a entrada de rede eram de 128×64 , com base na aleatoriedade da criação dos *templates*, esse era o tamanho que não excedia a imagem gerada. Após várias seções de treinamento e configuração de parâmetros, os parâmetros usados no treinamento das redes foram: tamanho do *batch* igual a 1, Otimizador Adadelta (ZEILER, 2012) com taxa de aprendizado inicial igual a 0,0001 e 10% de volumes do treinamento para validação.

Após o treinamento das três redes com os 30 volumes iniciais e a geração do modelo, elas foram aplicadas aos 6 volumes de teste (2 pacientes por instituto) e foram calculadas métricas de validação para avaliar a robustez do segundo método proposto. Primeiro, apresenta-se as métricas usando U-Nets com a função de perda de entropia cruzada (GODOY, 2019), que é comumente usada em abordagens usando redes neurais convolucionais. Estes resultados estão descritos na Tabela 3. O tempo aproximado de treino por época (em segundos) e o tempo aproximado de teste (para um paciente) também são expressos na tabela.

Tabela 3 – Resultado da etapa de segmentação final usando entropia cruzada como função de perda.

Rede	Dice(%)	Tempo de Treino (seg)	Tempo de Teste (seg)	SEN(%)	SPE(%)	ACC(%)
U-Net	79,18	320	20	83,62	99,47	99,18
ResU-Net	77,36	360	20	80,47	99,48	99,12
DenseU-Net	79,54	420	20	84,90	99,46	99,19

Ao analisar a Tabela 3, pode-se identificar que as três arquiteturas foram promissoras na tarefa de segmentação da medula espinhal. Ao avaliar a métrica de sensibilidade, fica claro que o método foi capaz de ajustar com eficiência os valores positivos da classe, alcançando a melhor sensibilidade usando o DenseU-Net com 84,90%. Analisando os valores de especificidade, percebe-se que havia poucos falsos positivos, desde que a métrica alcança 99%. Visualizando a acurácia, nota-se a boa relação entre os valores verdadeiro positivo, falso positivo, verdadeiro negativo e falso negativo.

Vale destacar que há um desbalanceamento entre a classe positiva e a classes negativa (i.e., o número de *voxels* que são medula espinhal e não medula espinhal, respectivamente), onde a classe negativa é superior. Assim, mais amostras de não medula espinhal é apresentado as U-Net, levando um melhor aprendizado desta classe. Ainda, na extração das métricas, por mais que alcance uma acurácia de 99%, deve-se observar

prioritariamente a métrica Dice como mais importante, uma vez que ela reflete o balanceamento entre verdadeiros positivos e falsos positivos.

Embora o Dice seja superior a muitos trabalhos mostrados na literatura (vide Capítulo 2), ela ainda está em desvantagem porque, em muitos casos, não há marcação na medula espinhal ao longo do volume. Ainda assim, são considerados resultados bastante promissores, com o melhor resultado encontrado no DenseU-Net, atingindo 79,54% nessa métrica.

Devido à deficiência na métrica de Dice, propôs-se o uso da função de *Diceloss* no treinamento em rede. De acordo com Sudre et al. (2017), essa métrica é melhor especialmente em classes desequilibradas, como é o nosso caso, uma vez que a medula espinhal compreende a menor parte das fatias. Resumidamente, essa métrica tende a ser otimizada nas interações da rede; portanto, são esperados melhores resultados em comparação à marcação de especialistas. Os resultados usando a função de *Diceloss* estão listados na Tabela 4. O tempo aproximado de treino por época (em segundos) e o tempo aproximado de teste (para um paciente) também são expressos na mesma tabela.

Tabela 4 – Resultado da etapa de segmentação final usando a função de *Diceloss*.

Rede	Dice(%)	Tempo de Treino (seg)	Tempo de Teste (seg)	SEN(%)	SPE(%)	ACC(%)
U-Net	80,83	320	20	92,64	99,30	99,18
ResU-Net	81,69	360	20	91,52	99,38	99,24
DenseU-Net	79,20	420	20	93,97	99,18	99,08

Analisando a Tabela 4, há uma melhora significativa nos resultados quando utilizou-se a função de *Diceloss*. Isso ocorre, pois durante o treinamento a função de perda é otimizada, e como busca-se resultados em relação ao Dice, que é a métrica principal em definir o quanto a marcação do especialista é similar ao método de segmentação, os resultados ao utilizar esta função de perda são superiores a Tabela 3.

Por sua vez, ao usar a função de *Diceloss*, a métrica de sensibilidade anteriormente usava a função de perda de entropia cruzada apresentava melhores resultados na DenseU-Net que tinha um resultado de 84,90%, agora os valores nas três arquiteturas são superiores a 91%, isso mostra que casos positivos estão sendo encontrados. Ao analisar a especificidade, é possível que a métrica permanece próxima de 100%. As métricas de acurácia também mantiveram métricas significativas.

Ainda, com uma melhoria de mais de 6% na sensibilidade (maior acerto da classe positiva, isto é, medula espinhal), a métrica Dice apresenta valores mais altos nas três

arquiteturas, sendo que usando blocos residuais (ResU-Net) apresenta o valor mais alto com o resultado de 81.69%. Conforme descrito em na Seção 4.2 e em Diniz et al. (2019), muitas fatias de pacientes não têm marcações, o que diminui esse tipo de métrica.

Os resultados visuais da aplicação do método na base de teste serão apresentado nos estudos de casos (Seção 4.4).

4.3.3.5 Considerações

Proteger a medula espinhal como um órgão de risco é uma tarefa fundamental no planejamento da radioterapia. A segmentação automática desse órgão auxilia o especialista na tarefa de proteção, uma vez que muita radiação nesse órgão pode causar danos irreversíveis aos pacientes. Assim, foi proposto um segundo método totalmente automático que utiliza abordagens U-Net para esta tarefa. O método visa compor o sistema CAD e foi testado em um banco de dados público para verificar sua eficácia. Assim, acredita-se que o método tenha várias vantagens, incluindo:

1. O método foi desenvolvido em uma base pública, o que possibilita a replicação e comparação;
2. Na segmentação inicial, foram feitos avanços na técnica proposta anteriormente na Seção 4.2 e em Diniz et al. (2019), tornando-a mais rápida e precisa na segmentação de uma região inicial;
3. Ao usar a segmentação inicial para limitar uma área de interesse a ser apresentada às arquiteturas de U-Nets, tornou-se o método mais robusto, pois não há necessidade de classificar todos os voxels no volume;
4. Apesar de ser uma dado 3D, o método é treinado em 2D, então resolveu-se o problema de ter poucas imagens para treinamento;
5. Usando abordagens da U-Net, pode-se suprimir várias etapas importantes no processamento e análise de imagens, como algoritmos de segmentação baseados em formas ou texturas, seleção e extração de características e classificação;
6. Um método robusto e comparativo, mostrando o resultado dos três principais tipos de abordagens de redes U-Nets encontrados na literatura;
7. O método ocupa um lugar de destaque na literatura, pois não utiliza nenhum conhecimento prévio, i.e., não necessitou de informação anatômica ou uso de atlas no

treinamento da rede, assim, o método torna-se automático e robusto na segmentação da medula espinhal;

8. O uso da função de *Diceloss* foi muito promissor, aumentando todas as métricas de validação e proporcionando melhor segmentação da medula espinhal;
9. A combinação das técnicas utilizadas na segmentação inicial e final é altamente precisa, com acurácia maior que 99%, sensibilidade maior que 91,5%, especificidade maior que 99,1%;
10. O método atinge um Dice de 81,69% em um banco de dados altamente diversificado, com problemas de marcação e aquisição provenientes de três institutos diferentes, provando que o método é muito promissor;
11. Além do uso da rede que já supera o estado da arte em vários domínios de imagem, propõe-se também o uso de duas técnicas de aprimoramento nessa rede: blocos residuais e camadas densas, tornando o método ainda mais robusto;

Mesmo com resultados promissores e provando ser eficiente na tarefa de segmentação da medula espinhal, o método ainda pode ser aprimorado, de modo que algumas limitações e pontos negativos possam ser exploradas no futuro:

1. Como destacado nos materiais (Seção 4.1), a base de imagens fornece apenas 36 pacientes, assim, acredita-se que quando testado em uma base maior, melhorias possam ser feitas nas configurações das etapas para aumentar a eficiência do método.
2. Embora muitos testes mostrem que o *template* padrão não interfere no resultado inicial da segmentação, um ponto a ser aprimorado no método é este, pois encontrar esse *template* automaticamente, excluiria o fator aleatório do método;
3. Embora todas as métricas tenham valores superiores a 90%, o Dice ainda é considerado baixo, uma sugestão de melhoria seria a redução de falsos positivos após a etapa de segmentação utilizando técnicas de processamento de imagens, como técnicas de reconstrução 3D baseado na curvatura da superfície gerada;
4. Finalmente, acredita-se que alterações nas arquiteturas ou o uso de informações de volume tridimensionais, em vez de fatias, podem melhorar ainda mais os resultados encontrados.

Todos esses pontos são de grande valor para o método proposto. Ao combinar as técnicas de segmentação inicial com a segmentação final, pode-se encontrar métricas

muito relevantes e proeminentes na literatura. Assim, o método mostrou-se robusto nessa tarefa, usando o ResU-Net, atingiu a métrica Dice de 81,69%, sensibilidade de 91,52%, especificidade de 99,38% e acurácia de 99,24%.

4.4 Estudo de caso dos métodos de medula espinhal

A fim de avaliar os dois métodos propostos para segmentação de medula espinhal, nesta seção são apresentados os estudos de caso. Basicamente, pretende-se avaliar o quão similar são as predições dos dois métodos comparados a marcação do especialista. Para isso, ao fim de todas as etapas de cada método, geram-se volumes resultantes. Para o segundo método, utilizou-se apenas a ResU-Net, uma vez que esta apresentava melhores resultados quantitativos. Como a base de imagens é composta por três institutos, serão apresentados os resultados por institutos.

4.4.1 Estudo de caso 1 - Instituto 1

Para a segmentação da medula espinhal foram propostos dois métodos. Sendo um baseado em uma técnica de *superpixel* e CNN e outro baseado em uma classificação pixel-a-pixel por uma ResU-Net. A Figura 30 apresenta o resultado dos dois métodos em volumes 3D juntamente com a marcação do especialista para o Instituto 1.

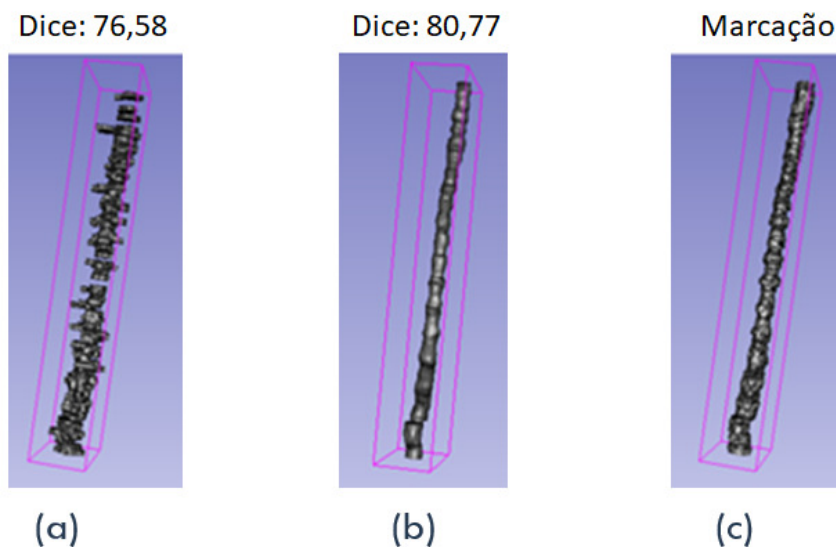


Figura 30 – Estudo de caso - Instituto 1: (a) volume predito pelo método IMSLIC + CNN, (b) predito pelo método ResU-Net e (c) marcação do especialista.

No resultado do primeiro método (Figura 30a), pode-se observar que o mesmo apresenta espaços vazios ao longo do volume (Dice de 76,58), isto pode ocorrer pela similaridade dos *superpixels* de medula espinhal e de não medula, acabando por confundir o modelo. Uma forma de contornar esse problema, pode ser a utilização de técnicas de interpolação com as fatias vizinhas marcadas pelo método. Além disso, o método apresenta diversos falsos positivos e bordas irregulares. Contudo, nota-se que há uma boa segmentação quando comparado a marcação do especialista.

Ao visualizar o segundo método (Figura 30b), observa-se que ele foi mais preciso na segmentação da medula espinhal (Dice de 80,77). Diferente do primeiro método, neste não há espaços vazios ao longo do volume, e ainda há uma maior regularidade entre as bordas. Isto se dá pelo fato que este método apresenta uma classificação pixel-a-pixel, mais precisa que o método que utiliza *superpixel*, visto que ao classificar todo um *superpixel* como medula espinhal, este pode compreender uma região muito maior que a marcação do especialista.

4.4.2 Estudo de caso 2 - Instituto 2

Análogo ao primeiro estudo de casos, neste é apresentado o resultado tridimensional dos dois métodos comparados a marcação do especialista. Estes resultados são apresentados na Figura 31.

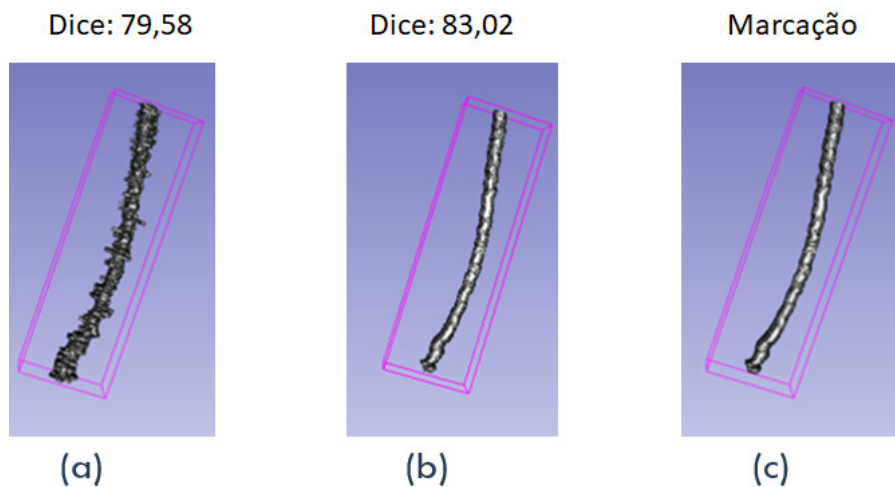


Figura 31 – Estudo de caso - Instituto 2: (a) volume predito pelo método IMSLIC + CNN, (b) predito pelo método ResU-Net e (c) marcação do especialista.

Mais uma vez o uso da ResU-Net mostrou superioridade comparado ao método de segmentação baseado em *superpixel* (Dice de 79,58 no primeiro método e Dice de 83,02 no segundo método). A classificação pixel-a-pixel mostra melhor regularidade das bordas e gera menos falsos positivos em relação aos *superpixel*. Ao utilizar o *superpixel*, o mesmo tende a utilizar as informações de textura para criar suas delimitações, por isso, há a possibilidade de exceder os limites ósseos do canal vertebral e, ao classificá-lo como medula espinhal, acaba gerando tais irregularidades nas bordas.

4.4.3 Estudo de caso 3 - Instituto 3

Por fim, são apresentados os resultados da aplicação dos dois métodos em um paciente da base de teste do Instituto 3. Os resultados dos dois métodos em uma visão tridimensional estão apresentados na Figura 32.

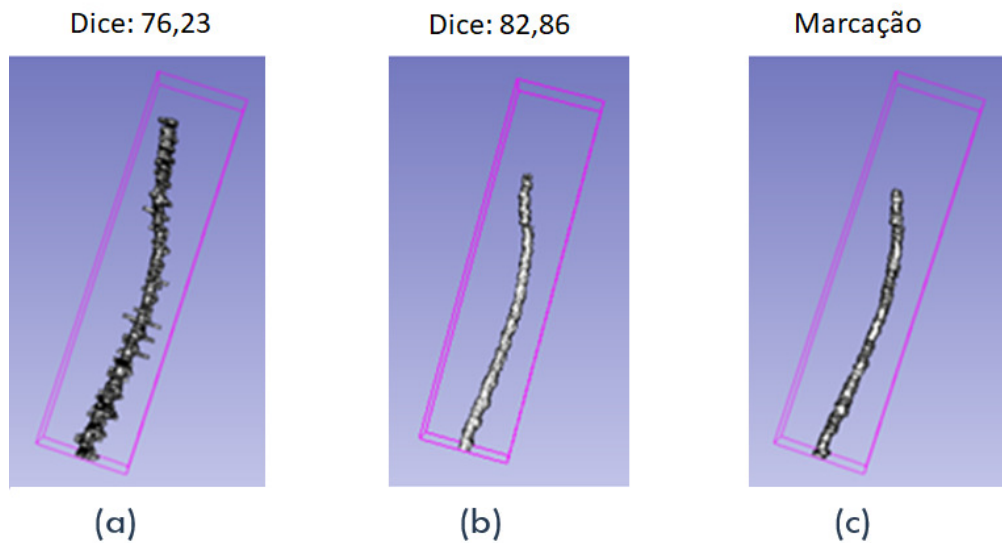


Figura 32 – Estudo de caso - Instituto 3: (a) volume predito pelo método IMSLIC + CNN, (b) predito pelo método ResU-Net e (c) marcação do especialista.

O método de segmentação de medula espinhal utilizando a ResU-Net mostra resultados superiores ao método que utiliza IMSLIC e CNN. Como citado anteriormente, existem algumas fatias que não possuem a marcação do especialista na medula espinhal. Isto impactou negativamente os dois métodos, todavia, o primeiro método sofreu mais com esse problema da base de dados. Ao observar a Figura 32(a), o método considerou uma grande parte das fatias que não havia marcação como medula espinhal.

Mesmo o segundo método considerando algumas fatias que não havia marcação como medula, ainda assim seus resultados foram mais satisfatórios. Observando a marcação do especialista, o método que utiliza a ResU-Net mostrou-se mais robusto ao segmentar a medula espinhal.

4.5 Comparação entre os resultados dos métodos de medula espinhal

Nesta seção, são discutidos e comparados os dois métodos para segmentação de medula espinhal apresentado neste trabalho. A Tabela 5 apresenta os melhores resultados dos dois métodos.

Tabela 5 – Comparação entre os resultados dos métodos de medula espinhal.

Rede	Dice(%)	SEN(%)	SPE(%)	ACC(%)
CNN	78,20	92,87	92,55	89,23
ResU-Net	81,69	91,52	99,38	99,24

Neste trabalho, foram desenvolvidos dois métodos para segmentação da medula espinhal. O primeiro método apresenta uma série de técnicas para a tarefa de segmentação de medula espinhal, consegue-se alcançar resultado satisfatório em todas as métricas. A etapa de segmentação inicial por uso do *template matching* é muito promissora, além de não haver perda da medula em nenhum exame (seja de teste ou treino), é possível reduzir o escopo para segmentação do órgão. Também, observou-se que o uso do *superpixel* é bastante robusto em encontrar os limites ósseos da medula espinhal, compreendidos no canal vertebral. O uso de técnicas de CNN para classificar esses *superpixel* entre medula espinhal e não medula espinhal também se mostra robusto, utilizando CNN consegue-se abstrair as etapas de extração e seleção de características. Observa-se na Tabela 5, que utilizando o primeiro método foi possível encontrar métricas relevantes, com Dice de 78,2%.

Todavia, ao analisar o fato que a classificação é em razão de *superpixel*, por mais que em muitos casos estejam bem delimitados os limites ósseos pelos *superpixel*, este acabam excedendo a marcação do especialista, apresentando um déficit nas métricas de validação. Logo, percebe-se que uma classificação pixel-a-pixel poderia tornar esse método mais robusto.

A classificação pixel-a-pixel mostra-se um importante avanço para o primeiro método, todavia, ao observar todo volume, a medula espinhal compreende apenas uma

região pequena de todo o volume. Utilizar uma FCN, como a U-Net, poderia gerar resultados ruins, como mostrado em Diniz et al. (2019), quando aplicada em todo o volume do paciente. Assim, nota-se, mais uma vez, a importância da diminuição do escopo do volume para uma VOI que compreendesse a estrutura da medula espinhal e órgãos circundantes. Desta forma, sugere-se no segundo método, uma segmentação inicial também baseada em *template matching*. Contudo, o *template matching* é suscetível a sua região de busca, tanto em custo computacional quanto em qualidade. Assim, uma melhoria foi feita nesta etapa, agora a região de busca do *template* só compreende ao dobro do melhor *matching*.

Ao gerar a segmentação inicial com essa nova abordagem de *template matching*, resta-se uma VOI menor, passível de ser classificada pelas abordagens U-Net. Assim, sugeriu-se o uso de três U-Net no segundo método. Os melhores resultados foram encontrados pela U-Net composta com blocos residuais (ResU-Net) apresentando resultados robustos, com um Dice de 81,69%. Ao olhar apenas para o Dice, há uma melhoria de mais de 2% se comparado com o método anterior, e ainda uma diminuição no número de etapas para a elaboração do método. Mais uma vez, destaca-se o balanceamento das amostras entre medula espinhal e não medula espinhal, onde uma melhoria de 2% significa grande avanço de resultado observando os estudos de casos apresentados.

Outro ponto que vale destacar, que a melhora das métricas de especificidade e acurácia. Observando a especificidade, com o método de CNN alcançou-se 92,55%, enquanto utilizando ResU-Net esta métrica sobe para 99,38%. Esses valores mostram que o segundo método apresenta menor número de falsos positivos e, conseqüentemente, uma segmentação mais precisa deste OAR.

Vale lembrar, que ambos os métodos propostos trazem uma série de avanços como já destacados nas seções anteriores, e que o uso deles pode ser fundamental em grandes centros médicos, onde a prática clínica é cansativa. Assim, auxiliando o especialista, esses métodos podem ser uma ferramenta de grande importância em tratamento que utilizam RT para combate ao câncer.

4.6 Comparação com os trabalhos relacionados a segmentação de medula espinhal

Existem poucos estudos que realizam a segmentação da medula espinhal em imagens de tomografia computadorizada de planejamento como destacado no Capítulo 2.

Archip et al. (2002) propuseram uma abordagem baseada em *knowledge-based* que consiste em um chamado mapa de estruturas anatômicas (ASM) e uma arquitetura orientada a tarefas que eles chamam de *plan-solver*. O trabalho foi testado em uma base de imagens privada de 23 pacientes e teve uma precisão de 92% na segmentação da medula espinhal. Neste estudo os pacientes tinham entre 9 e 87 fatias. Em nosso trabalho, ambos os métodos de segmentação da medula espinhal são feitos em uma base de imagens de 36 pacientes, onde as fatias variam até 288 fatias. Além disso, o método é totalmente automático sem a necessidade de qualquer abordagem prévia. Ainda, obteve-se uma acurácia de 92,55% no primeiro método e 99,24% no segundo, o que mostra a parte proeminente das nossas abordagens.

Os trabalhos propostos por Banik, Rangayyan e Boag (2010), Rangayyan, Deglint e Boag (2006), usaram abordagens baseadas em Hough para a segmentação do canal vertebral. A abordagem de Hough é baseada em círculos que, ao tentar delinear a medula espinhal, se tornam um problema. Portanto, nossa abordagem tanto usando o IMSLIC para delimitar esse órgão quanto a utilizando classificação pixel-a-pixel pela U-Net, mostram-se mais robustas nessa tarefa.

Por sua vez, Macchia et al. (2012) comparou três softwares comerciais que realizam segmentação de órgãos com base em atlas: VelocityAI, MIM 5.1.1 e ABAS 2.0. Os índices de Dice obtidos por ABAS, MIM e VelocityAI são 70%, 81% e 78%, respectivamente. Contudo, assim como outros trabalhos, utilizam de atlas para segmentar a medula espinhal, e mesmo sendo softwares comerciais, nosso trabalho encontra resultados de Dice equiparados aos softwares nos dois métodos propostos

Em outro trabalho, proposto por Leener, Cohen-Adad e Kadoury (2015) faz a segmentação da medula espinhal em imagens de ressonância magnética. Os resultados alcançados pelo Dice foi de 91%. Este método foi testado em apenas 17 pacientes, os métodos apresentados em nosso método para segmentação de medula compreendem mais que o dobro deste base de imagens. Já o método de Verhaart et al. (2014) apresenta um método automático para segmentar vários OARs. Para a segmentação da medula

espinhal, o método é utilizado em 27 TCs e obteve um Dice 87%. Contudo, as métricas calculadas nesse trabalho, levam em consideração apenas fatias que existam medula espinhal, desconsiderando o restante da CT. Nos métodos propostos são tratadas de todas as fatias do paciente, independentemente da existência ou não de marcação do especialista, alcançando resultados promissores.

Ibragimov e Xing (2017) propuseram um método para a segmentação da medula espinhal em TC de cabeça e pescoço usando um CNN. O método foi aplicado em 50 tomografias e obteve um índice de Dice de 87%. Contudo, este método é aplicado apenas em TCs de cabeça e pescoço. Nosso método é aplicado a TC do corpo todo do paciente, e ainda assim apresenta resultados promissores.

Em Dong et al. (2019), é proposto um método para segmentação de órgãos de risco em tomografia de planejamento. Os autores propõem o uso de uma rede adversária juntamente com uma rede convolucional completamente conectada para alcançar seus resultados. Os autores aplicam seu método em uma base com 35 pacientes e alcançam resultados de 90% de Dice. Este trabalho utiliza a mesma base utilizada em nossos métodos, contudo os autores utilizaram apenas 35 volumes e não justificam a exclusão de um paciente. Além disto, o resultado alcançado para a medula espinhal é dependente de uma segmentação ideal dos pulmões. Ainda, o trabalho relata uma ROI de 64×64 , mas não justificam o critério de escolha deste tamanho. Por fim, os autores utilizam para validação dos resultados um treinamento *leave-one-out*, logo para cada volume testado o treinamento é executado por 34 volumes, tendo mais exemplos para generalizar o modelo. Nossos métodos não necessitam de segmentação prévia dos pulmões, e garante-se o treinamento e teste com exemplos dos três institutos da base de dados.

Em Liu et al. (2020), é proposto um método para segmentação de OARs. O método utiliza como técnica de segmentação uma variação da U-Net (RONNEBERGER; FISCHER; BROX, 2015), apresentando como resultado para a segmentação da medula espinhal 82,70% de Dice. O trabalho apresenta métricas relevantes e mostra a importância de redes U-Net para a segmentação da medula espinhal. Vale ressaltar que a base de dados é superior a nossa, possibilitando maior número de amostras no conjunto de treino, e, consequentemente, um modelo mais generalista.

A Tabela 6 mostra a relação entre esses trabalhos e o método proposto.

Em relação aos trabalhos acima mencionados, mostra-se que os trabalhos sobre a segmentação de medula espinhal em TC de planejamento ainda são escassos, e muitos

Tabela 6 – Comparação com os trabalhos relacionados.

SEN - sensibilidade; SPE - especificidade; ACC - acurácia; Pt - número de pacientes.

Trabalhos	Tipo	Base de imagens	Técnica	Pt	SEN	SPE	ACC	Dice
Archip et al. (2002)	TC	Privada	Semi-automático	23	—	—	92%	—
Rangayyan, Deglint e Boag (2006)	TC	Privada	Semi-automático	4	—	—	—	—
Banik, Rangayyan e Boag (2010)	TC	Privada	Semi-automático	13	—	—	—	60%
VelocityAI(MACCHIA et al., 2012)	TC	Privada	Semi-automático	15	—	—	—	70%
MIM 5.1.1(MACCHIA et al., 2012)	TC	Privada	Semi-automático	15	—	—	—	81%
ABAS 2.0(MACCHIA et al., 2012)	TC	Privada	Semi-automático	15	—	—	—	78%
Leener, Cohen-Adad e Kadoury (2015)	RM	Privada	Semi-automático	17	—	—	—	91%
Verhaart et al. (2014)	TC	Privada	Semi-automático	27	—	—	—	87%
Ibragimov e Xing (2017)	TC	AAPM Thoracic Auto-segmentation	Automático	36	—	—	—	87%
Dong et al. (2019)	TC	AAPM Thoracic Auto-segmentation	Automático	35	93%	99%	—	90%
Liu et al. (2020)	TC	Privada	Automático	105	—	—	—	82,70%
Método proposto (CNN)	TC	AAPM Thoracic Auto-segmentation	Automático	36	92,87%	92,55%	89,23%	78,20%
Método proposto (ResU-Net)	TC	AAPM Thoracic Auto-segmentation	Automático	36	91,52%	99,38%	99,24%	81,69%

usam o atlas ou informações anatômicas dos modelos para auxiliar nessa segmentação. Alguns estudos também detectam o canal vertebral, que é onde a medula espinhal está localizada. Essa abordagem é interessante, pois, ao proteger a medula espinhal no processo de radioterapia, é interessante ter uma região de depuração para possíveis diferenças milimétricas posicionais do paciente.

Assim, ressalta-se que nosso trabalho apresenta dois métodos totalmente automáticos, sendo o primeiro baseado em *superpixels* e CNN, que além de melhor delimitar a região da medula espinhal, não requer conhecimento *a priori*. O segundo método apresenta a segmentação de medula espinhal baseado em U-Net. Também se destaca que existe uma grande dificuldade em fazer uma comparação fiel, já que as bases de imagens não são as mesmas. Algumas bases de dados são diferentes do trabalho apresentando, portanto não foi possível aplicar o método proposto a eles. Contudo, vale ressaltar que os métodos propostos são os que utilizam o maior número de pacientes para validar os métodos, o que também pode-se destacar como robustez. Além disso, em comparação com as métricas de validação, pode-se ver que nossas abordagens têm um lugar de destaque. Com isso, demonstra-se a viabilidade de usar os métodos propostos para a tarefa de segmentação de medula espinhal.

5 Segmentação de esôfago utilizando U-Net com blocos residuais

Para a segmentação de esôfago em TC de planejamento é proposto um método dividido em 4 etapas. Primeiro, é a etapa de aquisição de imagem, a base utilizada é a descrita na Seção 4.1. A marcação do esôfago, assim como muitos outros trabalhos relatam (FECHTER et al., 2017), é um dos OARs com maior dificuldade de delineamento. Mesmo os especialistas têm dificuldade em definir os limites do esôfago em TC de maneira confortável, uma vez que sua textura facilmente é confundida por outras estruturas em volta dele. A segunda etapa, na segmentação da VOI utilizam-se as informações espaciais de um atlas do esôfago para fazer um corte tridimensional no volume. Em seguida, a fim de realçar as estruturas do esôfago, faz-se um pré-processamento utilizando filtro bilateral seguido da equalização do histograma. Em seguida, na etapa de segmentação final, as fatias da VOI são apresentadas a uma U-Net com blocos residuais com o intuito de criar um modelo capaz de segmentar a região do esôfago e, por fim, é aplicada uma técnica de refinamento para melhorar a segmentação do esôfago. A Figura 33 descreve os passos desse método. Nas próximas seções, serão explicados detalhes de cada etapa.

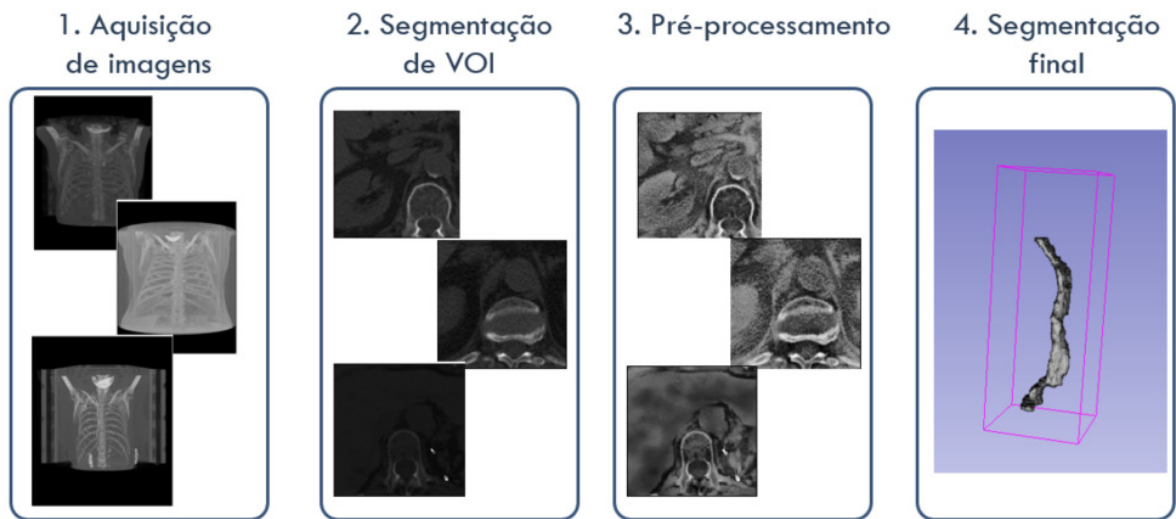


Figura 33 – Fluxograma do método do esôfago.

5.1 Segmentação de VOI

Como já mencionado no Capítulo 2, o esôfago é um dos órgãos mais problemáticos para segmentação. Um dos maiores agravantes é não haver uma definição visível das paredes do esôfago, aliado ao fato de que os órgãos circundantes ao esôfago possuem intensidade semelhante a ele nas TC de planejamento. Outro ponto que se deve levar em consideração é a abertura do esôfago, em algumas fatias isto é perceptível, em outras não. Ou seja, além da problemática já observada na base de imagens, ainda existem os problemas na estrutura anatômica do esôfago nos exames de TC de planejamento. Outro ponto é que o esôfago ocupa poucas fatias do exame de TC, e mais uma vez a literatura mostra a importância de delimitar uma região de interesse como etapa inicial do processo de segmentação (DONG et al., 2019; TRULLO et al., 2017a; LARSSON; ZHANG; KAHL, 2017; FEULNER et al., 2011). A fim de contornar o problema da diversidade da base de imagens e definição de uma região de interesse, é proposto uma etapa de segmentação de VOI. Esta etapa é composta por três subetapas, são elas o registro dos volumes, geração de atlas e corte na região do esôfago.

5.1.1 Registro de volumes

Como apresentado na Seção 4.1, a base de imagens é muito diversificada, apresentando exames de três institutos diferentes. Ainda, o número de fatias varia, alguns exames (como do instituto 3) apresentam fatias até a região da cabeça e outros não. Uma forma de padronizar o número de fatias é pela técnica de registro (Seção 3.3.2.3). Neste método, aplicou-se o registro de imagens em toda a base de imagens, tomando como imagem fixa o paciente que apresentava o menor número de fatias. Esse critério foi adotado pelo fato que aproximar as fatias de um paciente com poucas fatias a um com muitas geraria informações que não seriam reais, uma vez que alguma técnica de interpolação seria necessária para aproximação neste caso.

O registro de imagem foi aplicado tanto ao volume do paciente, quanto ao volume de marcação correspondente. Um exemplo de um volume registrado pode ser visto na Figura 34.

Com os volumes dos exames e marcações registradas, a próxima etapa é a geração do atlas da base de imagens.

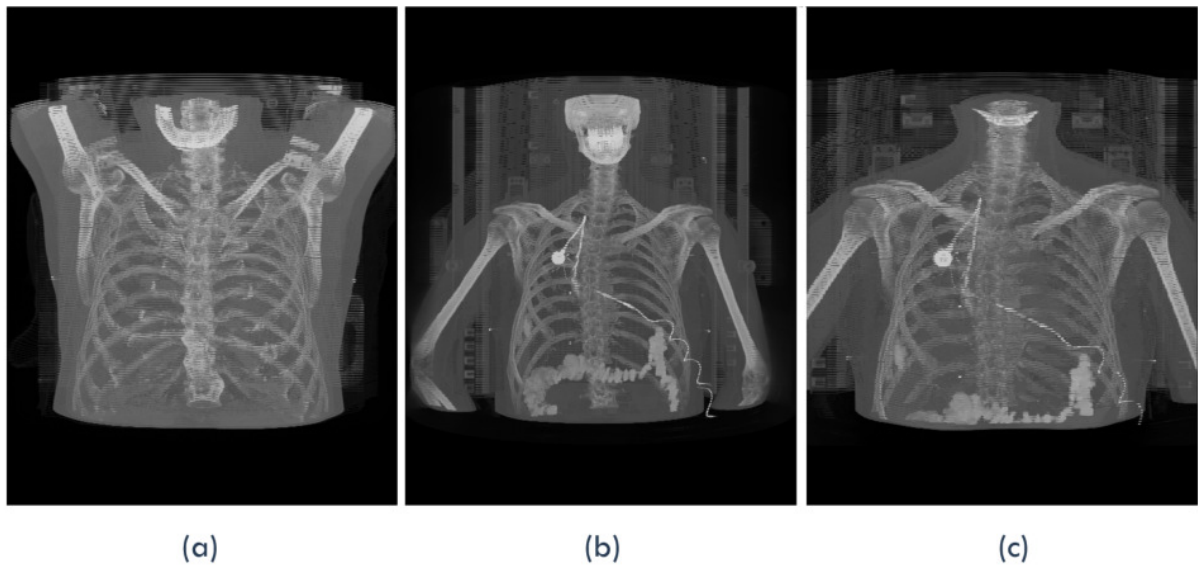


Figura 34 – Registro de volumes: (a) volume fixo com 134 fatias, (b) volume móvel com 279 fatias e (c) volume móvel registrado com 134 fatias.

5.1.2 Geração de atlas

A geração do atlas consiste em gerar um volume com as marcações registradas. A finalidade do atlas para a etapa de segmentação de VOI é gerar um volume com a localização das regiões dos esôfagos na base de imagens. Diferente do trabalho de Feulner et al. (2011) e Larsson, Zhang e Kahl (2017), onde os modelos probabilísticos e atlas são propostos para dirigir a classificação do OAR, nosso método utiliza o atlas apenas para auxiliar a próxima etapa de corte do volume. O atlas é construído conforme a Seção 3.3.3, e a Figura 35 mostra uma ilustração de como o atlas foi gerado.

Observa-se que apesar de ainda serem diversificadas as marcações do esôfago nos volumes, elas estão espacialmente localizadas em uma pequena porção do volume. Com essa informação é possível realizar um corte no volume gerando uma VOI.

5.1.3 Corte na região do esôfago

A partir da geração do atlas é possível obter informação espacial do esôfago. Assim, é feito um corte na região do esôfago em todos os volumes da base de imagens obedecendo ao critério de maior região encontrada no atlas. Assim, todos os volumes anteriormente registrados, agora são cortados juntamente com suas marcações, resultando em apenas

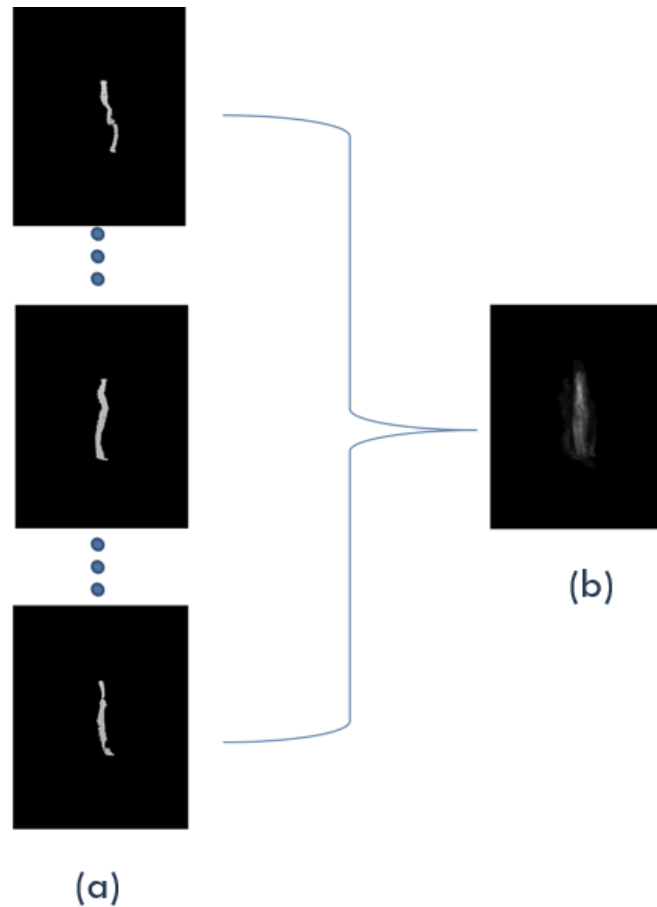


Figura 35 – Geração de atlas: (a) exemplo de volumes de marcações e (b) resultado da geração do atlas.

uma região que compreende a possível localização do esôfago. Um exemplo pode ser visto na Figura 36.

Assim, ao fim dessa etapa têm-se as VOIs de todos os pacientes e de todas as marcações. Com isso, reduz-se o escopo de trabalhar com todo o volume para a segmentação do esôfago. Ainda, apesar da existência de um desbalanceamento entre voxels de não esôfago, a proporção diminui de forma significativa. O próximo passo é o pré-processamento.

5.2 Pré-processamento

Ao observar a literatura (Capítulo 2), não se apresentam trabalhos preocupados em utilizar algum melhoramento da imagem para destacar a região do esôfago das demais regiões, uma vez que, em TC não há um contraste bem definido entre o esôfago e os demais tecidos circundantes. Assim, sugere-se o uso de pré-processamento que busque fazer o

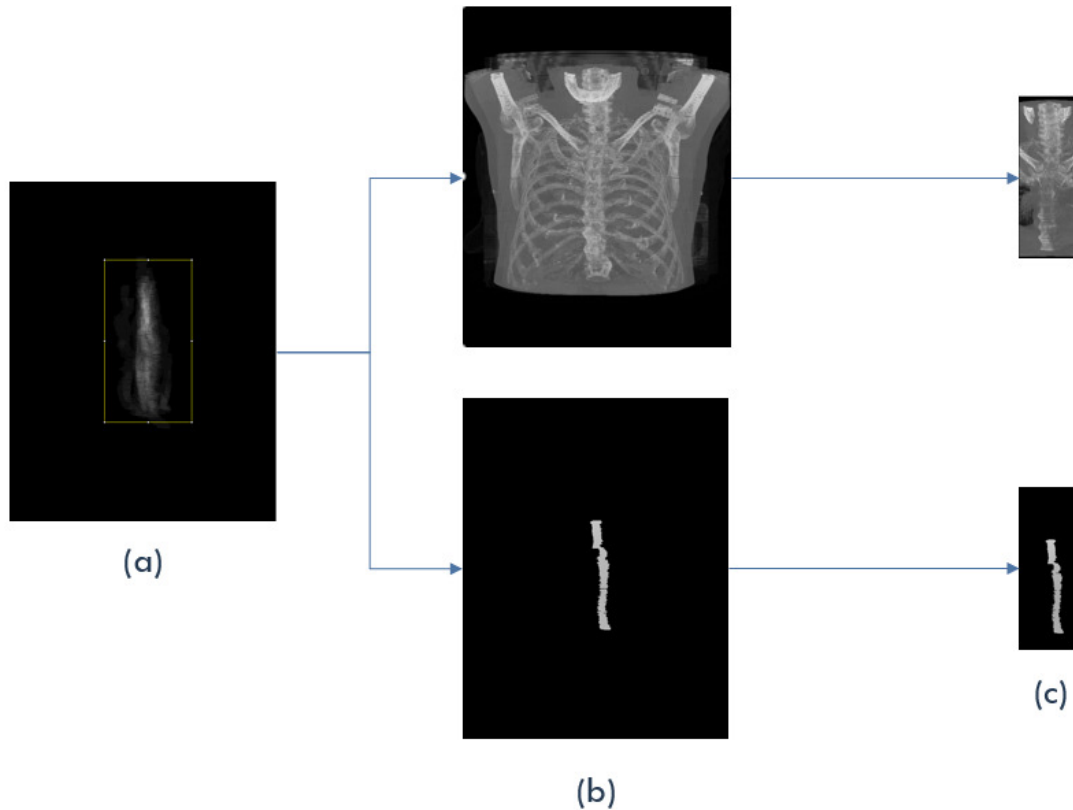


Figura 36 – Corte na região do esôfago: (a) atlas, em amarelo região correspondente ao volume do maior objeto, (b) volumes originais do paciente e da marcação (c) segmentação da VOI do paciente e da marcação.

melhoramento da região do esôfago a fim de melhorar a acurácia do classificador na tarefa de segmentação do esôfago.

Para a etapa de pré-processamento, propõe-se o uso da sequência de dois filtros (Seção 3.3.2.1). O primeiro filtro é o bilateral, a principal característica é a suavização das imagens enquanto preserva as bordas, por meio de uma combinação não linear de valores de imagem próximos. Pelo fato de não ser nítida a distinção do esôfago com os demais tecidos circundantes, este filtro mostra-se promissor ao destacar as bordas.

Em seguida, é utilizado uma equalização do histograma nas imagens realçadas pelo filtro bilateral. Por criar uma imagem com níveis de cinza igualmente distribuídos, pois acredita-se que melhore o contraste entre esôfago e demais tecidos. Um exemplo de uma fatia da VOI com a etapa de pré-processamento é mostrado na Figura 37.

Assim, com a VOI definida e pré-processada, o próximo passo é segmentar o esôfago.

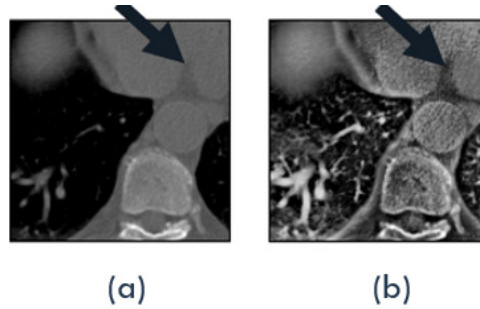


Figura 37 – Pré-processamento: (a) fatia da VOI sem pré-processamento, (b) fatia da VOI com pré-processamento. A seta indica o nítido realce das bordas do esôfago.

5.3 Segmentação final

A segmentação final é dividida em algumas subetapas. Primeiro, é definido a entrada da rede, em seguida é executado o treino e teste, por fim, é aplicado o refinamento da segmentação para eliminar os falsos positivos da imagem.

5.3.1 Gerando imagens de entrada

O resultado da segmentação das VOIs produz dois volumes, um do paciente e outro da marcação. O volume do paciente passa pela etapa de pré-processamento, a fim de realçar as estruturas internas e melhorar a representação do esôfago. Foi utilizada uma abordagem 2D, ou seja, a VOI de cada paciente foi passada fatia por fatia na rede. No entanto, ao final da classificação da rede, é construído um volume 3D.

Optou-se por usar as fatias em vez do volume, porque, se for considerado o número de volumes existentes, são apenas 36, o que significaria que a rede teria uma pequena amostra na fase de treinamento. Assim, ao utilizar as fatias, o número de conjuntos de imagens de treinamento/validação e teste é aumentado.

O próximo passo é o treinamento da U-Net (Seção 3.2.5.2) para criar um modelo robusto capaz de classificar novos pacientes. A rede utilizada para esta tarefa é uma U-Net com blocos residuais proposta por Kolařík et al. (2019) e que também mostrou melhor resultado no método proposto para segmentação da medula espinhal (Seção 4.3).

No método proposto para esôfago, cada convolução é seguida por uma camada de normalização de *batch* e uma camada de *Dropout*, com uma taxa de 20% para desativação de neurônios.

5.3.2 Treino e teste da U-Net com blocos residuais

Nesta seção, a base de imagens é dividida em dois grupos: o primeiro com as fatias de entrada de treinamento e o segundo com as fatias de entrada de teste. A base de treinamento é composta por fatias de pacientes e fatias correspondentes de marcações especializadas. Na fase de treinamento, 10% dos dados é separada para validação; é a partir da validação que são feitas simulações para encontrar os melhores parâmetros da rede. No final do treinamento e validação, o modelo é gerado. Este modelo é então aplicado nas imagens separadas para teste.

A fase de teste consiste em usar o modelo para classificar novas imagens, desconhecidas para o modelo. Para validar os resultados do teste, a imagem de marcação do especialista é obtida e as métricas de validação são calculadas, assim, é avaliado se o modelo foi bom o suficiente na segmentação do esôfago como um OAR.

Por se tratar de um OAR que não apresenta um limite entre ele e os demais tecidos bem definidos visualmente nas TCs de planejamento, é proposto mais uma subetapa para refinar a segmentação.

5.3.3 Refinamento da segmentação

Foi observado que apesar da ResU-Net apresentar resultados satisfatórios, ao gerar as imagens preditas pelo modelo, é possível identificar elementos que não correspondem ao esôfago. O caso mais comum é a geração de dois ou mais objetos preditos em uma fatia, normalmente um sendo o esôfago e outro uma estrutura menor. Para eliminar essas demais estruturas é mantido nas fatias apenas o objeto com maior área.

Como é observado na Figura 38a, muitas estruturas que não são esôfago foram classificadas como falsos positivos. A etapa de manter o maior objeto consegue eliminar estes objetos.

Outro ponto é que após manter apenas o maior objeto (Figura 38b), percebe-se que a região predita como esôfago é sempre maior que a região da marcação do especialista (Figura 38 em verde). Isto se dá pelo fato destacado pela maioria dos autores que propuseram métodos para segmentação do esôfago, este órgão não tem uma diferença perceptível que separe ele dos tecidos em sua volta. Além disso, os filtros 3×3 convolucionais aplicados nas etapas de convolução da ResU-Net extraem características associadas à

vizinhança, então acabam por considerar uma região maior no esôfago. Uma solução para refinamento final deste problema, foi a aplicação de uma erosão na marcação predita, isto proporcionou melhores resultados ao método proposto (Figura 38c). Foi utilizado como elemento estruturante um círculo com raio de 2×2 . A Figura 38 mostra um exemplo desta etapa.

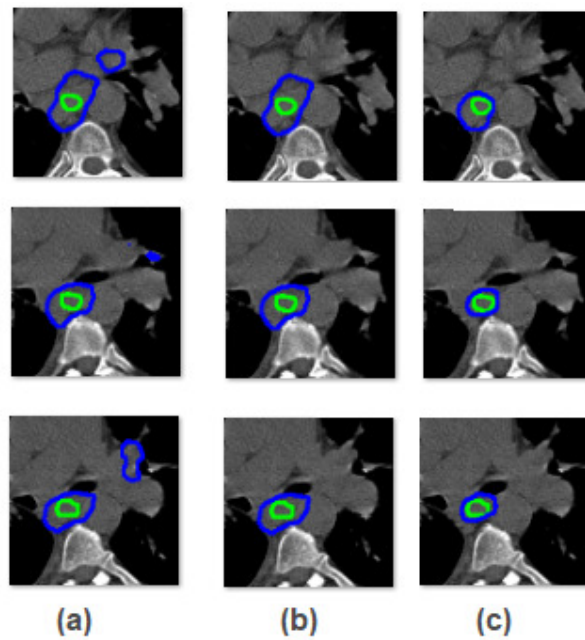


Figura 38 – Segmentação Final: (a) exemplos de marcações especializadas em verde e resultado de ResU-Net em azul, (b) marcações especializadas em verde e resultado de obtenção da maior região em azul, e (c) marcações especializadas em verde e resultado de erosão em azul.

Ao fim de todas as etapas da segmentação do esôfago, as métricas de validação foram calculadas - Dice, sensibilidade (SEN), especificidade (SPE) e acurácia (ACC). Os resultados para o método são apresentados na seção que segue.

5.4 Resultado

Os resultados alcançados na tarefa de segmentação do esôfago, usando o método proposto no planejamento de radioterapia, serão descritos nesta seção. Primeiramente, destaca-se o ambiente de execução de treino e teste.

5.4.1 Ambiente de treino/teste

O ambiente de treinamento para a etapa de Segmentação da VOI e Pré-processamento foi um computador com Intel Core i5-7300HQ a 2,50GHz, 8 gigabytes de memória DDR3. A etapa de segmentação da VOI composta por registro e corte da região do esôfago durou aproximadamente 30 segundos por exame. Por sua vez, a etapa de pré-processamento se mostrou mais custosa excedendo 1 minuto em cada exame, devido ao custo de execução do filtro bilateral. Na construção do modelo de rede e das arquiteturas U-Nets, foi usada uma máquina com duas placas gráficas NVIDIA Titan X, na qual foi utilizado a biblioteca Keras com o TensorFlow-GPU como back-end.

5.4.2 Aquisição de imagens

Como já mencionado na Seção 4.1, o banco de dados é composto por três institutos, com 12 pacientes por instituto, resultando em 36 volumes de pacientes. Para treinar e validar este método, o banco de imagens foi dividido em 2 conjuntos de dados. O primeiro para treinamento, composto por 30 pacientes (10 por instituto) e 6 por teste (2 por instituto). Análogo aos métodos de segmentação de medula espinhal, essa divisão foi feita aleatoriamente garantindo apenas a proporção dos institutos nos dois conjuntos de dados.

Além disso, após todas as etapas usando os 30 volumes no treinamento e o teste em 6 volumes, as métricas de validação são calculadas para mostrar a robustez do modelo.

5.4.3 Segmentação de VOI e Pré-processamento

Na etapa de segmentação de VOI, descrita na Seção 5.1 os dois conjuntos de treino e teste passaram pelo registro, sendo a imagem fixa a imagem com menor número de fatias do conjunto. Na geração do atlas, ao encontrar a maior região que se encontrava nas marcações do especialista, encontrou-se um tamanho de VOI de $128 \times 128 \times 112$, assim, todos os volumes da base de imagens foram cortados nesta proporção. Com isso, o que anteriormente eram volumes de $512 \times 512 \times (\text{número de fatias})$ agora torna-se uma VOI padronizada com menos estruturas para serem classificadas.

Em seguida, todas essas VOIs passaram pela etapa de pré-processamento, onde primeiramente foi aplicado o filtro bilateral seguido de uma equalização do histograma.

Agora, com os volumes dos pacientes e da marcação definidos, a próxima etapa é a segmentação final.

5.4.4 Segmentação final

Nesta etapa, os resultados encontrados nas etapas anteriores serão apresentados a U-Net com blocos residuais (ResU-Net). Primeiro, como destacado na Seção 5.3, foram apresentadas as fatias dos volumes de treino para a rede, esta estratégia é adotada como medida para aumentar o número de amostras no treinamento.

Como as imagens obtidas na segmentação inicial foram de $128 \times 128 \times 112$, as imagens passadas para a rede são fatias de 128×128 . Após várias seções de treinamento e configuração de parâmetros, os parâmetros usados no treinamento das redes foram: tamanho do *batch* igual a 3, otimizador Adam (KINGMA; BA, 2015) com taxa de aprendizado inicial igual a 0,0001 e 10% de volumes para validação. A rede foi treinada com 300 épocas e a função de perda foi a *Diceloss*.

Após o treinamento da rede, os volumes de testes são apresentados ao modelo, e as métricas de validação são calculadas. Vale ressaltar, que a base de teste é composta por 2 pacientes de cada instituto.

O modelo se encarrega de classificar os volumes de teste e como destacado na Seção 5.3, apesar de ser robusto em classificar os volumes, algumas fatias geram alguns falsos positivos, que primeiramente são descartados mantendo apenas o maior objeto nas fatias. Por fim, é aplicado uma erosão, a fim de reduzir o tamanho da marcação predita, uma vez que ela sempre excede o tamanho da marcação demarcada pelo especialista.

A fim de ter uma visão geral de cada uma das etapas e o quanto de melhora é obtida em cada uma, na Tabela 7, são apresentados os resultados da ResU-Net sem pré-processamento, com pré-processamento, mantendo o maior objeto, mantendo o maior objeto e aplicando a erosão no conjunto de teste.

Tabela 7 – Resultados das etapas da segmentação do esôfago.

Etapas	Tempo treino		Dice(%)	SEN(%)	ESP(%)	ACC(%)
Sem pré-processamento	320	20	44,35	100	98,02	98,03
Com pré-processamento	320	20	65,68	100	99,19	99,18
Com pré-processamento + Maior Objeto	320	31	67,05	99,89	99,22	99,23
Com pré-processamento + Maior Objeto + Erosão	320	43	82,15	90,61	99,76	99,69

Analisando a Tabela 7, a etapa de pré-processamento mostrou uma melhora significativa nos resultados. Acredita-se que pelo fato de definir melhor as paredes do esôfago, os resultados da segmentação em níveis de Dice subiu de 44,35% para 65,68% mostrando a importância da etapa. Apesar dos dois resultados estarem em 100% de sensibilidade, o resultado sem pré-processamento apresenta um maior número de falsos positivos, ao analisar a métrica de especificidade ela subiu quase 1%.

Observando a etapa de manter apenas o maior objeto, garante-se uma melhora em relação ao Dice e especificidade, ou seja, há menos falsos positivos e uma segmentação mais precisa. Apesar de haver uma perda na sensibilidade, vale lembrar que as métricas são medidas por pixels, ou seja, existem mais números de pixels pertencentes a classe de não esôfago do que esôfago, então qualquer melhora na especificidade (acertar os casos negativos) melhora significativamente o valor de Dice.

Logo, ao utilizar a erosão para reduzir o tamanho do objeto encontrado nas fatias, há uma melhora expressiva dos resultados. O Dice que era 67% sem essa técnica, subiu para 82,15%, e as métricas de especificidade e acurácia aumentam para próximo de 100%. Vale destacar, mais uma vez, que os pixels de não esôfago são maioria nas imagens, e melhoras no acerto desta classe traz resultados robustos ao Dice.

Ao observar a métrica de sensibilidade com o método completo, o mesmo apresenta uma queda de 10%, ao contrário do que acontece com a especificidade, pelo fato de haver menos pixels de esôfago nas fatias, qualquer erro nesta classe compromete a métrica de sensibilidade. Todavia, observando a Seção 5.3, os resultados sem a etapa de erosão acabam por classificar uma área muito extensa sendo esôfago, que apesar de acertar todos os pixels de esôfago, torna-se pouco útil por compreender diversos tecidos ao redor dele.

Assim, tendo como base Dice, que mostra o quanto próximo a predição é da marcação do especialista, acredita-se que o método para segmentação de esôfago é promissor, este método empregado em grandes centros pode ser de fundamental ajuda nos tratamentos de radioterapia.

5.4.5 Considerações

A segmentação do esôfago não é uma tarefa trivial. Desenvolver um método capaz de contornar todas as adversidades deste tipo de OAR e que consiga uma boa taxa de

acerto ainda é um desafio quando se trata de imagens médicas. O método proposto para segmentação de esôfago se mostrou bastante robusto e apresentou uma série de vantagens:

1. Pode-se destacar um método completamente automático, que não requer de conhecimento prévio (como informação anatômica ou inserção das informações do atlas para a U-Net) para conduzir a segmentação;
2. A etapa de segmentação de VOI se mostrou bastante pertinente, uma vez que o esôfago representa uma parcela bem pequena do exame de TC de planejamento, propor um método capaz de reduzir o escopo da segmentação final é de crucial importância para os resultados;
3. Um dos principais desafios do esôfago, já destacado em todo o trabalho, é que este OAR não possui uma distinção clara entre seu tecido e os tecidos ao redor dele, analisando a literatura, observou-se que a maioria dos trabalhos se preocupava em localizar este OAR a partir de informações espaciais, seja por modelos probabilísticos ou técnicas de redução de escopo, assim, uma das contribuições para essa tarefa foi a etapa de pré-processamento;
4. A etapa de pré-processamento se mostrou crucial para alcançar melhores resultados, o método conseguiu uma melhora de mais de 20% de Dice ao utilizar a sequência dos dois pré-processamentos (filtro bilateral e equalização do histograma);
5. A estratégia de utilizar fatias na etapa de treinamento da rede também se mostrou efetiva. Uma abordagem 3D reduziria consideravelmente o número de amostra, ao utilizar fatias, contorna-se esse problema;
6. A rede proposta para segmentação do esôfago também foi uma etapa crucial para alcançar bons resultados. Acredita-se que, pelo fato de o esôfago compreender pouca região da VOI, ao utilizar a técnica de blocos residuais entre as camadas, mantém-se as características deste OAR propiciando melhores resultados;
7. Ainda, não encontrou-se na literatura trabalhos que utilizam blocos residuais na U-Net para a segmentação do esôfago, mais uma vez, apresentando avanços com o método proposto;
8. A etapa para manter o maior objeto também é importante para o método. Observou-se que por vários tecidos em volta do esôfago apresentarem características semelhantes a ele, a rede acaba predizendo alguns objetos menores como esôfago, logo, excluir esses objetos melhora o Dice do método;

9. É possível notar que pelo fato do tecido em volta ao esôfago ser muito semelhante a textura do próprio esôfago, a rede tende a segmentar uma região maior, porém, na maioria dos casos, não há perda do esôfago, fato que gera 100% de sensibilidade em alguns os testes. Ao considerar isto, aplicou-se uma etapa de erosão, onde foi possível melhorar significativamente as métricas;
10. A combinação de todas as técnicas empregadas neste método, resultou em um Dice de 82,15%, sensibilidade de 90,61%, especificidade de 99,76% e acurácia de 99,69%. Este resultado supera os demais trabalhos apresentados até então na literatura;
11. Por fim, ressalta-se mais uma vez a problemática da base de imagens que além de diversificada, apresenta a estrutura de esôfago como um OAR muito difícil de segmentar, apesar de todas as adversidades o método conseguiu alcançar um Dice de 82,15%.

Apesar do método ser muito promissor, ainda apresenta algumas limitações e pontos negativos:

1. A etapa de treinamento conta com uma rede de segmentação composta por inúmeros parâmetros e camadas, esses parâmetros são ajustados na fase de validação, o que geralmente é uma tarefa árdua;
2. Outro ponto é o refinamento da segmentação, apesar de produzir resultados de Dice muito expressivos, a medida que a técnica de erosão reduz o objeto segmentado pela ResU-Net, há uma certa perda na métrica de sensibilidade;
3. Por fim, observa-se uma grande diferenciação entre os três institutos utilizados, o que de fato ocorre no processo de aquisição real. Sabe-se que cada centro de triagem e planejamento a radioterapia tem sua forma e protocolo de fazer aquisição das imagens, assim, uma forma de padronizar a textura dos exames poderia melhorar os resultados.

Assim, afirma-se que todas essas vantagens e limitações agregam importante valor ao método proposto para segmentação do esôfago. A combinação das técnicas propiciou resultados promissores, e acredita-se que o uso deste método em sistemas CAD seria de fundamental importância na demarcação deste OAR pelo especialista.

5.5 Estudo de caso da segmentação de esôfago

Nesta seção, é apresentado alguns estudos de caso. Cada caso mostrará um paciente de um dos três institutos. Além disso, discute-se os resultados alcançados pelo método em cada um desses casos.

5.5.1 Estudo do caso - Instituto 1

O primeiro estudo de caso é um paciente do Instituto 1 da base de imagens. Neste caso, pode-se observar o quanto o método de segmentação do esôfago foi robusto. A Figura 39 mostra uma imagem de várias VOIs do mesmo paciente segmentada pelo método.

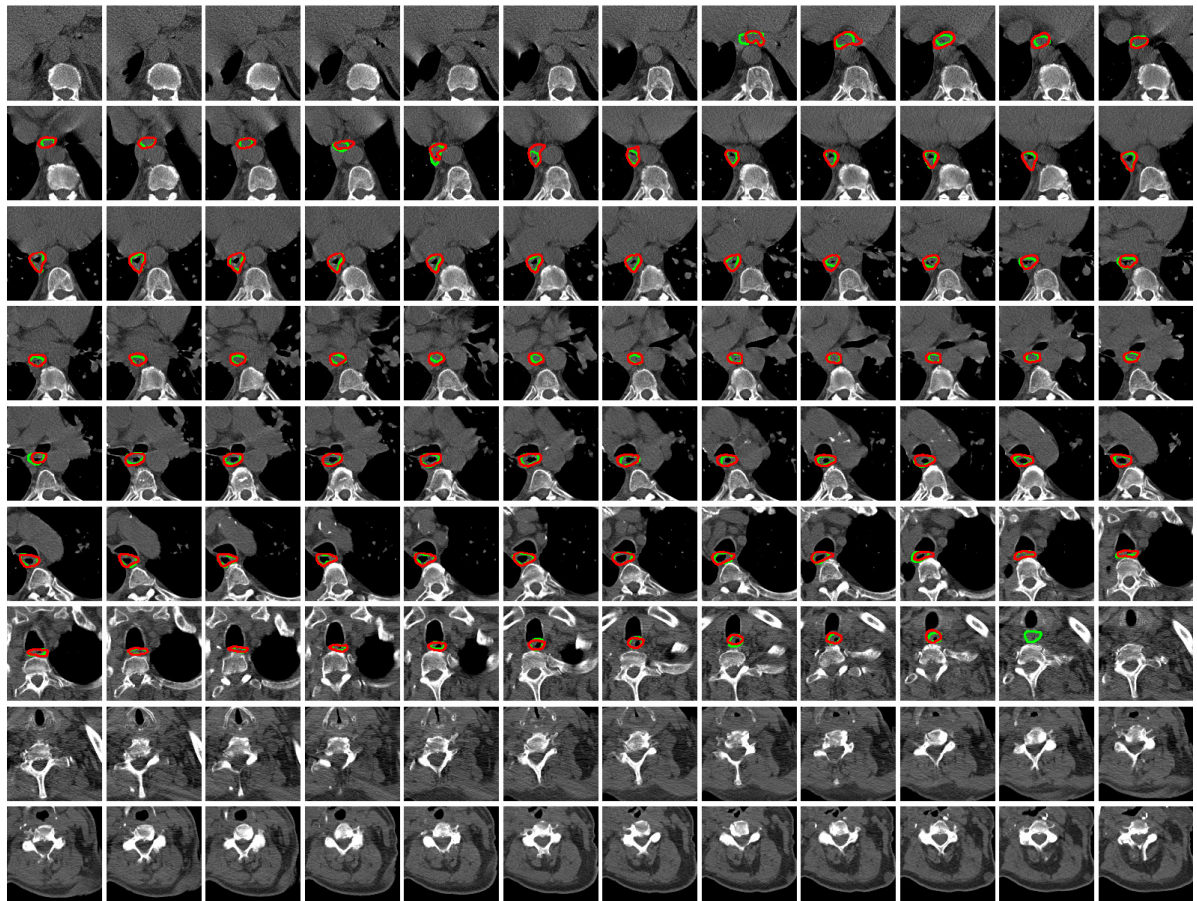


Figura 39 – Estudo de caso - Instituto 1. Em verde a marcação do especialista e em vermelho a segmentação do esôfago predita pelo método.

A Figura 39 mostra o quanto é próxima a método proposto com a marcação do especialista. O método foi capaz de segmentar uma região muito parecida com o especialista

gerando o mínimo de falsos positivos. Contudo, pode-se observar que houve perda do esôfago em uma das fatias. Esta perda, apesar de apenas uma fatia, afeta negativamente o valor das métricas, principalmente do Dice. Ainda assim, as fatias que não havia esôfago o método não segmentaram regiões a mais.

Mesmo o método sendo submetido a uma rede 2D, propõe-se uma visualização tridimensional dos resultados, tanto como forma de validação quanto para uma melhor visualização do especialista. Esta visualização pode ser vista na Figura 40.

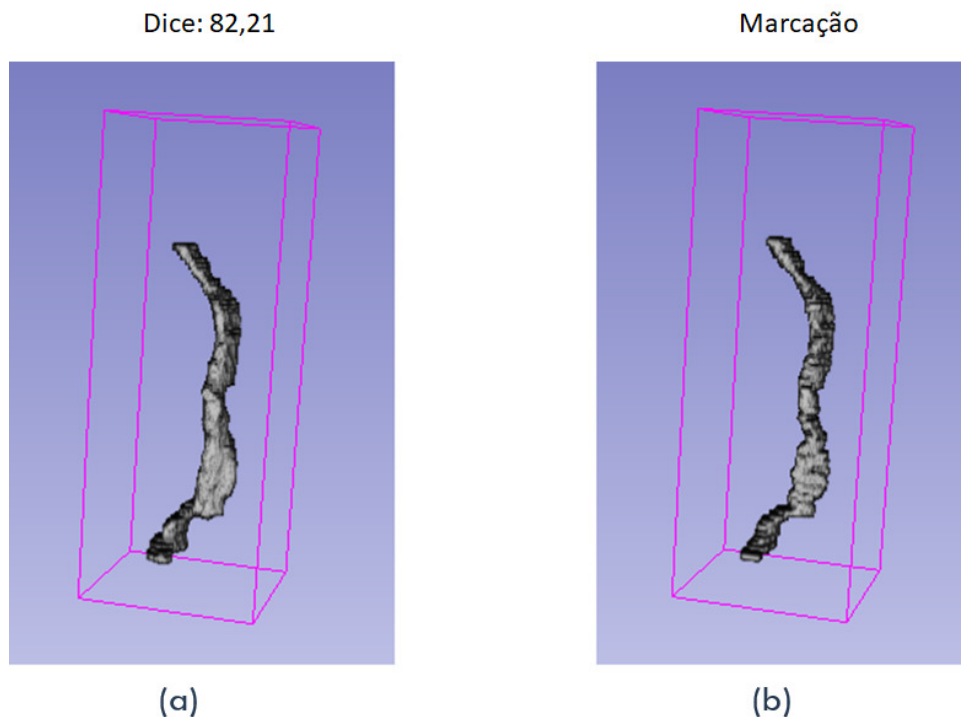


Figura 40 – Estudo de caso - Instituto 1. (a) volume predito pelo método e (b) volume da marcação do especialista.

Analisando as Figuras 40 e Figura 39, a perda de apenas uma fatia não comprometeu a visualização tridimensional da segmentação do esôfago (Dice de 82,21). Ao comparar a marcação do especialista com o resultado do método, mostra-se resultados robustos e eficientes para a segmentação do esôfago em TC de planejamento.

5.5.2 Estudo do caso - Instituto 2

Análogo ao primeiro estudo de caso, neste mostram-se os resultados de um paciente da base de teste do Instituto 1. Novamente, o método foi considerado robusto na tarefa de

segmentação do esôfago. A Figura 41 mostra uma imagem do resultado da segmentação do esôfago.

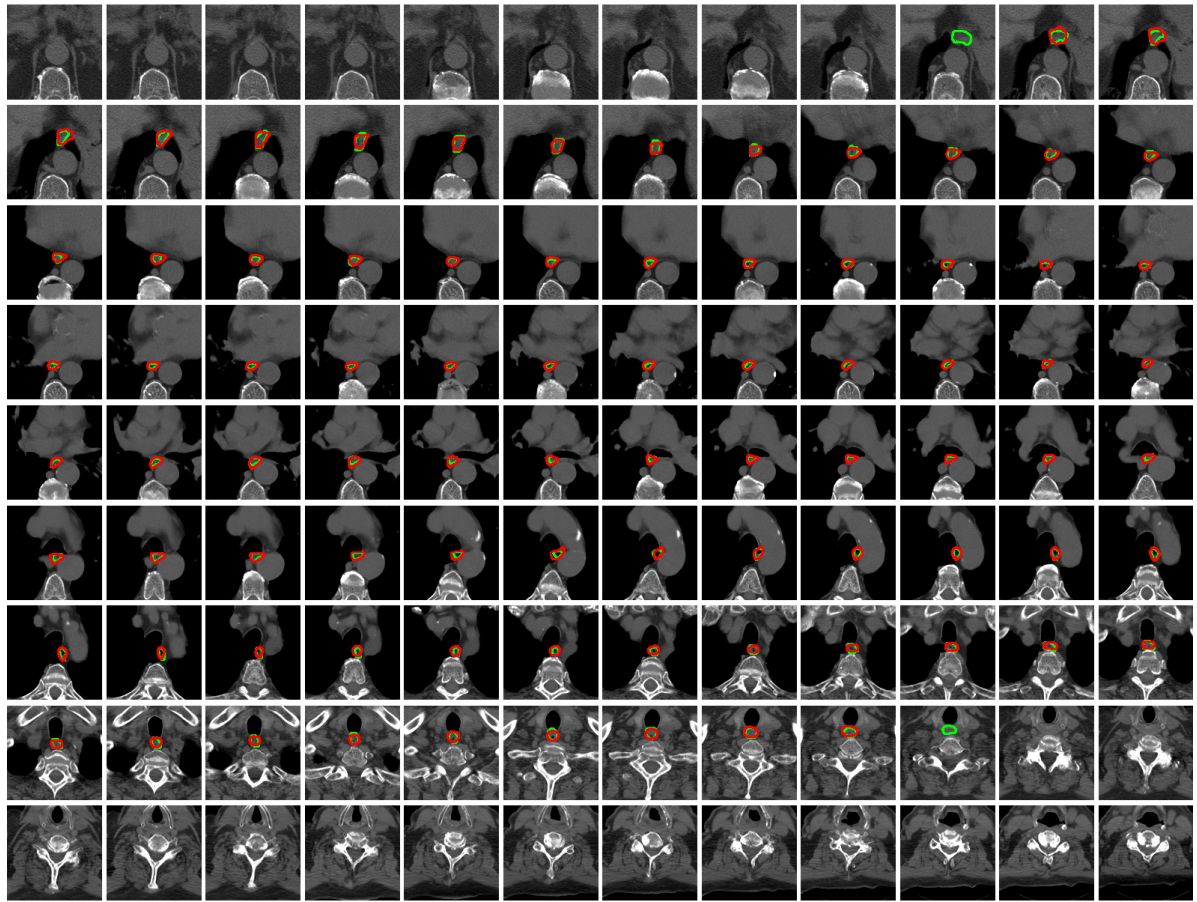


Figura 41 – Estudo de caso - Instituto 2. Em verde a marcação do especialista e em vermelho a segmentação do esôfago predita pelo método.

Mais uma vez é mostrado a eficiência do método. Nota-se que neste caso houve perda de duas fatias marcadas pelo especialista. Porém, ao analisar o volume como um todo, são 112 fatias submetidas a segmentação, e houve perda apenas de duas e não foram geradas marcações em fatias que não havia marcações do especialista. Desta forma, mostra-se o quanto o método proposto é robusto na segmentação de esôfago em TC de planejamento.

Assim como no estudo de casos do primeiro instituto, também é proposto uma visualização 3D do resultado alcançado no paciente do instituto 2. A Figura 42 apresenta uma visualização 3D do resultado.

Assim como no primeiro estudo de caso, neste há uma grande similaridade entre a marcação do especialista e o volume predito pelo método (Dice de 81,63). Esta visualização tridimensional mostra uma visão macro e, consequentemente, uma análise de forma global

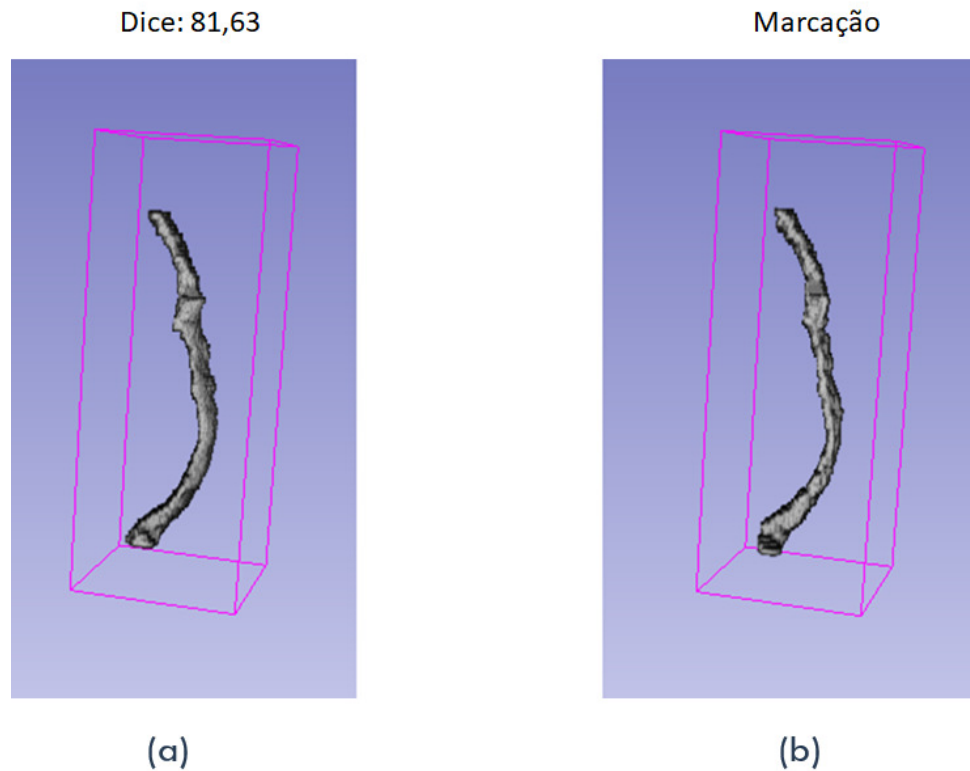


Figura 42 – Estudo de caso - Instituto 2. (a) volume predito pelo método e (b) volume da marcação do especialista.

do resultado do método. Acredita-se que a perda de poucas fatias, apesar de problemático, não invalida a segmentação em uma visão 3D do problema.

5.5.3 Estudo do caso - Instituto 3

Por fim, é mostrado um estudo de caso de um paciente do Instituto 3. Análogo aos casos anteriores, a Figura 43 mostra as VOIs segmentadas pelo método e a marcação do especialista.

Novamente, houve a perda do esôfago em uma fatia. Todavia, como já mencionado, o exame possui 112 fatias, ao considerar que as demais fatias foram segmentadas de forma correta, ou mesmo semelhante ao especialista, garante-se que o método é muito robusto na segmentação do esôfago.

Análogo aos casos de teste anteriores, neste também é gerado uma visualização tridimensional do resultado do método comparado a marcação do especialista. Esta visualização está apresentada na Figura 44.

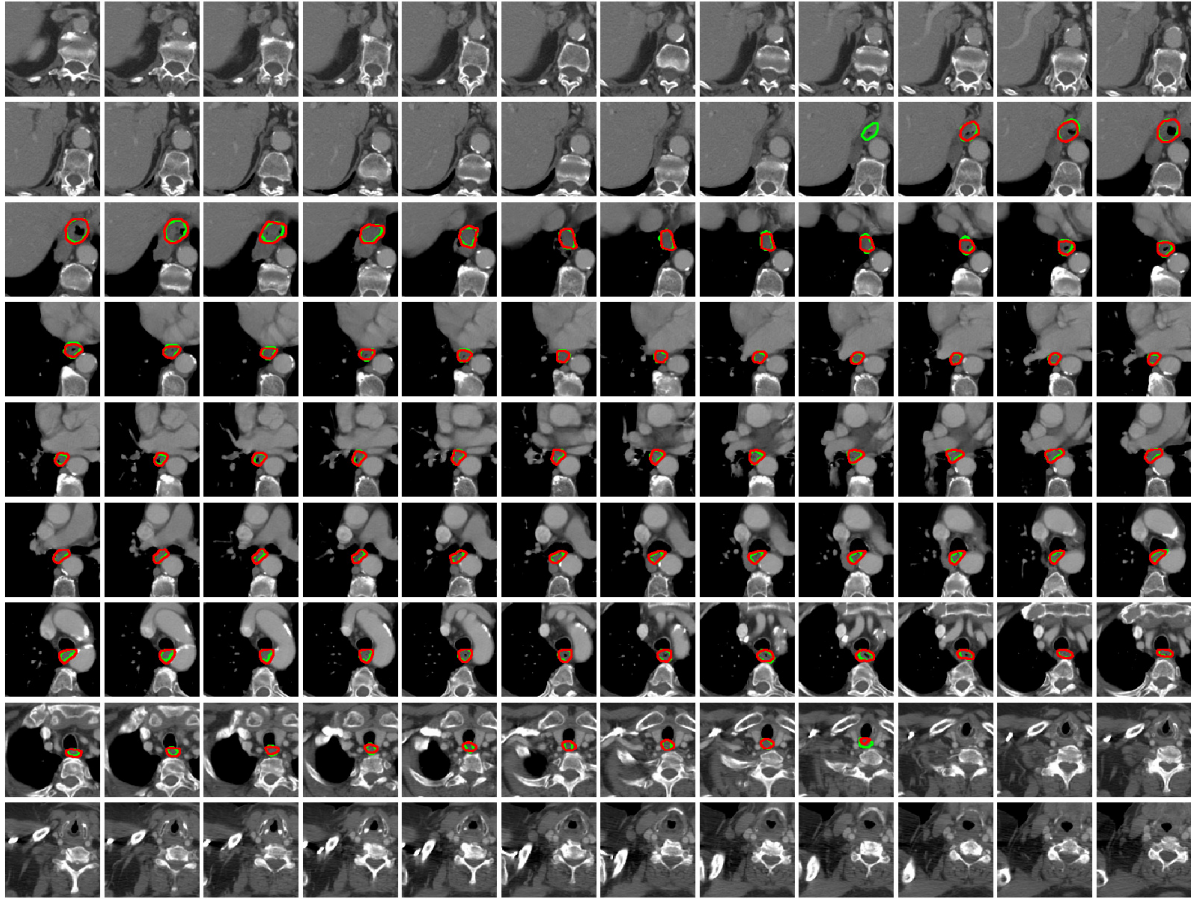


Figura 43 – Estudo de caso - Instituto 3. Em verde a marcação do especialista e em vermelho a segmentação do esôfago predita pelo método.

Novamente, é possível observar uma excelente segmentação tridimensional quando comparado a marcação do especialista (Dice de 82,01).

5.6 Comparação com os trabalhos relacionados a segmentação de esôfago

A literatura apresenta vários métodos de segmentação de esôfago em TC, seja como OAR ou não. A maioria dos autores relatam a dificuldade de distinção do esôfago em TC, até mesmo quando se trata de especialista. Destaca-se que, por não haver um contraste bem definido deste órgão em TC, acaba-se por se tornar uma tarefa árdua a segmentação do mesmo e, conseqüentemente, torna-se um processo exaustivo a detecção/segmentação manual.

Em Feulner et al. (2011), é proposto um método probabilístico para a segmentação do esôfago. Os experimentos dos autores são executados em uma base particular com 144

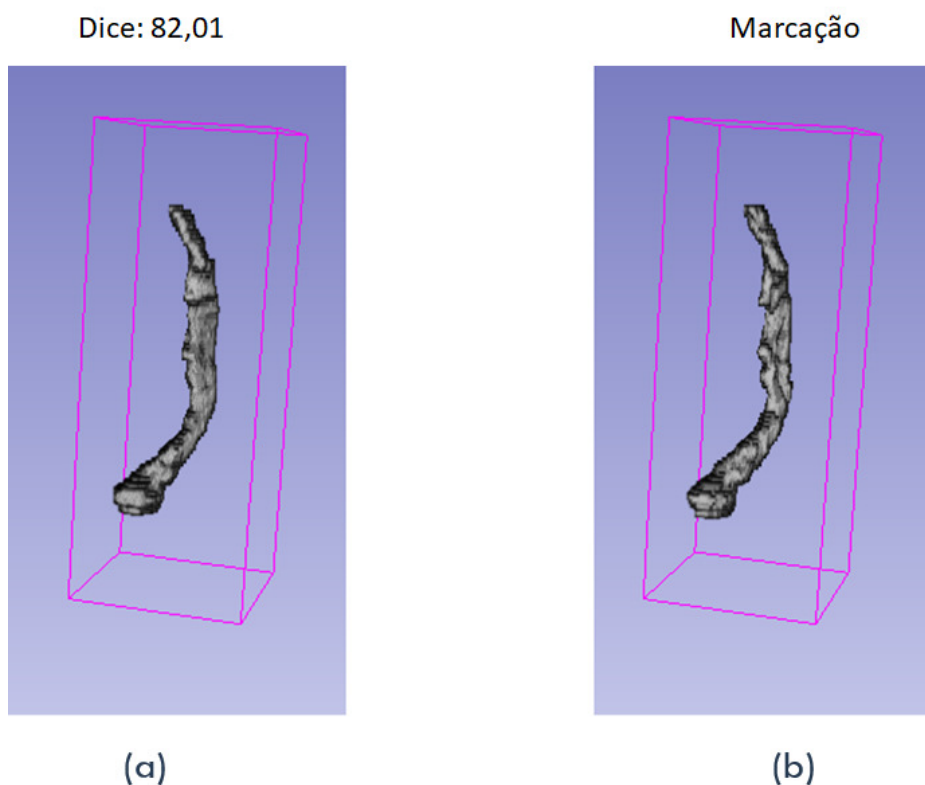


Figura 44 – Estudo de caso - Instituto 3. (a) volume predito pelo método e (b) volume da marcação do especialista.

CT e apresentam um resultado de 74% de Dice. Percebe-se que este método é altamente dependente dos modelos descritivos e do conhecimento prévio, ainda, é necessário um registro não rígido, causando deformidades na estrutura inicial do esôfago. Um dos principais pontos negativos do trabalho é o uso dos registros deformáveis que acabam por modificar a estrutura anatômica individual de cada paciente. Além disto, a segmentação é guiada por um modelo probabilístico. Nosso trabalho, apesar da utilização do registro, este é rígido e o atlas utilizado apenas é para geração de uma VOI que será posteriormente apresentada para a ResU-Net.

O trabalho proposto por Grosgeorge et al. (2013) é aplicado a uma base de imagens com 6 pacientes e apresenta Dice igual a 61%. Apesar de ser dito um método automático, o mesmo depende da criação de um modelo de esqueleto a partir da marcação de especialista. O método também utiliza apenas técnicas de processamento de imagens sem aplicar uma classificação e é aplicado em 6 pacientes. Nosso trabalho utiliza uma base de imagens muito diversificada com o total de 36 pacientes. Ainda, nosso trabalho alcança Dice de 82%.

O trabalho proposto por Larsson, Zhang e Kahl (2017) apresenta um Dice de 66,2%. Vale ressaltar, que o trabalho utiliza o atlas como um segmentador inicial, definindo a probabilidade de um voxel ser ou não um determinado órgão se este estiver “casado” com o atlas. Em seguida, a CNN se encarrega de dizer se esses voxels são ou não o determinado órgão. Neste tipo de abordagem, além de não classificar o volume inteiro, a classificação dos voxels é completamente dependente do atlas. Em nosso método, foi utilizada uma rede própria de segmentação U-Net com blocos residuais, e alcançou resultados mais expressivos do que o trabalho mencionado.

Trullo et al. (2017b) publicou um primeiro trabalho apresentando Dice de 66% para o esôfago. Em um segundo trabalho, Trullo et al. (2017a)) extrai uma *bounding box* da região do esôfago encontrado pelo primeiro trabalho, e a mesma rede é treinada e classifica apenas essa pequena região. O método foi testado em 30 CT e apresentou Dice de 72%. Ao observar o fato de que a primeira rede possa ignorar a localização do esôfago, ou marcá-lo em região errada, a segunda rede também não será capaz de segmentá-la. No nosso trabalho, para a segmentação da VOI primeiramente utilizou-se o atlas, em seguida, a ResU-Net classifica as regiões da VOI em esôfago, desta forma nosso trabalho supera os valores de Dice de ambos os trabalhos.

Fechter et al. (2017) propuseram uma CNN para elaboração de um modelo probabilístico que em seguida é utilizado como entrada para um modelo de contorno ativo alcançando resultado de 76% de Dice. Apesar do método utilizar CNN como parte de suas etapas, a segmentação final depende da escala de Hounsfield da imagem, isso torna o modelo suscetível a erros quando há uma mudança brusca de HU em casos que o processo de aquisição das TCs não seja bem definido, como é o caso da base de imagens utilizada no nosso trabalho, está possui diferentes tipos de aquisição. Apesar deste método apresentar o melhor valor de Dice da literatura, nosso trabalho também o supera neste quesito.

Em Dong et al. (2019), os autores propõem o uso de uma rede adversária juntamente com uma rede convolucional completamente conectada para alcançar seus resultados. Para o esôfago, o método apresenta 75% de Dice. O trabalho se mostra bastante promissor, contudo, a necessidade de uma segmentação precisa do pulmão é necessária para a extração do volume de interesse onde possa estar o esôfago. Este trabalho apresenta os resultados na mesma base de imagens utilizada por nós, contudo, o método é validado pela técnica de leave-one-out, isto garante que apenas um indivíduo é testado por vez. Logo, o treinamento tem mais amostrar para ser gerado. Ainda, o método é dependente de uma boa segmentação

dos pulmões, o que no nosso caso não interfere. O trabalho foi executado na mesma base de imagens utilizada em nosso trabalho, contudo os autores utilizam apenas 35 volumes, mas não afirmam o porquê disto. Nosso método supera este trabalho em todas as métricas de validação.

O trabalho proposto por Chen et al. (2019) realizou uma segmentação semântica do esôfago usando um U-Net. Seu banco de dados contém 15 TCs, totalizando mais de duas mil fatias. O resultado encontrado é um valor médio de dados de 79%. Apesar de ser um trabalho recente e de utilizar uma rede bastante conhecida, só se aplica a 15 exames. Por sua vez, Feng et al. (2019) apresenta um método de segmentação usando abordagens de CNN para vários órgãos. Para o esôfago, o método alcança um valor Dice de 72%. Neste trabalho, observa-se a importância dos cortes nas regiões de interesse do volume para que as segmentações sejam mais precisas. No entanto, o método tem baixo valor de Dice em comparação com outros estudos da literatura.

Lou et al. (2020) também propõe um método de segmentação de esôfago em TC. O método faz um estudo do uso de U-Net e suas variações para esta tarefa. O método usa técnicas de aumento de dados para contornar o problema de pouca amostragem da base de imagens. Após o uso das U-Net, é feito um refinamento da segmentação através de processos semi automáticos para criar a marcação manual. Por fim, é comparado a segmentação da rede com a marcação semi automática e extraídas as métricas de validação, o resultado do Dice foi de 86.80% em 17 pacientes. Apesar da métrica relevante, o trabalho compara a segmentação do método com seu próprio resultado ajustado pelo especialista.

A Tabela 8 mostra a relação entre esses trabalhos e o método proposto.

Tabela 8 – Comparação com os trabalhos relacionados.

SEN - sensibilidade; SPE - especificidade; ACC - acurácia; Pt - número de pacientes.

Trabalhos	Base de imagens	Pt	SEN	SPE	ACC	Dice
Feulner et al. (2011)	Privada	144	—	—	—	74%
Grosgeorge et al. (2013)	Privada	6	—	—	—	61%
Larsson, Zhang e Kahl (2017)	MICCAI 2015	30	—	—	—	66,2%
Trullo et al. (2017b)	MICCAI 2015	30	—	—	—	66%
Trullo et al. (2017a)	MICCAI 2015	30	—	—	—	72%
Fechter et al. (2017)	MICCAI 2015	50	—	—	—	76%
Dong et al. (2019)	AAPM Thoracic Auto-segmentation	35	73%	99%	—	75%
Chen et al. (2019)	Privada	15	—	—	—	79%
Feng et al. (2019)	AAPM Thoracic Auto-segmentation	36	—	—	—	72%
Lou et al. (2020)	Privada	17	—	—	—	86,80%
Método proposto (ResU-Net)	AAPM Thoracic Auto-segmentation	36	90,61%	99,76%	99,69%	82,15%

Pode-se observar que a literatura apresenta diversos trabalhos recentes para segmentação de esôfago. Por não ser uma tarefa trivial, nem mesmo para os médicos especialista, o maior valor encontrado na literatura até então é de 76% de Dice. Todos os trabalhos mostram a necessidade de uma redução do escopo do volume do paciente para a segmentação. Esta informação foi valiosa ao construir nosso método, onde utilizou-se de técnicas de registro e atlas para fazer a segmentação da VOI.

Também foi possível observar que nenhum destes trabalhos fazem melhoramento na imagem, então uma etapa crucial do nosso método foi encontrar técnicas de realce que pudesse melhorar a visualização do esôfago nas imagens. Outro ponto importante é o fato que após a classificação da rede, por ser um OAR muito similar em textura aos demais órgãos e tecidos, fez-se necessário uma etapa de refinamento, que gerou resultados bastante promissores.

Assim, ressalta-se que nosso trabalho apresentou um método automático para segmentação do esôfago. Este é composto por técnicas observadas na literatura e por técnicas desenvolvidas por nós para garantir melhores resultados. Apesar das bases de dados utilizadas serem diferentes da empregada em nosso trabalho, ele consegue superar quase todos os trabalhos apresentados até então na literatura. Com isso, demonstra-se a viabilidade de utilizar o método propostos na tarefa de segmentação de esôfago.

6 Conclusão

A segmentação de órgãos de risco é uma importante etapa no planejamento da radioterapia. Esta segmentação é feita em um exame tridimensional, comumente conhecido como tomografia de planejamento. Muitos pesquisadores são instigados a propor métodos para segmentação destes OARs. Neste trabalho, foram propostos dois métodos de segmentação de medula espinhal e um método de segmentação de esôfago como OARs em TC de planejamento.

6.1 Segmentação de medula espinhal

Para os dois métodos de segmentação de medula espinhal, utilizou-se da técnica de *template matching* para fazer uma segmentação inicial da região da medula espinhal, diminuindo o escopo para os algoritmos de classificação. No primeiro método, esta técnica ajudou tanto nas outras etapas quanto no custo computacional de não ter que executar a técnica IMSLIC em todo o volume. No segundo método, a técnica de *template matching* auxiliou reduzir as imagens de entrada das U-Nets e, consequentemente, o custo computacional.

Para o primeiro método, o uso do IMSLIC provou ser muito promissor para ajudar na geração de imagens de entrada da CNN. O IMSLIC foi utilizado pela primeira vez no contexto de imagens médicas e, mesmo sem utilizar uma técnica de redução de falso positivo, foi possível obter um excelente resultado apenas com essas imagens clusterizadas. Também foi proposto o uso da CNN para tarefa de classificação. O uso da CNN mostrou-se bastante eficaz. Mesmo com uma arquitetura simples, foi possível detectar cerca de 89,23% da medula espinhal. As outras métricas também mostraram quão boa a CNN é na tarefa de classificação.

Para o segundo método, propor e usar três arquiteturas de U-Nets compostas pelos principais parâmetros e inovações já encontrados na literatura, como blocos residuais e camadas densas, tornam o método ainda mais promissor. As três redes propostas alcançaram métricas de Dice próximas a 80% em uma base de dados completamente diversificada. E, ao analisar outras métricas, o método é ainda mais robusto, atingindo quase 100% em medidas como acurácia e especificidade.

Para melhoramento dos dois métodos, sugere-se a busca por algoritmos que consigam automatizar os parâmetros, tanto da CNN quanto das U-Nets, dessa forma, podendo apresentar melhores resultados. Ainda, o teste em outras bases de dados poderá validar ainda mais os métodos.

6.2 Segmentação do esôfago

Para a segmentação do esôfago, buscou-se associar as técnicas mais utilizadas na literatura e propor melhorias nas existentes de modo a gerar resultados promissores. A etapa de segmentação de ROI foi crucial para os resultados iniciais, com ela foi possível reduzir o escopo do volume e desta forma propiciar melhoras nos resultados. A etapa de pré-processamento se mostrou fundamental, com ela foi possível realçar as paredes do esôfago e, conseqüentemente, passar informações valiosas para a ResU-Net.

A rede proposta para a segmentação também foi promissora, uma vez que o esôfago ocupa uma pequena região da VOI, as camadas de subamostragem tendem a reduzir ainda mais esse OAR, logo, aliar informações residuais das camadas anteriores trouxe melhores resultados ao método.

Outro ponto importante foi a etapa de refinamento. Esta etapa foi capaz de tratar os problemas observados com a similaridade da textura dos tecidos em volta do esôfago, tanto a etapa de manter o maior objeto quanto a erosão aplicada foram importantes para alcançar o resultado de 82,15%.

Um ponto importante a se ressaltar é que nos trabalhos encontrados na literatura, este foi o melhor resultado de Dice, isto em uma base de dados completamente diversificada, o que valida ainda mais a eficácia do método proposto.

Porém, como já mencionado, a etapa de treinamento conta com uma rede de segmentação composta por inúmeros parâmetros e camadas, esses parâmetros são ajustados na fase de validação, o que toma bastante tempo. Automatizar essa etapa poderia produzir melhores resultados. Observa-se uma grande diferenciação entre os 3 institutos utilizados, uma forma de padronizar a textura também poderia melhorar os resultados.

Assim, destaca-se que os métodos apresentados são bastante promissores. Acredita-se que tais métodos possa ser úteis se empregada em grandes centros de radioterapia, onde a tarefa de proteção da medula espinhal e esôfago são essenciais. Deve-se notar que os danos à medula espinhal podem ser irreversíveis, de modo que os métodos computacionais,

como o proposto, são de extrema ajuda para o especialista. Para o esôfago, além de ser um órgão extremamente sensível a radiação, é um órgão de difícil segmentação até mesmo pelos médicos especialistas.

6.3 Comprovação das hipóteses

Com o método proposto e os resultados obtidos, foi possível alcançar as hipóteses propostas. De forma que:

- A utilização do *template matching* adaptativo nos dois métodos para medula espinhal foram promissores, reduzindo a região de interesse a ser analisada no volume dos pacientes. No esôfago, a utilização do atlas para corte do volume resultando apenas na VOI também apresentou melhora, tanto para análise do volume, quanto pelo custo computacional;
- Foi possível criar uma técnica automática para ser segmentação inicial nos dois métodos de medula espinhal, ambas técnicas utilizam *template matching* adaptativo e produziram resultados satisfatórios;
- A utilização do registro juntamente com o atlas permitiu uma padronização dos volumes evitando problemas causados pelo processo de aquisição de imagens;
- O esôfago é um OAR que não apresenta contraste em relação aos tecidos circundantes nas TCs. Desta forma, as técnicas de processamento de imagens para realce (filtro bilateral e equalização do histograma) foram eficientes em destacar essa região e produzir melhores resultados para a rede de segmentação;
- Tanto na medula espinhal quanto no esôfago, as técnicas de aprendizado de máquina (CNN e U-Net) foram eficientes na segmentação destes OARs, mostrando local de destaque na literatura;
- Ao utilizar técnicas de processamento de imagens na fase refinamento da segmentação, foi possível produzir resultados promissores na segmentação do esôfago, comprovando a importância desta etapa para a segmentação.

6.4 Produções científicas

Durante o desenvolvimento dos três métodos, foi possível alcançar uma série de contribuições no meio científico, principalmente na confecção de artigos, seja como autor ou como coautor. Diretamente ligado aos métodos apresentados neste trabalho, foram produzidos cinco artigos científicos destacados na Tabela 9.

Tabela 9 – Produções científicas como autor relacionadas ao método proposto.

Paper	Tipo	Status	Qualis	Ano
Spinal cord detection in planning CT for radiotherapy through adaptive template matching, IMSLIC and convolutional neural networks. Em: Computer Methods and Programs in Biomedicine	Journal	Publicado	A1	2019
Automatic spinal cord segmentation as organ at risk in planning CT using adaptive template matching and U-Net. Em: Escola Regional de Computação Aplicada à Saúde	Regional	Publicado	—	2019
Spinal cord segmentation as OAR in planning CT for radiotherapy using histogram matching, template matching, and U-Net. Em: Revista de Sistemas e Computação	Journal	Publicado	B5	2020
Automatic segmentation of spinal cord in planning computed tomography for radiotherapy using U-Net approaches. Em: Computer Methods and Programs in Biomedicine	Journal	Em revisão	A1	2019
Esophagus segmentation from planning CT images using an atlas-based deep learning approach. Em: Computer Methods and Programs in Biomedicine	Journal	Publicado	A1	2020

Ainda, foram produzidos artigos em vários campos, resultando em 1 congresso regional, 12 congressos nacionais, 5 congressos internacionais, 1 capítulo de livro, 1 revista nacional e 17 jornais internacionais.

6.5 Reconhecimento científico

Por fim, em reconhecimento a contribuição científica do trabalho intitulado *Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks* publicado na revista *Computer Methods and Programs in Biomedicine* (DINIZ et al., 2018), este foi considerado pela *International Medical Informatics Association Yearbook editorial board* um dos quatro trabalhos mais relevantes no meio científico (em uma seleção mundial de 1.459 métodos publicados no ano de 2018) no campo de *Sensor, Signal and Imaging Informatics* (HSU et al., 2019).

Referências

- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSTRUNK, S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 34, n. 11, p. 2274–2282, 2012.
- ADEBAHR, S.; SCHIMEK-JASCH, T.; NESTLE, U.; BRUNNER, T. B. Oesophagus side effects related to the treatment of oesophageal cancer or radiotherapy of other thoracic malignancies. *Best Practice & Research Clinical Gastroenterology*, Elsevier, v. 30, n. 4, p. 565–580, 2016.
- ARCHIP, N.; ERARD, P.-J.; EGMONT-PETERSEN, M.; HAEFLIGER, J.-M.; GERMOND, J.-F. A knowledge-based approach to automatic detection of the spinal cord in ct images. *IEEE Transactions on medical imaging*, IEEE, v. 21, n. 12, p. 1504–1516, 2002.
- BANIK, S.; RANGAYYAN, R. M.; BOAG, G. S. Automatic segmentation of the ribs, the vertebral column, and the spinal canal in pediatric computed tomographic images. *Journal of digital imaging*, Springer, v. 23, n. 3, p. 301–322, 2010.
- BENGIO, Y. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, Now Publishers Inc., v. 2, n. 1, p. 1–127, 2009.
- BENGIO, Y.; LAMBLIN, P.; POPOVICI, D.; LAROCHELLE, H. et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, MIT; 1998, v. 19, p. 153, 2007.
- BHATTACHARYYA, S. Spinal cord disorders: Myelopathy. *The American Journal of Medicine*, Elsevier, 2018.
- BROOKS, R. A. A quantitative theory of the hounsfield unit and its application to dual energy scanning. *Journal of computer assisted tomography*, v. 1, n. 4, p. 487–493, 1977.
- BROSCH, T.; YOO, Y.; LI, D. K.; TRABOULSEE, A.; TAM, R. C. Modeling the variability in brain morphology and lesion distribution in multiple sclerosis by deep learning. In: *MICCAI (2)*. [S.l.: s.n.], 2014. p. 462–469.
- BROWN, L. G. A survey of image registration techniques. *ACM computing surveys (CSUR)*, ACM, v. 24, n. 4, p. 325–376, 1992.
- BRUNELLI, R. *Template matching techniques in computer vision: theory and practice*. [S.l.]: John Wiley & Sons, 2009.
- BUTTON, M.; STAFFURTH, J. Clinical application of image-guided radiotherapy in bladder and prostate cancer. *Clinical Oncology*, Elsevier, v. 22, n. 8, p. 698–706, 2010.
- CASTANEDA, S. A.; ROMAK, L. B. Radiotherapy for anal cancer: Intensity-modulated radiotherapy and future directions. *Surgical Oncology Clinics*, Elsevier, v. 26, n. 3, p. 467–475, 2017.
- CAUDELL, J. J.; TORRES-ROCA, J. F.; GILLIES, R. J.; ENDERLING, H.; KIM, S.; RISHI, A.; MOROS, E. G.; HARRISON, L. B. The future of personalised radiotherapy for head and neck cancer. *The Lancet Oncology*, Elsevier, v. 18, n. 5, p. e266–e273, 2017.

- CHEN, M.; CARASS, A.; OH, J.; NAIR, G.; PHAM, D. L.; REICH, D. S.; PRINCE, J. L. Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. *Neuroimage*, Elsevier, v. 83, p. 1051–1062, 2013.
- CHEN, S.; YANG, H.; FU, J.; MEI, W.; REN, S.; LIU, Y.; ZHU, Z.; LIU, L.; LI, H.; CHEN, H. U-net plus: Deep semantic segmentation for esophagus and esophageal cancer in computed tomography images. *IEEE Access*, IEEE, v. 7, p. 82867–82877, 2019.
- CHOLLET, F. et al. Keras: The python deep learning library. *Astrophysics Source Code Library*, 2018.
- CIREŞAN, D.; MEIER, U.; MASCI, J.; SCHMIDHUBER, J. A committee of neural networks for traffic sign classification. In: IEEE. *Neural Networks (IJCNN), The 2011 International Joint Conference on*. [S.l.], 2011. p. 1918–1921.
- CRUM, W. R.; HARTKENS, T.; HILL, D. Non-rigid image registration: theory and practice. *The British Journal of Radiology*, British Institute of Radiology, 2014.
- DINIZ, J. O. B.; DINIZ, P. H. B.; VALENTE, T. L. A.; SILVA, A. C.; PAIVA, A. C. de; GATTASS, M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 156, p. 191–207, 2018.
- DINIZ, J. O. B.; DINIZ, P. H. B.; VALENTE, T. L. A.; SILVA, A. C.; PAIVA, A. C. Spinal cord detection in planning ct for radiotherapy through adaptive template matching, inslic and convolutional neural networks. *Computer methods and programs in biomedicine*, Elsevier, v. 170, p. 53–67, 2019.
- DINIZ, P. H. B.; VALENTE, T. L. A.; DINIZ, J. O. B.; SILVA, A. C.; GATTASS, M.; VENTURA, N.; MUNIZ, B. C.; GASPARETTO, E. L. Detection of white matter lesion regions in mri using slic0 and convolutional neural network. *Computer methods and programs in biomedicine*, Elsevier, v. 167, p. 49–63, 2018.
- DONG, X.; LEI, Y.; WANG, T.; THOMAS, M.; TANG, L.; CURRAN, W. J.; LIU, T.; YANG, X. Automatic multiorgan segmentation in thorax ct images using u-net-gan. *Medical physics*, Wiley Online Library, v. 46, n. 5, p. 2157–2168, 2019.
- DOWLATI, E. Spinal cord anatomy, pain, and spinal cord stimulation mechanisms. In: ELSEVIER. *Seminars in Spine Surgery*. [S.l.], 2017. v. 29, p. 136–146.
- EVANS, E.; STAFFURTH, J. Principles of cancer treatment by radiotherapy. *Surgery-Oxford International Edition*, Elsevier, v. 36, n. 3, p. 111–116, 2018.
- EVANS, L.; SPRUCK, J. Motion of level sets by mean curvature i. *preprint*, 1992.
- FECHTER, T.; ADEBAHR, S.; BALTAS, D.; AYED, I. B.; DESROSIERS, C.; DOLZ, J. Esophagus segmentation in ct via 3d fully convolutional neural network and random walk. *Medical physics*, Wiley Online Library, v. 44, n. 12, p. 6341–6352, 2017.
- FENG, X.; QING, K.; TUSTISON, N. J.; MEYER, C. H.; CHEN, Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3d images. *Medical physics*, Wiley Online Library, v. 46, n. 5, p. 2169–2180, 2019.

- FERNANDES, B. J. T. Redes neurais com extração implícita de características para reconhecimento de padrões visuais. Universidade Federal de Pernambuco, 2013.
- FEULNER, J.; ZHOU, S. K.; HAMMON, M.; SEIFERT, S.; HUBER, M.; COMANICIU, D.; HORNEGGER, J.; CAVALLARO, A. A probabilistic model for automatic segmentation of the esophagus in 3-d ct scans. *IEEE transactions on medical imaging*, IEEE, v. 30, n. 6, p. 1252–1264, 2011.
- GODOY, D. *Understanding binary cross-entropy/log loss: a visual explanation*. [S.l.]: Látogatva, 2019.
- GÓMEZ-MILLÁN, J.; LARA, M. F.; GENEROSO, R. C.; PEREZ-ROZOS, A.; LUPIÁÑEZ-PÉREZ, Y.; CARMONA, J. A. M. Advances in the treatment of prostate cancer with radiotherapy. *Critical reviews in oncology/hematology*, Elsevier, v. 95, n. 2, p. 144–153, 2015.
- GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. [S.l.]: Prentice hall Upper Saddle River, 2002.
- GONZALEZ, R. C.; WOODS, R. E. *Pocessamento digital de imagens*. [S.l.]: Pearson Prentice Hall, 2010.
- GROSGEORGE, D.; PETITJEAN, C.; DUBRAY, B.; RUAN, S. Esophagus segmentation from 3d ct data using skeleton prior-based graph cut. *Computational and mathematical methods in medicine*, Hindawi, v. 2013, 2013.
- GUO, Y.; LIU, Y.; OERLEMANS, A.; LAO, S.; WU, S.; LEW, M. S. Deep learning for visual understanding: A review. *Neurocomputing*, Elsevier, v. 187, p. 27–48, 2016.
- GWYNNE, S.; WEBSTER, R.; ADAMS, R.; MUKHERJEE, S.; COLES, B.; STAFFURTH, J. Image-guided radiotherapy for rectal cancer—a systematic review. *Clinical Oncology*, Elsevier, v. 24, n. 4, p. 250–260, 2012.
- HAFEMANN, L. G. An analysis of deep neural networks for texture classification. 2014.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- HORNE, J. C. V.; PARKER, G. G. The random-walk theory: an empirical test. *Financial Analysts Journal*, Taylor & Francis, v. 23, n. 6, p. 87–92, 1967.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural networks*, Elsevier, v. 2, n. 5, p. 359–366, 1989.
- HSU, W.; BAUMGARTNER, C.; DESERNO, T. et al. Advancing artificial intelligence in sensors, signals, and imaging informatics. *Yearbook of medical informatics*, Georg Thieme Verlag KG, v. 28, n. 01, p. 115–117, 2019.
- IBRAGIMOV, B.; XING, L. Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks. *Medical physics*, Wiley Online Library, v. 44, n. 2, p. 547–557, 2017.

ISHIYAMA, H.; SHUTO, N.; TERAZAKI, T.; NODA, S.; ISHIGAMI, M.; YOGO, K.; HAYAKAWA, K. Risk factors for radiotherapy incidents: a single institutional experience. *Medical Dosimetry*, Elsevier, 2018.

JOHNSON, H.; HARRIS, G.; WILLIAMS, K. et al. Brainsfit: mutual information rigid registrations of whole-brain 3d images, using the insight toolkit. *Insight J*, v. 57, n. 1, p. 1–10, 2007.

KHAN, J. *Anatomic Basis of Neurologic Diagnosis*. [S.l.]: AAN Enterprises, 2010.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: *International Conference on Learning Representations (ICLR)*. [S.l.: s.n.], 2015.

KOLAŘÍK, M.; BURGET, R.; UHER, V.; ŘÍHA, K.; DUTTA, M. K. Optimized high resolution 3d dense-u-net network for brain and spine segmentation. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 9, n. 3, p. 404, 2019.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

LAAN, J.; LONKHUIJZEN, L. van; OS, R. van; TYTGAT, K.; FAJARDO, R. D.; PIETERS, B.; STALPERS, L.; WESTERVELD, G. Socioeconomic status as an independent risk factor for severe late bowel toxicity after primary radiotherapy for cervical cancer. *Gynecologic oncology*, Elsevier, v. 147, n. 3, p. 684–689, 2017.

LANDMAN, B.; XU, Z.; IGELSIAS, J.; STYNER, M.; LANGERAK, T.; KLEIN, A. Miccai multi-atlas labeling beyond the cranial vault—workshop and challenge. In: *Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge*. [S.l.: s.n.], 2015.

LAROCHELLE, H.; ERHAN, D.; COURVILLE, A.; BERGSTRA, J.; BENGIO, Y. An empirical evaluation of deep architectures on problems with many factors of variation. In: *ACM. Proceedings of the 24th international conference on Machine learning*. [S.l.], 2007. p. 473–480.

LARSSON, M.; ZHANG, Y.; KAHL, F. Robust abdominal organ segmentation using regional convolutional neural networks. In: SPRINGER. *Scandinavian Conference on Image Analysis*. [S.l.], 2017. p. 41–52.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–44, 05 2015.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.
- LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. et al. Convolutional networks and applications in vision. In: *ISCAS*. [S.l.: s.n.], 2010. p. 253–256.
- LEENER, B. D.; COHEN-ADAD, J.; KADOORY, S. Automatic segmentation of the spinal cord and spinal canal coupled with vertebral labeling. *IEEE transactions on medical imaging*, IEEE, v. 34, n. 8, p. 1705–1718, 2015.
- LI, R.; ZHANG, W.; SUK, H.-I.; WANG, L.; LI, J.; SHEN, D.; JI, S. Deep learning based imaging data completion for improved brain disease diagnosis. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2014. p. 305–312.
- LITJENS, G.; KOOL, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017.
- LIU, Y.-J.; YU, M.; LI, B.-J.; HE, Y. Intrinsic manifold slic: a simple and efficient method for computing content-sensitive superpixels. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 3, p. 653–666, 2018.
- LIU, Z.; LIU, X.; XIAO, B.; WANG, S.; MIAO, Z.; SUN, Y.; ZHANG, F. Segmentation of organs-at-risk in cervical cancer ct images with a convolutional neural network. *Physica Medica*, Elsevier, v. 69, p. 184–191, 2020.
- 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.
- LOU, X.; ZHU, Y.; PUNITHAKUMAR, K.; LE, L. H.; LI, B. Esophagus segmentation in computed tomography images using a u-net neural network with a semiautomatic labeling method. *IEEE Access*, IEEE, v. 8, p. 202459–202468, 2020.
- MACCHIA, M. L.; FELLIN, F.; AMICHETTI, M.; CIANCHETTI, M.; GIANOLINI, S.; PAOLA, V.; LOMAX, A. J.; WIDESOTT, L. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiation Oncology*, Springer, v. 7, n. 1, p. 160, 2012.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.
- MCCUNNIFF, A. J.; LIANG, M. J. Radiation tolerance of the cervical spinal cord. *International Journal of Radiation Oncology* Biology* Physics*, Elsevier, v. 16, n. 3, p. 675–678, 1989.
- MCVICAR, N.; POPESCU, I. A.; HEATH, E. Techniques for adaptive prostate radiotherapy. *Physica Medica: European Journal of Medical Physics*, Elsevier, v. 32, n. 3, p. 492–498, 2016.

MILOSEVIC, M.; GOSPODAROWICZ, M.; ZIETMAN, A.; ABBAS, F.; HAUSTERMANS, K.; MOONEN, L.; RÖDEL, C.; SCHOENBERG, M.; SHIPLEY, W. Radiotherapy for bladder cancer. *Urology*, Elsevier, v. 69, n. 1, p. 80–92, 2007.

MINSKY, B. D.; PAJAK, T. F.; GINSBERG, R. J.; PISANSKY, T. M.; MARTENSON, J.; KOMAKI, R.; OKAWARA, G.; ROSENTHAL, S. A.; KELSEN, D. P. Int 0123 (radiation therapy oncology group 94-05) phase iii trial of combined-modality therapy for esophageal cancer: high-dose versus standard-dose radiation therapy. *Journal of clinical oncology*, v. 20, n. 5, p. 1167–1174, 2002.

NETTO, S. M. B. et al. Métodos computacionais para identificação, quantificação e análise de mudanças no tecido da lesão pulmonar através de imagens de tomografia computadorizada. Universidade Federal do Maranhão, 2016.

RANGAYYAN, R. M.; DEGLINT, H. J.; BOAG, G. S. Method for the automatic detection and segmentation of the spinal canal in computed tomographic images. *Journal of Electronic Imaging*, International Society for Optics and Photonics, v. 15, n. 3, p. 033007, 2006.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241.

ROTH, H. R.; LU, L.; SEFF, A.; CHERRY, K. M.; HOFFMAN, J.; WANG, S.; LIU, J.; TURKBAY, E.; SUMMERS, R. M. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2014. p. 520–527.

RUPPERT, B. N.; WATKINS, J. M.; SHIRAI, K.; WAHLQUIST, A. E.; GARRETT-MAYER, E.; AGUERO, E. G.; SHERMAN, C. A.; REED, C. E.; SHARMA, A. K. Cisplatin/irinotecan versus carboplatin/paclitaxel as definitive chemoradiotherapy for locoregionally advanced esophageal cancer. *American journal of clinical oncology*, LWW, v. 33, n. 4, p. 346–352, 2010.

SILVA, A. C. Algoritmos para diagnostico assistido de nodulos pulmonares solitarios em imagens de tomografia computadorizada. *Rio de Janeiro: PUC*, v. 1, n. 1, 2004.

SILVA, G. L. da; NETO, O. P. da S.; SILVA, A. C.; PAIVA, A. C. de; GATTASS, M. Lung nodules diagnosis based on evolutionary convolutional neural network. *Multimedia Tools and Applications*, Springer, v. 76, n. 18, p. 19039–19055, 2017.

SUDRE, C. H.; LI, W.; VERCAUTEREN, T.; OURSELIN, S.; CARDOSO, M. J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. [S.l.]: Springer, 2017. p. 240–248.

TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, BioMed Central, v. 15, n. 1, p. 1–28, 2015.

- TAIGMAN, Y.; YANG, M.; RANZATO, M.; WOLF, L. Deepface: Closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 1701–1708.
- TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. In: IEEE. *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. [S.l.], 1998. p. 839–846.
- TOMPSON, J.; GOROSHIN, R.; JAIN, A.; LECUN, Y.; BREGLER, C. Efficient object localization using convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2015. p. 648–656.
- TONG, T.; LI, G.; LIU, X.; GAO, Q. Image super-resolution using dense skip connections. In: *Proceedings of the IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2017. p. 4799–4807.
- TRULLO, R.; PETITJEAN, C.; NIE, D.; SHEN, D.; RUAN, S. Fully automated esophagus segmentation with a hierarchical deep learning approach. In: IEEE. *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. [S.l.], 2017. p. 503–506.
- TRULLO, R.; PETITJEAN, C.; RUAN, S.; DUBRAY, B.; NIE, D.; SHEN, D. Segmentation of organs at risk in thoracic ct images using a sharpmask architecture and conditional random fields. In: IEEE. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. [S.l.], 2017. p. 1003–1006.
- TSANG, Y. M.; HOSKIN, P. The impact of bladder preparation protocols on post treatment toxicity in radiotherapy for localised prostate cancer patients. *Technical Innovations and Patient Support in Radiation Oncology*, Elsevier, v. 3, p. 37–40, 2017.
- VERHAART, R. F.; FORTUNATI, V.; VERDUIJN, G. M.; WALSUM, T. van; VEENLAND, J. F.; PAULIDES, M. M. Ct-based patient modeling for head and neck hyperthermia treatment planning: Manual versus automatic normal-tissue-segmentation. *Radiotherapy and Oncology*, Elsevier, v. 111, n. 1, p. 158–163, 2014.
- WANG, T.; WU, D. J.; COATES, A.; NG, A. Y. End-to-end text recognition with convolutional neural networks. In: IEEE. *Pattern Recognition (ICPR), 2012 21st International Conference on*. [S.l.], 2012. p. 3304–3308.
- WANG, Z.; WANG, E.; ZHU, Y. Image segmentation evaluation: a survey of methods. *Artificial Intelligence Review*, Springer, v. 53, n. 8, p. 5637–5674, 2020.
- WERNER-WASIK, M.; YORKE, E.; DEASY, J.; NAM, J.; MARKS, L. B. Radiation dose-volume effects in the esophagus. *International Journal of Radiation Oncology* Biology* Physics*, Elsevier, v. 76, n. 3, p. S86–S93, 2010.
- YANG, J.; VEERARAGHAVAN, H.; III, S. G. A.; FARAHANI, K.; KIRBY, J. S.; KALPATHY-KRAMER, J.; ELMPT, W. van; DEKKER, A.; HAN, X.; FENG, X. et al. Autosegmentation for thoracic radiation treatment planning: a grand challenge at aapm 2017. *Medical physics*, Wiley Online Library, v. 45, n. 10, p. 4568–4581, 2018.
- ZEILER, M. D. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

ZELEFSKY, M. J.; FUKS, Z.; HAPPERSETT, L.; LEE, H. J.; LING, C. C.; BURMAN, C. M.; HUNT, M.; WOLFE, T.; VENKATRAMAN, E.; JACKSON, A. et al. Clinical experience with intensity modulated radiation therapy (imrt) in prostate cancer. *Radiotherapy and Oncology*, Elsevier, v. 55, n. 3, p. 241–249, 2000.