Automated Segmentation of Lymph Nodes on Neck CT scans using Deep Learning

Saba Ghazimoghadam
Department of Medicine
Division of Experimental Medicine
McGill University, Montreal
August 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Abstract

Cervical lymph node detection and accurate classification are essential for the staging and optimal management of patients with head and neck malignancies. While experts can accurately identify large abnormal lymph nodes, the sensitivity for detecting early-stage pathologic or metastatic lymph nodes, measuring 5-10 mm, remains suboptimal. Consequently, invasive neck dissections or sentinel node biopsies are often required for patients with head and neck cancers at high risk sites. To address this issue, there is growing interest in employing radiomic and artificial intelligence (AI) methodologies for precise diagnosis and classification of neck lymph nodes. A critical step in any classification task is object detection and, depending on the approach, lesion segmentation. Deep learning (DL) techniques have demonstrated exceptional performance in various image analysis tasks, including semantic segmentation and object detection. Nonetheless, lymph nodes present unique challenges due to their small size and limited representation in computed tomography (CT) scans. This thesis aims to develop a non-invasive algorithm for the detection and automatic segmentation of cervical lymph nodes on contrast-enhanced CT scans, paving the way for future diagnostic classification algorithms in research and clinical settings. To achieve this, we employed various DL approaches to detect and segment small lymph nodes, ranging from 5 to 10 mm in size. After evaluating multiple architectures, we attained a Dice

score of 0.8014, demonstrating that despite their small size, cervical lymph nodes can be detected and segmented automatically using deep learning techniques.

Résumé

La détection et la classification précise des ganglions lymphatiques cervicaux sont essentielles pour la stadification et la prise en charge optimale des patients atteints de tumeurs à la tête et au cou. Alors que les experts peuvent identifier avec précision les gros ganglions lymphatiques anormaux, la sensibilité pour détecter les ganglions lymphatiques pathologiques ou métastatiques à un stade précoce, mesurant 5 à 10 mm, reste sous-optimale.Il existe donc un intérêt grandissant pour l'utilisation d'approches radiomiques et de l'intelligence artificielle (AI) pour le diagnostic de précision et la classification des ganglions lymphatiques au cou. Cependant, une étape essentielle pour toute tâche de classification est la détection d'objets et, en fonectionde l'approche spécifique utilisée, la segmentation des lésions. Les techniques d'apprentissage en profondeur (DL) ont donné d'excellents résultats pour une variété de tâches d'analyse d'images, y compris la segmentation sémantique et la détection d'objets. Cependant, les ganglions lymphatiques posent des défis uniques parce que la majorité sont petits et constituent un très petit pourcentage de pixels ou de voxels globaux sur une tomodensitométrie. L'objectif de cette thèse est de construire un algorithme non-invasif visant la détection et la segmentation automatique des ganglions lymphatiques cervicaux dans les tomodensitométrie avec injection de contraste. Un tel outil pourrait alors constituer l'élément de base de futurs algorithmes de classification de diagnostique en recherche et, éventuellement, en clinique. À cette fin, nous avons

utilisé différentes techniques de DL pour la détection et la segmentation des ganglions lymphatiques, nous concentrant sur les petits ganglions lymphatiques mesurant de 5 à 10 mm. De plusieurs architectures fûrent évaluées, atteignant un coefficient de Dice de 0.8014. Grâce à diverses techniques d'apprentissage en profondeur, nous démontrons que malgré leur petite taille, la détection et la segmentation automatiques des ganglions lymphatiques cervicaux peuvent être effectuées en utilisant l'apprentissage en profondeur.

Acknowledgements

I want to express my sincere gratitude to my supervisor Dr. Reza Forghani for his tremendous mentorship, support, and invaluable guidance throughout my graduate studies. I have been fortunate to explore a wide range of fields under his supervision, which has enriched my academic journey. I hope to one day be able to dedicate and contribute to medicine as he does. I also appreciate my co-supervisor's support, Dr. Gerald Batist.

I would also like to thank my dear friend Dr. Padcha Tunlayadechanont, from whom I have learned much. Her patience and dedication to this project are worth a lot to me. Furthermore, I am grateful to her for our fantastic memories in Montreal.

I am deeply thankful to Mahfuz Al Hasan and Tahsin Mostafiz for their tremendous assistance and support in the completion of my thesis.

Also, I would like to thank Kevin Pierre and Manas Gupta for all their valuable contributions.

I am grateful to my dear friend Chantal for helping me write the French abstract of this thesis. Also, I am thankful to the members of AIPHL, especially Dr. Caroline Reinhold, Dr. Farhad Maleki and Ms. Rita Zakarian.

Lastly, I am extremely thankful to my family and friends for their continuous support and love, especially my parents and brother, whose countless sacrifices have been the foundation of my success.

Table of Contents

List of Figures

List of Tables

Contribution of Authors

Saba Ghazimoghadam conducted part of the literature review and contributed to the design of the study, working collaboratively with the multi-disciplinary team. She participated in the curation of data and annotated/segmented the objects of interest (lymph nodes), providing the data used for the training of the deep learning models. Other members of the laboratory team wrote the code for the experiments. Saba Ghazimoghadam contributed to the analysis of the data and preparation of the manuscript that will be submitted for publication.

List of Acronyms

2D	Two-dimensional
3D	Three-dimensional
AG	Attention Gate
AI	Artificial Intelligence
ASPP	Atrous Spatial Pyramid Pooling
BCE	Binary Cross Entropy
CECT	Contrast-enhanced CT
CNN	Convolutional Neural Networks
CT	Computed Tomography
DCAN	Deep Contour Aware Network
DCNN	Deep convolutional neural network
DL	Deep Learning
DICOM	Digital Imaging and Communications in Medicine
DNN	Deep Neural Network
DSC	Dice Similarity Coefficient
FCN	Fully Convolutional Network
FOV	Field-of-view
GAN	Generative Adversarial Network

GPU Graphics Processing Units

HNSCC Head and Neck Squamous Cell Carcinoma

IOU Intersection Over Union

LN Lymph Nodes

ML Machine learning

MLP Multilayer perceptrons

MRI Magnetic Resonance Imaging

OLNM Occult Lymph Node Metastases

OSCC Oral Squamous Cell Carcinoma

RECIST Response Evaluation Criteria in Solid Tumors

ReLU Rectified Linear Unit

SAD Short Axis Diameter

Chapter 1

Introduction

Head and neck cancers encompass primary tumors found in the lips and oral cavity, nasopharynx, oropharynx, hypopharynx, larynx, salivary glands, paranasal sinuses, in addition to non-mucosal cancers such as those arising from the thyroid gland and other less common sites (9). The Global Burden of Disease study revealed that 5.3% of all cancer-related deaths could be attributed to head and neck cancers. Between 1990 and 2017, incidence rates declined for larynx and nasopharyngeal cancers but increased for lip, oral cavity, and other pharyngeal cancers. The global burden of head and neck cancers is projected to rise for both men and women by 2030 (10, 11). Tobacco use and alcohol consumption are the leading risk factors (11).

Over 90% of mucosal head and neck cancers are classified as squamous cell carcinomas or related variants. These cancers exhibit aggressive behaviour and are prone to developing early cervical lymph node metastasis or late-stage distant metastasis, even with effective treatment (12).

1.1 Motivation and approach

Head and neck cancers frequently metastasize to regional lymph nodes (LN) through lymphatics, making routine cervical LN assessment essential (13). LN metastases significantly impact disease staging and treatment strategy. Patients with head and neck cancer and metastatic LNs may be treated with surgery, chemoradiation, or both. Nodal staging influences treatment planning, including the extent of neck dissection surgery and radiation therapy (14). Pathological cervical LNs are also crucial prognosticators for overall survival (15, 16). For the sample, bilateral or multiple metastatic neck nodes, as well as nodes contralateral to the primary tumour, significantly decrease patients' survival rates (17).

In current clinical practice, the evaluation of LNs on CT is based on 2-dimensional measurements. Despite significant advances in LN evaluation and classification, expert discrimination of abnormal from normal LNs is imperfect, and the accuracy can be even less when interpretation is performed by radiologists not subspecialized in head and neck imaging. Particularly, detecting early nodal metastases in small LNs measuring less than 1 cm remains a significant challenge (5, 14).

Managing patients without evidence of cervical LN metastasis upon clinical examination, or “clinically negative (N0) necks,” remains controversial due to the possibility of occult LN metastasis, which is undetectable radiologically and clinically. Neck dissections are routinely performed for patients with clinically N0 necks and high-risk tumours, but this may lead to overtreatment with potential for complications in many

cases (17, 18). Consequently, there is significant interest in predictive models to improve diagnostic accuracy and reduce unnecessary elective neck dissections (14, 19). Figure 1.1 demonstrates the anatomical stations of the neck LNs (3).

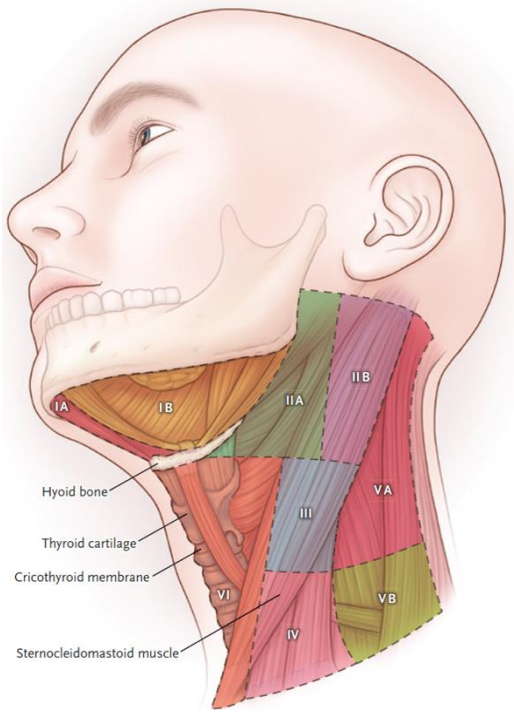

Figure 1.1. Lymph node Levels in the Neck. The metastatic nodes' presence and location can strongly affect the disease prognosis and potential therapeutics (3)

The reliability of a physical examination involving cervical palpation can be affected by factors such as cervical node size, subcutaneous fat presence, and the examiner's experience (18). Importantly, abnormal nodes located in deep nodal stations may not be

palpable on physical exam. Therefore, after conducting a comprehensive medical history and thorough physical examination, medical imaging is performed for a more accurate assessment, potentially upstaging the clinical assessment (3, 15, 19). Imaging assists in determining clinical staging and the extent of neck dissection before tumors resection surgery (20). Both contrast-enhanced CT (CECT) scans and magnetic resonance imaging (MRI) can be used for the evaluation of cervical LN in head and neck squamous cell carcinomas (HNSCCs) (18, 19). At many centers, CT is the first line modality for the evaluation of mucosal cancers below the level of the hard palate. CECT scans provide valuable quantitative information on LN sensity, shape, and texture (16). However, LN identification can be time-consuming and challenging, requiring specific expertise (13, 15). Importantly, the sensitivity for detection of early nodal metastasis falls significantly in small lymph nodes less than 1 cm on anatomic imaging. There is great interest in computer assisted evaluation of lymph nodes to assist diagnosis. However, the first step for such analysis would be LN segmentation. Manual segmentation of LNs in CT scans is complex and varies depending on the radiologist's experience, and would be prohibitively time consuming for routine radiological practice, highlighting the need for an automated detection and segmentation system in the medical image analysis (13, 21). Automated segmentation approaches can also be beneficial for therapy planning.

There has been a specific interest in using different radiomic or computer vision approaches, including deep learning (DL), to improve LN detection and classification

accuracy. However, due to the small dimensions of LNs, the primary emphasis of current approaches is on identifying and segmenting these objects. To achieve success, it is essential to automate these processes. This prerequisite is necessary for the integration of such tools into the busy clinical workflow, enabling their adoption in clinical practice and ultimately improving patient care. The automated segmentation of LNs using DL poses distinctive challenges due to their typically small size, occupying only a small fraction of pixels on a given scan. This difficulty is especially prominent for the nodes that have the potential for the most significant impact through machine-assisted classification – specifically, the small nodes measuring less than 1 cm, where expert evaluation tends to be less precise (13).

DL techniques have demonstrated outstanding performance in computer vision tasks, including semantic segmentation, object detection, and regression prediction, and have become popular for automated segmentation on medical images (22). Deep convolutional neural network (DCNN) algorithms have gained popularity for automatic segmentation in the medical images (21). However, its application in the evaluation of small LNs in head and neck cancer remains limited. (23, 24). This thesis aims to pioneer LN segmentation as a preliminary step in this field, with the objective of enhancing patient management and quality of life by minimizing the extent or necessity for elective neck dissections in head and neck cancer patients with a clinically N0 neck.

1.2 Outline of Thesis

To the best of our knowledge, no studies have implemented a model for the automatic segmentation of small non-metastatic cervical LNs in healthy individuals. Therefore, in this project, we aim to develop a non-invasive clinical tool for segmenting neck LNs in contrast-enhanced CT scans, enabling the extraction of high-level quantitative features in the future. We will employ DL models to create algorithms for this purpose.
Previous studies have typically considered LNs with a minimal axial diameter greater than 10 mm as abnormal, potentially overlooking metastases in LNs with a minimal axial diameter of less than 10 mm (13, 25). We have broadened our criteria to include LNs with a maximal axial diameter of 5 mm or larger to increase precision in detecting smaller LNs.
This thesis aims to apply advanced DL methods for LN detection and segmentation in contrast-enhanced neck CT scans and investigate state-of-the-art DL algorithms for LN segmentation. Evaluating the proposed objective will help highlight experimental research aspects in both medicine and computer science, ultimately leading to enhanced disease identification and improved health outcomes over time.

Given that cancer is a leading cause of death worldwide and its prevalence is steadily increasing, accurate identification and analysis of LNs are crucial for faster diagnosis and more precise treatment decision-making.

The structure of this thesis is as follows: Chapter 2 provides a comprehensive overview of the necessity of this work and reviews relevant literature in the field. Chapter 3 details the project's methodology, outlining the steps performed. Chapter 4 presents the visualization of experimental results and discussion. Lastly, Chapter 5 offers our conclusions and directions for future research related to this project.

1.3 Contribution

We aim to contribute to the fields of artificial intelligence (AI) and medical imaging analysis by demonstrating the promising results of our proposed algorithm in LN localization and segmentation. The main contributions of this thesis can be summarized as follows:

- Generating manual ground truth by annotating medical image data.
- Applying a preprocessing technique to the image dataset for effective and accurate analysis by the proposed DL model.
- Developing a DL-based model using CECT scan images to aid physicians and radiologists in rapidly diagnosing small LNs and facilitating accurate detection and segmentation of cervical LNs.
- Conducting a comparative performance analysis of our proposed model with other state-of-the-art methodologies, demonstrating that our algorithm can detect

and segment cervical LNs with high accuracy in contrast-enhanced CT scan datasets using metrics such as Dice score and Jaccard index.

Chapter 2

Background

This section discusses the significance of cervical LNs in head and neck cancers, the rationale for this project, and the various DL methods employed for LN segmentation. We will also cover computational image analysis, which consists of data preparation, data preprocessing, and DL model development.

2.1 Cervical Lymph nodes

In solid cancers, regional LN metastasis is an important prognostic indicator. Clinicians routinely evaluate LN shape, morphology, and size in malignancies to assess disease progression and determine therapeutic strategies (21, 26). In head and neck cancers, metastasis to cervical LNs has a significant impact on prognosis and treatment. LN metastasis is a crucial prognostic factor, and LN analysis plays an essential role in cancer staging and treatment effectiveness (15, 20, 21). Thus, accurate detection of metastatic LNs is important (26).

LN size, morphology, and functional metabolic activity help differentiate metastatic from reactive LNs (5). Diagnostic imaging is essential for head and neck cancer evaluation and staging, as it can identify metastatic LNs that are not clinically detectable (8).

Although comparative research is limited, CECT scans are often considered the first-line, optimal imaging modality due to their reliability and accessibility (8). Figure 2.1 demonstrates a CECT of a patient with a metastatic LN detected through this imaging technique.

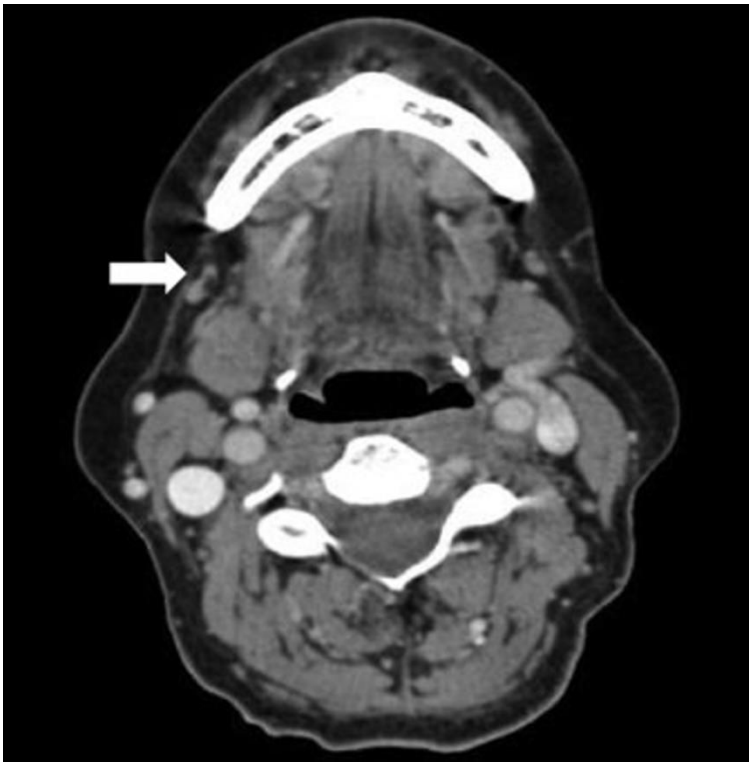

Figure 2.1. Imaging of a patient with a normal LN. The white arrow in CECT demonstrates an example of a normal level IB LN (8).

Imaging-based classification for cervical LNs is divided into seven levels, as illustrated in Figure 2.2. Detailed descriptions of the landmarks for each level are beyond the scope of this thesis and can be found in reference (5).

These levels can often help find the primary tumor site due to characteristic anatomic lymphatic drainage patterns. Levels I, II, and III are frequently involved in the oral cavity and lip cancers; levels II, III, and IV are commonly involved in laryngeal, oropharyngeal, hypopharyngeal, and thyroid cancer; and level V LNs are mostly involved in nasopharyngeal cancers (27).

In head and neck cancers, the necessity for invasive LN evaluation through dissection depends on the tumor's site and stage. For instance, due to the extensive lymphatic network in supraglottic cancers, occult LN metastasis is highly probable. Consequently, for T1 and T2 stage supraglottic cancers, it is strongly advised to consider radiotherapy or neck dissection of bilateral LNs in levels II and III, even in the absence of clinical and radiological signs of metastasis.

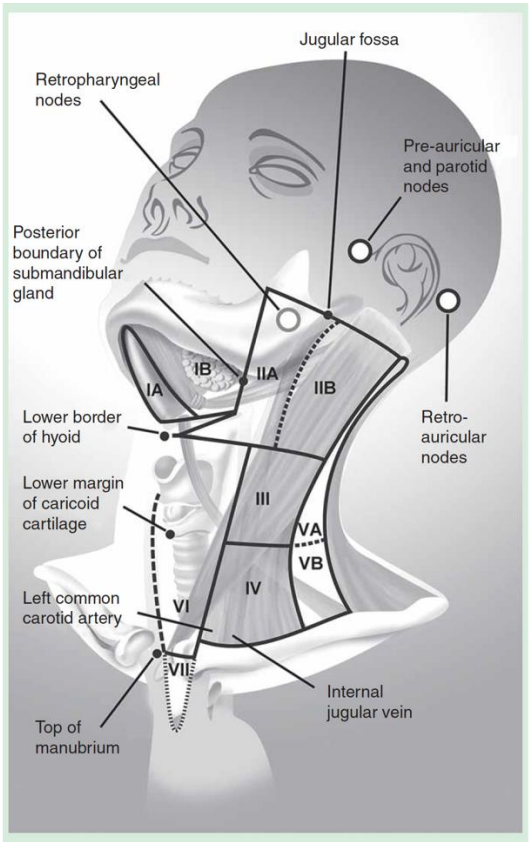

Figure 2.2. Imaging-based classification of deep cervical nodes (5)

For T2b and T3 stage glottic cancers, the same elective treatment approach, radiotherapy or surgery, is applied for bilateral LNs in levels II, III, and IV. Depending on nodal involvement in glottic tumors, treating LNs in levels 1b and V may be suggested (28).

In oral squamous cell carcinoma (OSCC), the most common oral cavity malignancy, there is a risk of occult metastasis to neck nodes even when there is no clinical or radiological evidence of nodal disease. Neck nodal metastasis can reduce survival by

50%, highlighting its role as the most significant prognostic factor (29). Literature suggests that elective neck dissections are recommended for levels I, II, and III as metastatic LNs at levels IV and V are rarely seen (29, 30). One study found that 54% of OSCCs with radiological N0 had occult metastasis in neck nodes, with 44% in level I, 32% in level II, 14% in level III, and 4% in level IV, while none were found in level V (29). Another study evaluated cervical occult LN metastases (OLNM) in primary parotid carcinoma in patients with N0 based on clinical and imaging examination. These patients underwent elective neck dissection, revealing that 30.3% had OLNM, with 69% in level II, 22.5% in level III, 20% in level I, 16% in level V, and 7.5% in level IV. Levels II, III, and V were the most common locations for OLNM (69%, 22.5%, and 16%, respectively) (31). As malignant parotid carcinoma is rare, level V LNs are less clinically significant regarding metastasis (31, 32). We did not include level V cervical LNs due to their limited clinical significance.

Considering the possibility of harboring micrometastases, small LNs should be evaluated for disease staging. Manual detection and segmentation of cervical LNs can be time-consuming, error-prone, and dependent on the observer's experience (33). The reasons for this are explained in more detail in the subsequent section.

2.2 The challenging task of lymph node detection and segmentation

Detecting LNs can pose challenges for several reasons:

- LNs exhibit various sizes, shapes, and locations in CT scans, and their appearances can be influenced by disease effects.
- Different radiologists or medical professionals may have significant variations in interpreting LNs based on their expertise and experience. Additionally, processing vast amounts of data and the large size of each scan can be time-consuming and may lead to errors even among skilled professionals.
- Small lymph nodes can be difficult to distinguish from some other normal structures such as small vessels.
- Noise and artifacts pose further challenges in LN detection, which can arise from factors like scanner malfunctions, imaging protocols, and patient movement during CT scans.

Consequently, fully automated approaches are necessary for rapid and accurate detection and segmentation of the LNs (33). There is evidence in the literature that AI can potentially transform the healthcare sector, particularly in the image recognition (1, 34). Thus, this project aims to develop a model for automatically segmenting head and neck LNs.

2.3 AI in Medicine

The application of AI techniques is a growing trend for performing health-related data and imaging analysis. Due to its significant advancements, AI has been utilized for problem-solving and decision-making in various areas of medicine aimed at improving the healthcare domain (35, 36).

Medical image analysis is a complex task that requires considerable specialist effort. It may be affected by human error, and different experts can have varying interpretations. Consequently, it is crucial to employ machine learning (ML) algorithms to automate the image analysis (1). One of the essential steps before model development is image preprocessing, where the proper application of different preprocessing techniques results in more accurate image analysis. For example, a large portion of each CT slice might be irrelevant and, therefore, not used in image analysis. By cropping the image and reducing its size, preprocessing increases the relative amount of relevant imaging data for analysis.

2.3.1 Preprocessing

Various preprocessing techniques are employed to make images more suitable for analysis, such as cropping, resizing, artifact removal, and filtering. All images may need to be scaled down to accelerate training. Cropping CT images should also be performed to extract the main parts and remove redundant components before feeding the images

into the model (37). The data size of an original Digital Imaging and Communications in Medicine (DICOM) image is 512 x 512, which can lead to a heavy computing workload. Therefore, the CT images were resized and cropped to 224 x 224 only to include the region of interest. Figure 2.3 demonstrates an example of the cropping operation (15).

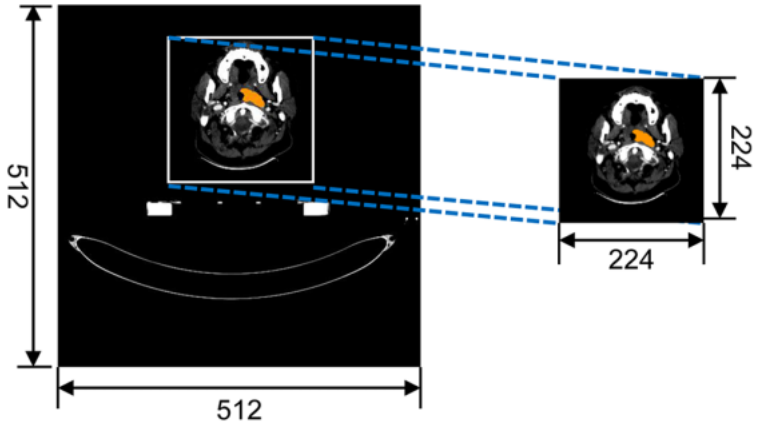

Figure 2.3. Region of interest cropping (15).

Another preprocessing technique for CT images is normalization. Different CT scan equipment may have variations in the field of view and multiple configurations. Normalization operations can be applied to remove these differences using the following formula:

$$\text{Pixel norm} = \frac{\text{Pixel} - \text{Pixel min}}{\text{Pixel max} - \text{Pixel min}}$$

In the mentioned formula, Pixel norm refers to the normalized pixel data in CT images, while Pixel max and Pixel min represent the maximum and minimum gray values in CT images, respectively (15).

2.3.2 Deep Learning

Machine learning and DL models can be classified into different types of learning, such as supervised, unsupervised, and reinforcement learning. Supervised learning uses labelled datasets for classification or predictions, requiring human intervention to label input data accurately. In contrast, unsupervised learning deals with unlabeled datasets, discovering previously undetected patterns. Reinforcement learning is a process where a model learns to make a sequence of decisions based on feedback to maximize the reward (38).

DL is a subset of machine learning algorithms known for efficient performance in various healthcare domains. It uses multiple layers to extract features from raw input, with increasing interest in automated applications (Figure 2.4) (1, 39). Unlike machine learning, DL does not require human expertise for feature selection, as the model itself determines which features to use for the classification (40).

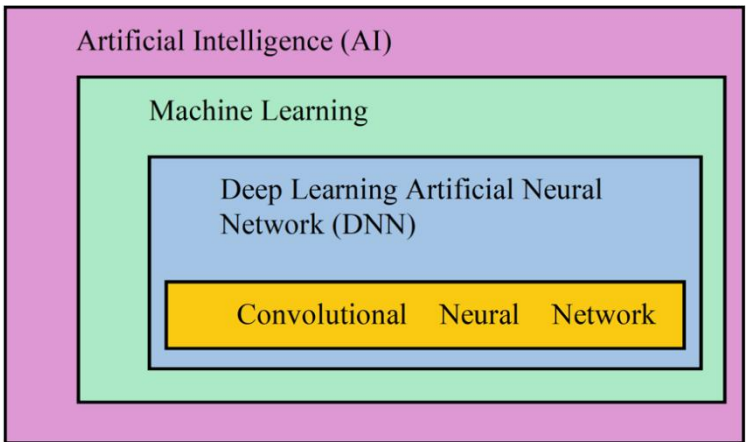

Figure 2.4. The relationship between artificial intelligence, machine learning, and deep learning (1)

DL powers numerous AI applications that enhance automation and physical tasks (41). Requiring minimal human input, DL uses raw pixel data and has outperformed existing methods for detection and classification problems (42, 43).

A deep neural network (DNN) is an artificial neural network consisting of multiple layers between input and output layers inspired by the human biological nervous system. These networks simulate the human brain's functionality and can learn from large amounts of data (41). Depending on the problem's complexity, one or more layers are added between the input and output layers, called hidden layers. The number of layers indicates the model's complexity and capacity (1).

The input data of a DNN undergoes processing, with the increasing complexity of data representations in subsequent layers. The current output is then transferred to the next

layer. DNNs use a mathematical algorithm with various layered parameters in a hierarchical fashion, trained and tested on labelled data through an iterative process to minimize prediction error compared to the actual label. After training and developing the model on labelled databases, neural networks can predict labels on new, unseen data (1, 44-46).

In a feedforward neural network, each neuron of one layer connects to every neuron of the next layer as a directed acyclic graph, which is inefficient with a large number of inputs. Multilayer perceptrons (MLPs), generative adversarial networks (GANs), convolutional neural networks (CNN), and autoencoders are based on feedforward networks (47).

CNN is a subclass of DNN and a state-of-the-art model with excellent performance for object detection, segmentation, classification, and many other image-processing tasks (7). CNN algorithms have been implemented for automated segmentation in different anatomical sites with promising results, including CT scans of the liver and bladder, PET/CT images of skeletal structures, and prostate magnetic resonance imaging (48-51).

CNN contains a set of layers, each doing a particular operation such as convolution, loss calculation and pooling. For functionality, CNN uses spatial associations of the data and replaces neurons with convolutions; therefore, only a limited number of connections between layers is needed, reducing memory requirements. After the beginning layer, which is an input layer, a stack of convolutional layers performs convolutions on the

input. The convolution filters, also known as kernels, act as feature extractors. The results of convolving on the input would be feature maps. Defined by the kernel size, each neuron in the feature map responds only to a region of the previous layer, called the receptive field.

Activation function, such as the rectified linear unit (ReLU), is applied to the network to add nonlinearity. Then, pooling layers are applied to reduce the input dimensions and increase the shift-invariance in the feature detection. The sequence of convolutional, activation and pooling layers are stated as hidden convolutional units (figure 2.5) (7). For the learning process, an input is fed to the neural network. The network analyzes the input and predicts an output. The loss function indicates the difference between the prediction output by the model and the expected output. The DL network training is performed with the objective of optimizing and modifying the weights and biases based on loss value (or cost function). Matrix convolution repetitively happens and subsequently produces new hidden layers of neural maps. In the optimization process, backpropagation and gradient descent are repeatedly applied to the network, updating the network parameters until reaching a constant criterion. Commonly used evaluation metrics to assess the performance of segmentation tasks are the Dice coefficient, intersection over union and pixel accuracy (1, 47).

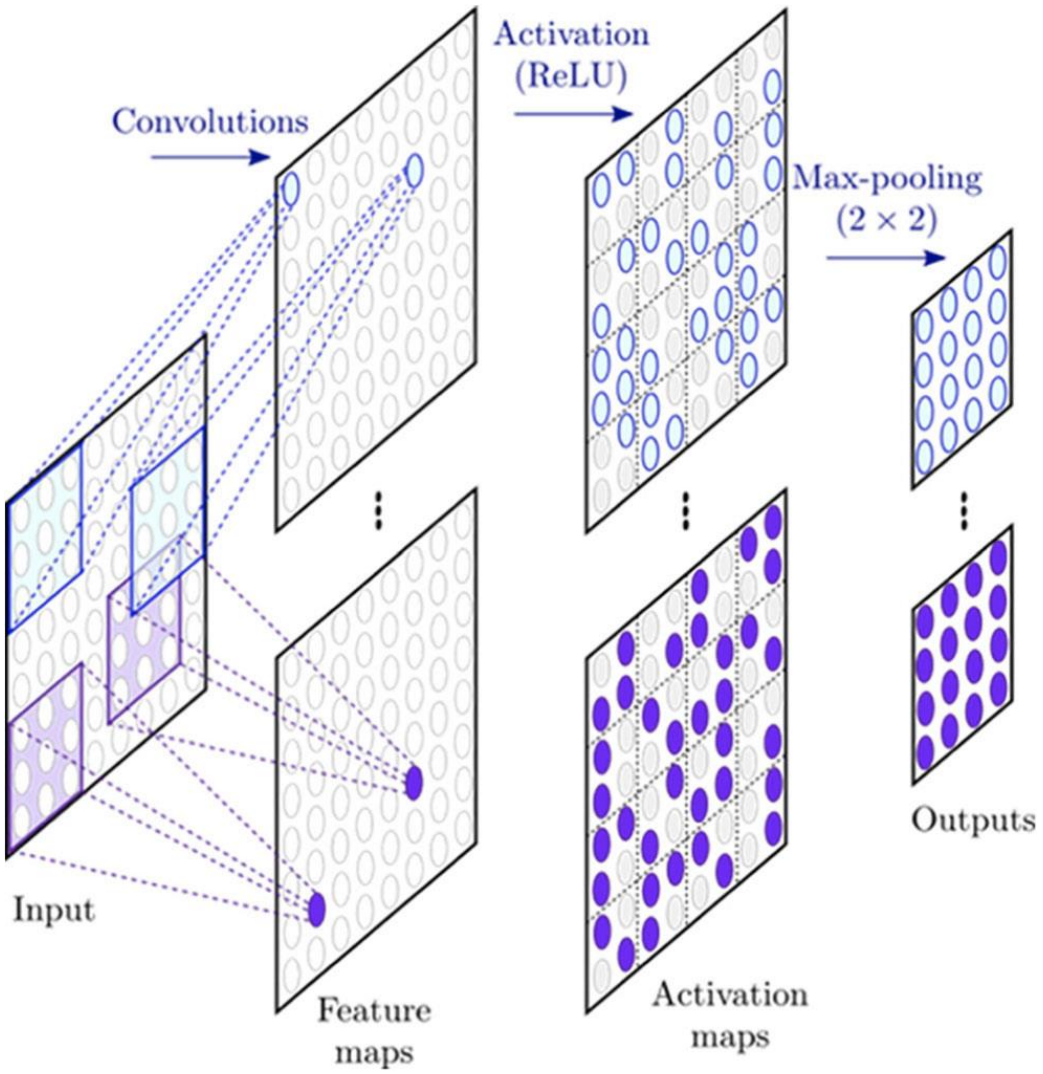

Figure 2.5. The hidden convolutional unit structure (7)

2.3.3 Image Segmentation

According to the American Cancer Society's 2022 report, an estimated 608,570 cancer-related deaths occurred in the United States, amounting to approximately 1,670 daily deaths. Early detection of cancerous cells and taking appropriate measures can save millions of lives. In object detection, only bounding boxes are generated, which do not provide information about the shape of the cells. This is where the importance of image segmentation arises ((1, 52).

Image segmentation refers to the process of dividing an image into several areas based on different features such as geometric shapes, spatial texture, grayscale, and color. Image segmentation is divided into semantic segmentation, instance segmentation and panoramic segmentation. Semantic segmentation is the task of medical image segmentation. There is no universal segmentation method appropriate for all images (53).

Segmentation is a necessary step in obtaining more precise results for the following steps in machine learning, such as image measurement and feature extraction (14).

In medical imaging segmentation, the area of interest is separated from the rest of the image contents (54), detecting the pixels of organs or lesions from the original images, such as MRI or CT scans (55). Medical image segmentation has been utilized for a wide range of applications, such as treatment planning and tumor detection (56, 57)

Manual segmentation is a laborious and time-consuming task that requires expert input, which may not be practically feasible as the number of images necessary for analysis is

increasing exponentially due to advancements in imaging technology (1, 34). These challenges highlight the need for artificial intelligence methods to address these issues. The rapid progress of DL has allowed DL-based image segmentation algorithms to achieve impressive results in the image segmentation (53).

Depending on the model complexity, segmentation can represent either the whole object or one of its components. DL has surpassed traditional machine learning methods regarding segmentation speed and accuracy, with CNNs demonstrating notable superiority by providing unlimited accuracy and feature learning (34). Advances in CNNs have improved both the classification and segmentation of images. CNNs have been implemented in segmenting blood vessels in various areas, including the retina and heart (58). They have also achieved state-of-the-art performance for segmentation and classification in MRI images of head and neck cancers and brain tumors (59, 60). This method has attained high Dice coefficients for segmenting other structures, such as the bony orbit and its background (61).

Despite the significant advancements in CNNs, they do not adequately address the complex challenges of image segmentation, leading to the emergence of other segmentation networks, including U-net (1). The following sections will discuss DL-based networks used for medical image segmentation.

Fully Convolutional Network

The pioneering work of using the most advanced DL models for image segmentation was the fully convolutional network (FCN) proposed by Long et al. (62). The typical FCN consists of a series of convolution layers followed by a softmax layer to attain the classification information of each pixel. The network takes an input of any size and generates a correspondingly sized output. The feature map of the last convolution layer is upsampled, and the classification of each pixel is carried out to accomplish image segmentation. The FCN has been used widely used for segmentation, such as brain tumor segmentation in multimodal 3D MRIs (63), optic disc and cup segmentation in fundus images (64) and pathological lung tissue segmentation in CT scans (65). However, FCN has some limitations, as its upsampling results might be fuzzy and not sensitive enough to image details, leading to less accurate segmentation results (53).

SegNet

SegNet is a semantic segmentation model based on the FCN's segmentation task. In essence, SegNet is a variant of FCN in the decoder part. SegNet consists of a stack of encoders and a corresponding decoder network, followed by a softmax classification layer to assign a class label to each pixel in the image. The main difference between SegNet and FCN lies in the upsampling method during the decoding phase. While FCN uses transposed convolution layers during decoding to upsample feature maps, SegNet stores max-pooling indices during the encoding phase and uses them to upsample low-

resolution feature maps during decoding. The upsampled feature maps are then convolved with a trainable convolution kernel to produce a dense feature map. Finally, the feature maps are upsampled to their original resolution and fed into the softmax classifier, generating the final high-resolution segmentation (66). Figure 2.6 illustrates the difference between SegNet and FCN decoders. Examples of SegNet applications include automated brain tumor segmentation on 3D MRI dataset (67) and infected tissue region segmentation in CT lung images (68).

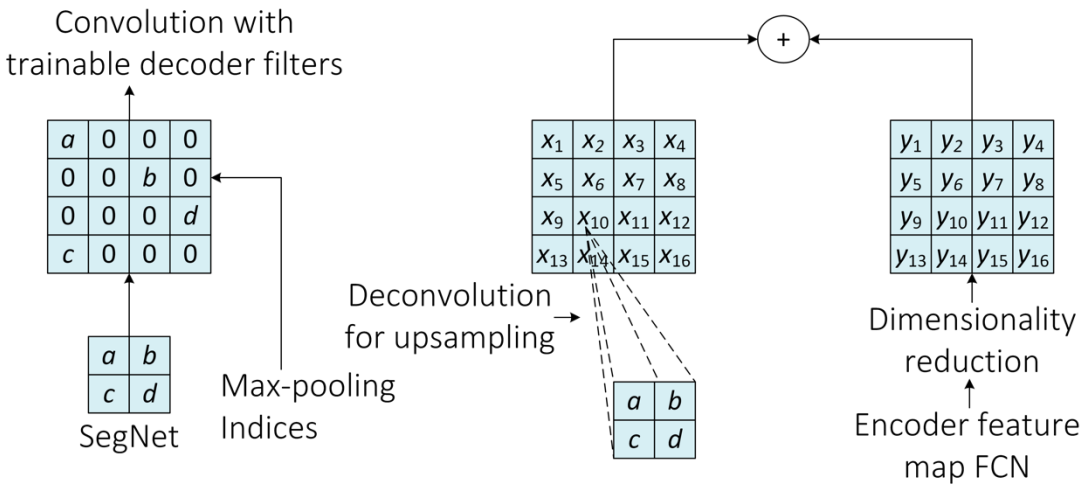

Figure 2.6. Comparison of SegNet with FCN. a, b, c, and d represent the values in a feature map. While FCN learns deconvolution operations, SegNet uses the max-pooling indices to upsample the feature maps. Figure adapted from (6).

U-Net

To segment whole images, a network may not be able to process images beyond a certain resolution due to memory limitations. Dividing large images into regions allows the network to segment larger images. Ronneberger et al. modified the architecture of a fully convolutional network and proposed a model that works with only a few annotated images for training while delivering a more accurate segmentation (2). This novel architecture, called U-Net due to its U-shaped structure, has gained traction in medical image segmentation, leading to the development of its variants, such as 3D U-Net and deep contour-aware network (DCAN) (55, 69, 70).

U-Net comprises the contraction section, the bottleneck section, and the expansion section (Figure 2.7). The contraction section consists of several contraction blocks, each with two 3x3 convolutions followed by a ReLU and a 2x2 max-pooling operation, doubling the number of feature channels. The contracting path generates a dense representation of the input image. Each stage of the expansion phase, consisting of two 3x3 convolutions followed by a ReLU, is followed by a 2x2 up-sampling layer, halving the number of feature channels (1, 2).

Attention Gates in U-Net Model

CNN models have excessive use of model parameters and computational resources due to repeated extraction of similar low-level features. To overcome this problem,

attention gates (AG) were developed, which can be easily incorporated into CNN structures and increase their prediction accuracy and sensitivity. AG in U-Net improves its prediction performance while maintaining computational efficiency because the model learns to focus on the most significant regions of the image while suppressing the irrelevant or unimportant parts (71). More details on this model are described in Chapter 3.

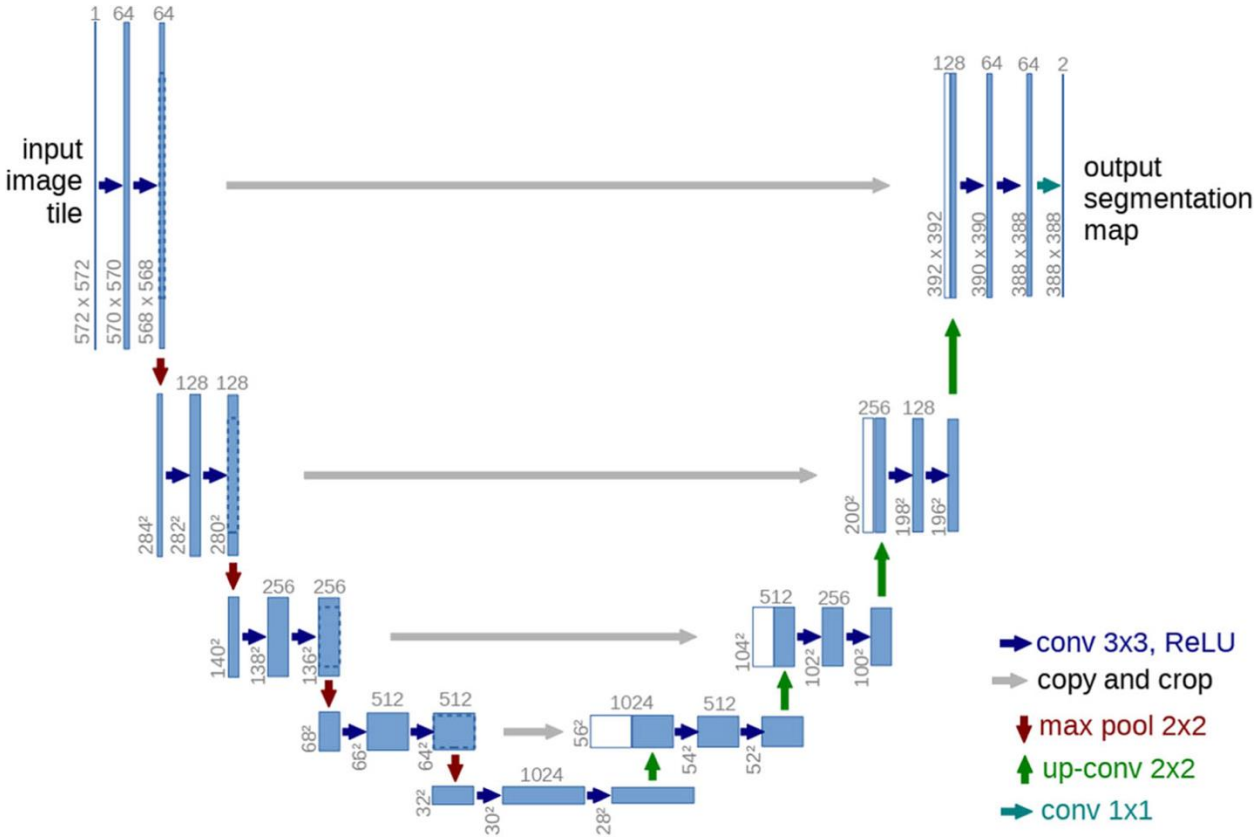

Figure 2.7. U-Net structure (2)

2.3.4 Evaluation Metrics

The performance of a segmentation model should be assessed using appropriate and standardized metrics to ensure the model makes a significant contribution to the field. Commonly used evaluation metrics include Accuracy, Precision, Recall, F1-score, intersection over union (IOU), and Dice coefficient. The following equations represent each of these metrics:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 Score} = \frac{2 \text{ x Precision x Recall}}{\text{Precision} + \text{Recall}}$$

TP, FP, TN, and FN refer to true-positive, false-positive, true-negative, and false-negative, respectively (71).

IOU is calculated as the intersection of predicted regions and ground truth regions divided by the union of both regions. The IOU can be represented with the following formula:

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

Dice similarity is an overlap metric often used to evaluate the quality of segmentation maps by comparing generated segmentation against the ground truth. The following formula is used to calculate the Dice similarity coefficient (DSC):

$$\mathrm{DSC} = \frac{2|\mathrm{G} \cap \mathrm{T}|}{|\mathrm{G}| + |\mathrm{T}|}$$

Where G represents the generated segmentation, and T represents the ground truth. The numerator is the number of overlapping pixels from G and T multiplied by two, while the denominator is the total number of pixels in both G and T. A DSC of 1 signifies perfect overlap, and a DSC of 0 indicates no overlap (72, 73).

Other potential metrics for a proposed model include execution time and memory footprint. While providing the time needed for training the network may not be essential, it can assist other researchers or contribute to reproducibility. Additionally, Graphics Processing Units (GPUs), which are commonly used, may have limited memory. Thus, reporting the peak and average memory footprint of a model can be helpful for understanding implementation-dependent aspects (66).

2.3.5 Challenges of Deep Learning for Medical Image Segmentation

Limited Annotated Data and Overfitting

DL networks necessitate large datasets to achieve high accuracy in complex conditions, such as medical image segmentation. Collecting a vast amount of annotated data for medical images is both time-consuming and expensive (74). When the training dataset is small, overfitting may occur, causing the model to perform well on the training data but poorly on new, unseen data, resulting in a weak generalization (75)

Various approaches can help increase data size, improve DL model performance, and address the overfitting problem:

1. Data Augmentation: This technique involves increasing the size of the training data by applying transformations such as mirroring, rotating, cropping, flipping, or adding noise to the original samples (76).
2. Patch-wise Training: In this method, an image is divided into smaller patches, and the model is trained on each patch separately. However, this approach might not be suitable for small organ segmentation, as random patching could lead to a loss of contextual information and, consequently, an inaccurate segmentation (55).
3. Sparse Annotation: This approach involves labelling only a small portion of the data to reduce the time and cost associated with data annotation (77). However,

the annotated data may not be representative of the entire dataset, leading to poor generalization.

4. Transfer Learning: Fine-tuning a model, initially trained with a large dataset, on a smaller dataset may enhance the model's performance.

Class Imbalance

Class imbalance occurs when only a small number of pixels belong to the region of interest, while most patches correspond to the less important background area. This issue is particularly prevalent in medical image processing. A network trained on such a dataset could be biased toward the background, resulting in poor performance. To overcome this problem, higher weights can be assigned to minority patches during training (78, 79), and patch-wise training may also address class imbalance (80).

Computational Cost and training time

DL's ability to outperform humans comes at a cost. Justus et al. developed a model to predict the computational cost of DL networks, specifically execution time (81). Training networks to learn complex patterns from a dataset requires time and resources. Many studies have focused on reducing execution time and achieving faster convergence. Techniques such as batch normalization, down-sampling, and pooling have been employed to facilitate faster convergence (48, 82).

2.3.6 Related Work

Various DL architectures, such as U-Net and its variants, FCN, GAN, and others, have been used for the task of organ segmentation (83). Numerous studies have utilized DL algorithms to classify and predict cervical LN metastasis (84-87). However, to the best of our knowledge, no studies have employed a model for the automatic segmentation of non-metastatic cervical LNs in healthy individuals. Therefore, in this study, we aimed to develop a DL model for robust LN segmentation in CECT scans of healthy individuals. Ariji et al. used Detectnet and U-Net, for automatic segmentation and metastasis detection in cervical LNs of patients with oral squamous cell carcinoma. Recall, precision, and F1 scores were used to evaluate the model's performance for detecting metastatic LNs, both overall and for each level separately. The recall values of metastatic and non-metastatic LN segmentation were 0.742 and 0.782, respectively, indicating insufficient performance that requires improvement (88, 89).

In another study by Tomita et al., CNN and transfer learning were employed to differentiate between benign and metastatic cervical LNs in patients with squamous cell carcinoma. The model's area under the curve (AUC) was calculated at 0.898, which was higher than the radiologists' performance (90).

Iuga et al. (33) collected a dataset of 89 contrast-enhanced CT scans of the thorax containing 4275 LNs. A radiologist segmented all the LNs semi-automatically, evaluating the 3D volume of the LNs. A 3D fully convolutional neural network was

trained on this dataset using four-fold cross-validation. The total detection rates for enlarged LNs were 76.9% in the training set and 69.9% in the testing set, respectively. The detection rate of enlarged LNs, with a short-axis diameter (SAD) ≥ 20 was much better than that in the small LNs, with a SAD 5–10 mm, 91.6%
versus 62.2%, respectively.

Using a 3D CNN for extranodal extension detection in head and neck squamous cell carcinoma by Kann and colleagues demonstrated that the CNN algorithm outperformed radiologists (25).

A 3D foveal fully CNN (U-Net) was applied for automated detection and segmentation of thoracic LNs using contrast-enhanced CT scans. The output was a probability map indicating the likelihood of each voxel being an axillary or mediastinal LN. The algorithm achieved excellent detection performance with reasonable generalizability and a DSC value for segmentation accuracy, facilitating LN detection in routine clinical work (33).

It has been noted that while unidimensional measurement of LN SAD is routinely used for nodal disease staging, two-dimensional (2D) approaches may underestimate lesion size. Therefore, considering the entire volume of the LN is crucial for an accurate segmentation (33, 91).

Manjunatha et al. proposed a two-stage approach for CT scans of mediastinal and abdominal LNs. In Stage I, they used modified U-Net with ResNet architecture to have high sensitivity, which was achieved with the cost of increased false positives, with

sensitivities of 87% at 2.75 false positives per volume. For false positive reduction, they used a 3D convolutional neural network classifier in stage II (92).

Cai et al. developed a slice-wise label-map propagation algorithm on response evaluation criteria in solid tumors (RECIST), inspired by weakly supervised image segmentation and due to the expensive LN segmentation cost. They reached a mean DSC of 92% on RECIST slices and 76% on the lesion volume (93).

Sartor et al. employed a CNN to automatically segment the clinical target volume of LNs in patients with anorectal or cervical cancer. Using Dice scores and the distribution of Mean Surface Distance for model evaluation, the CNN method achieved a high performance (94).

Zhou et al. developed an FCN architecture for the segmentation of multiple organs in 3D CT images. Their model demonstrated promising results, achieving an acceptable accuracy of 88.1% voxels for the training dataset and 87.9% voxels for the testing dataset in segmenting 19 structures of interest. However, the authors acknowledged a limitation of their network, which was its lower accuracy in segmenting smaller structures (95).

Although DL models have been effectively employed in numerous fields, their use for assessing small LNs in head and neck cancer is still scarce (16). Also, the segmentation of objects which occupy only a small fraction of pixels remains a challenging task (83). We will discuss this problem in more detail in the methodology section.

Chapter 3

Methodology

This chapter describes the whole workflow in detail. The first section contains the dataset description and its preparation. The following sections are dedicated to a comprehensive overview of the steps involved in this work, including the preprocessing, training, and evaluation metrics in detail.

3.1 Dataset

3.1.1 Dataset Acquisition

The dataset consists of 221 head and neck contrast-enhanced CT scans obtained from the Augmented Intelligence and Precision Health laboratory (AIPHL), which belongs to the research institute and Department of Radiology of McGill University Health Center. The institutional review board was approved at the McGill University Health Centre Research Institute. The criteria for including participants in the study were as follows: (1) The participants who have undergone a contrast-enhanced CT scan of the neck, (2) the scan would have been interpreted as normal or with minor inconsequential incidental findings, and (3) the participants would have been adults aged 18 years or older. The criteria for excluding participants were: (1) The presence of any nodal disease or

abnormality on the scans, (2) the presence of any known or suspected primary malignancy on the scans, (3) the presence of significant inflammatory change or abscess on the scans, and (4) any history of known malignancy.

Those CT scans belong to healthy individuals without any head and neck cancer.

3.1.2 Dataset preparation

The dataset was stored as DICOM files imported to an open-source software for medical image visualization called 3D-slicer version 5.0.3. Imaging-based classification for cervical LNs is divided into seven anatomical levels initially proposed by Som et al. (20). Considering these levels, levels I to IV are the areas with the most remarkable propensity for LN metastasis from HNSCC and of most significant clinical interest; therefore, all levels I to IV of these CT scans, including 1A, IB, II, III and IV, were reviewed. Normal LNs with a long-axis diameter of $\geq$ 5mm on axial planes were manually segmented in the dataset by a trainee (S.G.) and a neuroradiologist (P.T.). Figure 3.1 illustrates an example of this operation. Coronal planes would also be reviewed when LN detection was challenging due to the surrounding soft tissues. The remnants of each LN, which would be smaller than 5mm, were contoured in the subsequent slices.

These annotations were then reviewed and modified as needed by a fellowship-trained neuroradiologist and head and neck radiologist with over ten years of clinical practice

experience (R.F.). The CT scans and their corresponding mask images were stored in the format of NRRD files.

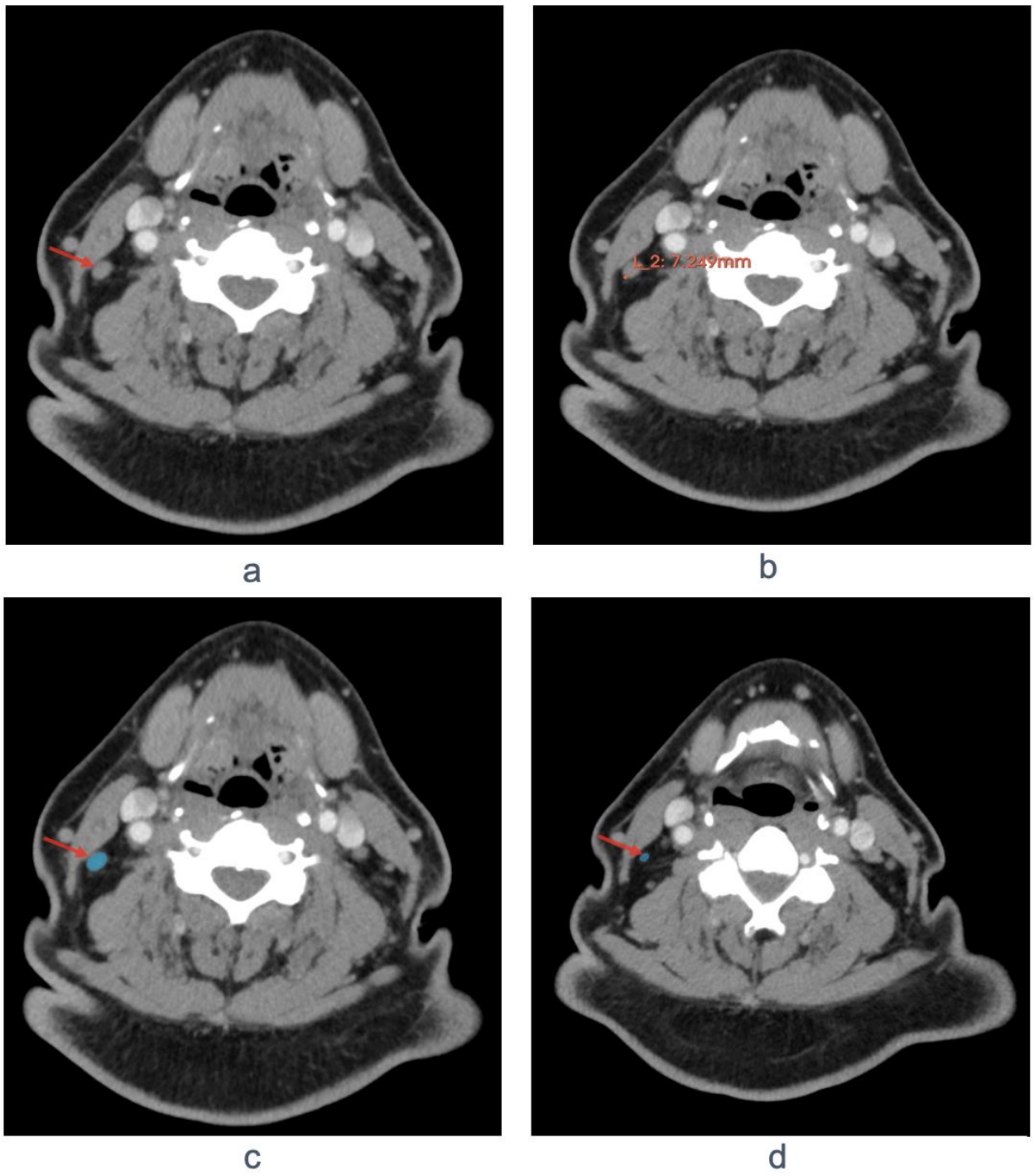

Figure 3.1. The steps taken to prepare our ground truth: A. First, we find a LN (the red arrow is representative of a cervical LN at level II). B. Then, we measure the size of the longest axis of the LN. We have to move forward and backward of the subsequent slices on the CT image in order to find the slice where the LN is at its largest size. C. If the longest axis is ^{3} 5 mm, we would contour the whole LN. D. We would annotate the subsequent slices containing the mentioned LN. Image D shows the last slice of the mentioned LN.

3.2 Image preprocessing

Preprocessing the data is one of the essential steps for getting the best image analysis. This is why we took several steps for preprocessing consisting of center cropping, windowing, clipping, and normalization.

In each CT image, slices above the bottom of the orbits and below the top of the lungs were discarded. Each original DICOM image is 512 × 512 pixels. Many parts of each slice would be redundant and should be removed to have a more accurate image analysis. Also, during the training phase, the irrelevant regions in the original CT image can cause a significant computational workload. Therefore, we center-cropped all images to 384 × 384 pixels from the original image, which included the main anatomical structures and our targets. For the final training, the process of the region-of-interest extraction was automated and it was performed in the same way across all images. Windowing maps the original pixel values of the image to a new specified range of values, which helps improve the visibility of certain structures or tissues in the CT image. We considered Slope 1, intercept 0, window center 40, and width 400, which are the typical window center and window width values for the soft tissue in the head and neck (96). We can visualize the LNs in these values properly.

$$Image = (Image \times slope) + Intercept$$

$$window\_min = level - (window\_width / 2)$$

$$window\_max = level + (window\_width / 2)$$

Based on the above formulas, the minimum and maximum values would be calculated as:

$$window\_min = 40 - (400 / 2) = -160$$

$$window\_max = 40 + (400 / 2) = 240$$

Clipping is one of the common preprocessing techniques in image preprocessing. Using the above formulas, the pixel values of CT slice images were limited to the range of -160 to 240. Any pixel values below a lower bound or above an upper bound were set to the corresponding bound values. By restricting the pixel values, clipping enhances image contrast and visibility, making the image more appropriate for the following processing. The formula for clipping an image can be written as:

$$clipped\_image = max(min(original\_image, 240), -160)$$

Where the original image is the input image and max and min are functions to compare the pixel values in the input image to the maximum and minimum, returning them to these values, respectively. After applying soft tissue windowing, the slices containing air regions were removed from further analysis to improve the model's accuracy and reduce the computational burden.

Different CT scans might have various configurations. By deploying normalization in CT images, the range of pixel intensity of the remaining values would be scaled between zero and one using the following formula:

$$Slice\ norm = \frac{slice - \min(slice)}{max\ \ (slice) - \min(slice)}$$

Where slice is the source pixel data in CT images, slice norm is normalized CT images pixel data, and max slice and min slice are the original CT images' maximum and minimum gray values, respectively.

3.3 Lymph Node Segmentation

We trained and analyzed two different state-of-the-art architectures for LN segmentation: U-Net with attention and Focus Net.

3.3.1 U-net Architecture

U-net and its variants have been widely used for segmentation in medical imaging. The ability to work with small datasets and achieve high accuracy provides this model with high utility in the analysis of medical images (2, 97, 98).

In the contracting path of the U-net, the input image goes through a series of convolutional and pooling layers, which reduces the spatial resolution of the image and

gives a compressed representation. Then the image goes through the expansive path, a series of convolutional and upsamlping layers, which recover the spatial resolution and result in the final prediction map. The training process is performed by calculating the loss by comparing the model's predicted output with the ground-truth segmentation mask. Figure 3.2 represents the fundamental architecture of the U-Net model.

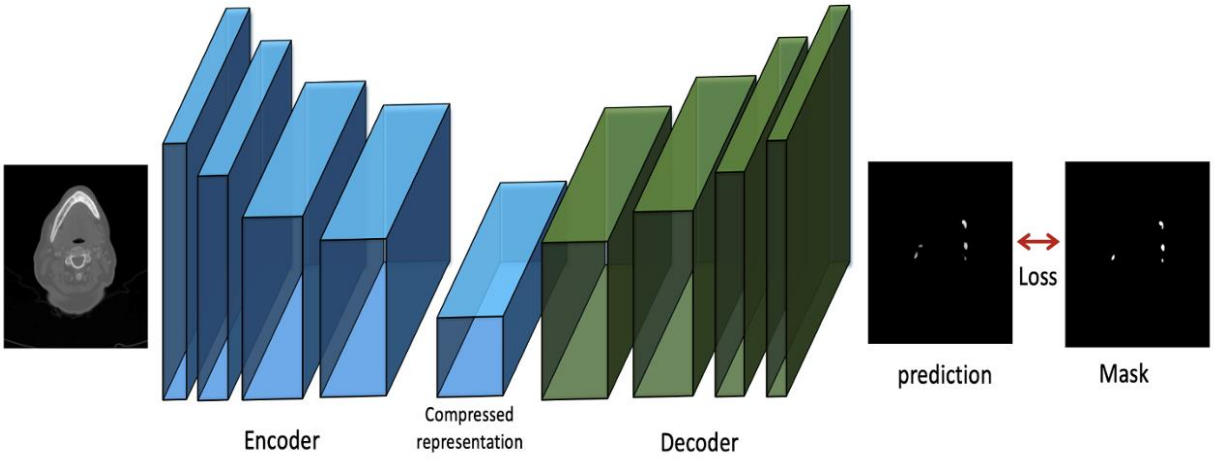

Figure 3.2. U-Net Structure used for the ground truth

Attention U-Net

Focusing on specific objects that are of importance and ignoring the irrelevant areas is a desirable trait in the image processing network. The attention U-net achieves this trait by using an AG. The expansive path of attention U-net has an AG implemented in the skip connection of each layer. The corresponding features from the contracting path

have to go through this AG before combining with the up-sampled features in the expansive path. This AG also suppresses activations at irrelevant regions. Figure 3.3 describes an additive AG.

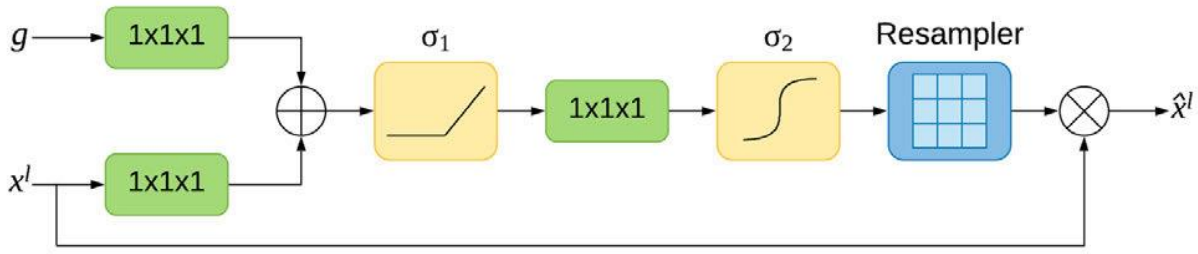

Figure 3.3. The schematic of additive attention U-net (4).

Based on Figure 3.3, each AG has two inputs: the gating signal (g) and the connection from the corresponding encoder layer (x). Both inputs pass through separate 1 × 1 × 1 convolutions. Then the signals are combined; the aligned weights get larger while the unaligned weights get relatively smaller. After passing through the ReLU activation, 1 × 1 × 1 convolution and the sigmoid activation, the output is up-sampled or resampled to the same size as the input x. Finally, this output is multiplied element-wise to the original x vector. By this mechanism, the network focuses on the most informative features, which would improve its segmentation performance without requiring excessive computational complexity (4).

Based on the above definition of attention, we used a U-Net-based structure as our baseline model. Figure 3.4 represents a schematic architecture of our U Net.

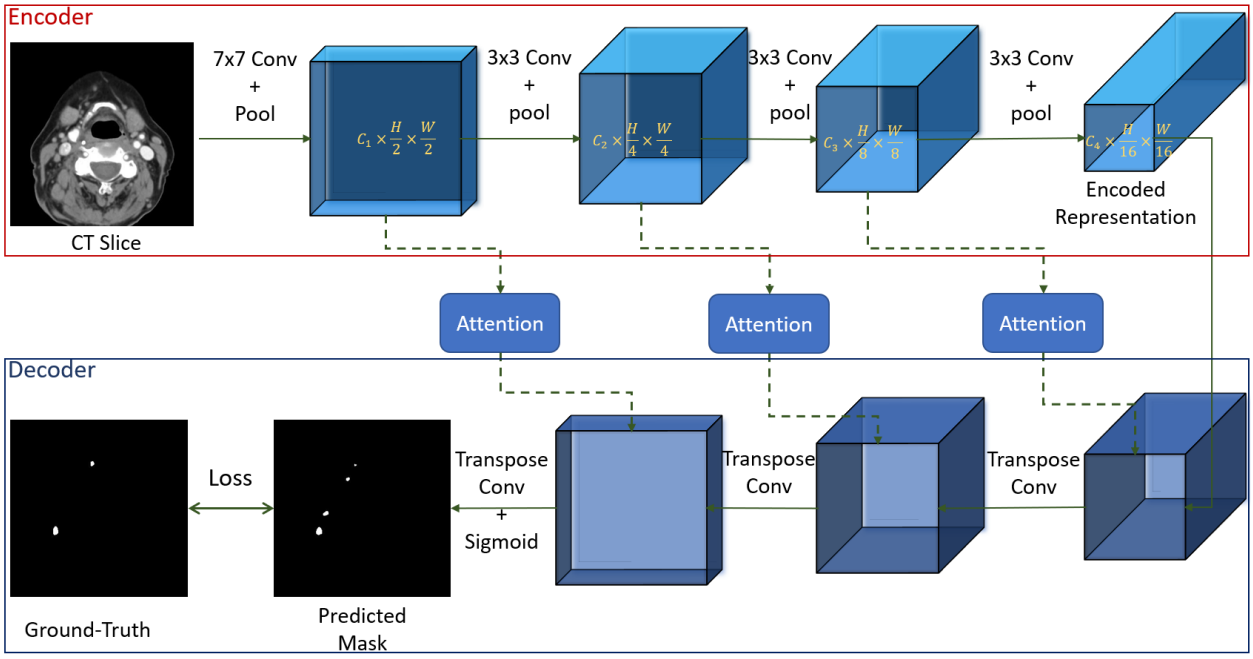

Figure 3.4. The architecture of Attention U-Net for LN segmentation.

The breakdown of the different components of the U-Net architecture is as follows:

1. Encoder path: The input image, as illustrated above, goes through multiple convolutional blocks in the encoder. Convolutional operations are applied in each block to extract features from the image. The output of the encoder path is a compressed representation of the input image containing deep features.

2. Decoder Path: This encoded representation from the encoder is then passed through the decoder layers, consisting of a series of Transpose Convolution blocks, also known as deconvolution blocks. These blocks upsample features with the goal of reconstructing the original input image while preserving the extracted features.
3. Skip Connections with Attentional Module: To improve contextual information flow and retain fine-grained details, skip connections with attentional modules are employed between the encoder and decoder layers.
4. Segmentation Mask Prediction: The final layer in the decoder applies the sigmoid activation function to predict a segmentation mask. The sigmoid activation function is a mathematical function that maps the input value to a value between 0 and 1 and generates a pixel-wise probability for each class. Finally, the model would produce a segmentation mask which is a binary mask that assigns a prediction value to each pixel, highlighting the regions of interest.

In summary, this structure enables the extraction of significant image features and maintains spatial information by employing skip connections with attentional modules, resulting in producing a segmentation mask that identifies and categorizes different regions within the image (99).

We used binary cross-entropy, which is a traditional loss function, to compare the predicted segmentation map with the original segmentation map. This loss function is commonly used for binary segmentation tasks in U-Net. For this purpose, we converted

these segmentation maps to a vector of probabilities. Binary cross entropy is mainly used for classification tasks, but it can also be used for segmentation as a pixel-level classification using the following formula:

$$Binary\ Cross\ Entropy\ Loss(y, p) = -(y\log(p) + (1 - y)\log(1 - p))$$

Where p is the predicted segmentation probabilities by the prediction model, and y is the true segmentation probabilities (100).

Limitations of Attention U-net

While attention U-Net has been widely used in the medical domain and has achieved promising results for segmentation tasks, it has some limitations. The objects we are targeting for segmentation are cervical LNs which are too small. The U-Net model applies convolution and down-samples the image several times, 16 times down-sampling as shown in figure 3.4. Too much down-sampling and excessive input compression can result in the loss of high-resolution information. As each LN occupies only a few voxels, this significant down-sampling causes the loss of information, which is crucial to produce an accurate segmentation map. Combining low-level and high-level features would provide only a partial solution and cannot address this issue thoroughly (101).

Also, the foreground or the main region of interest in our images is very small compared to the background area. For this imbalance, the traditional loss function might have poor performance. To overcome these challenges, we deployed Focus Net and used binary Tversky loss.

3.3.2 Focus Net

To address the mentioned limitations of U-Net, we adopted the spatial context network from FocusNet. FocusNet applies down-sampling only twice in the encoder path, which would help retain as much detail as possible. However, minimizing the information loss comes with the cost of a limited receptive field. Downsampling the input images only twice leads to a relatively small receptive field, hindering the network from capturing high-level features and extensive contextual information. To overcome this,
we applied dense atrous spatial pyramid pooling (dense ASPP) module, which captures contextual information from the same feature map at multiple scales. Before going further, we delved into the concept of atrous convolution.

Atrous convolution, also known as dilated convolution, was developed to overcome the limitation of traditional convolutions in capturing context at different scales and increase the receptive field of a convolutional layer.

In a traditional convolution, a given kernel, also called filter, slides over the input feature map using a fixed stride. Dissimilar to fixed stride convolutions, atrous convolution

includes gaps between filter weights, leading to receptive field expansion without requiring adding parameters. Traditional convolution concentrates on local features, while atrous convolution provides the network with capturing a broader scope of information by a parameter called the dilation rate. Figure 3.5 demonstrates their difference.

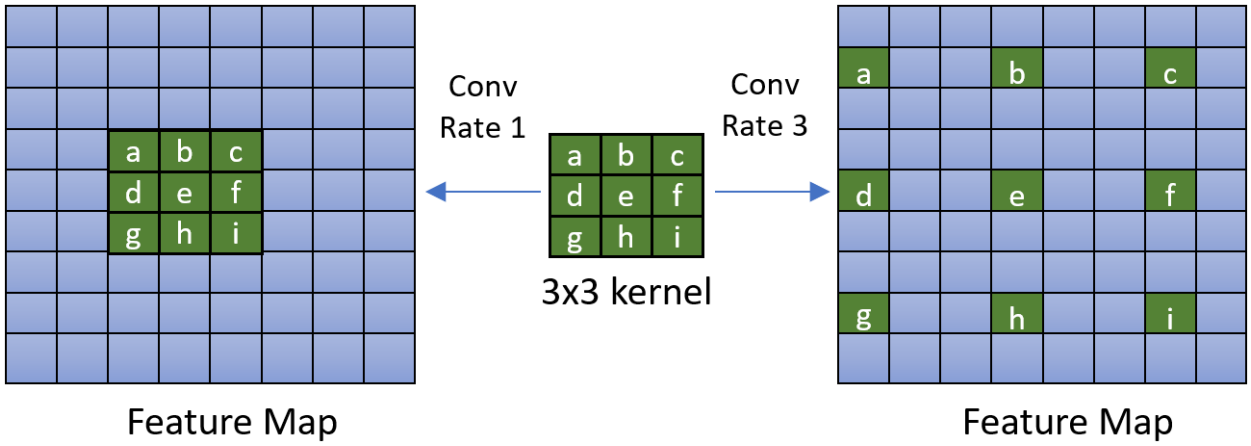

Figure 3.5. Atrous Convolution

For a two two-dimensional signal x, applying atrous convolution with a filter w yields the output y following the equation:

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k]w[k].$$

where the atrous rate r corresponds to the stride used to sample the input signal, equivalent to convolving the input x with unsampled filters produced by inserting r - 1

zero between two consecutive filter values along each spatial dimension. Standard convolution is a particular case of atrous convolution with r = 1. We can modify the filter's field-of-view (FOV) by changing the rate value.

Atrous Spatial Pyramid Pooling (ASPP) employs atrous convolutions to merge multiple atrous-convolved features, each with a different dilation rate, to create a final feature map representation. The dilation rates enable the network to capture information at various scales (102). Due to a limited number of down-sampled features, ASPP would have a small receptive field. To address this limitation, we adopted the densely connected ASPP (DenseASPP), which connects a set of atrous-convolved features in a dense way, enhancing the input image representation by information aggregation and exchange across multiple scales (103).

Also, in DenseASPP, skip connections were introduced between features of the same scales from the encoder to the decoder to prevent the loss of contextual information. These connections aim to ensure the continuous flow of relevant information across the network. Rather than a simple addition of the features, a technique called reverse axial attention was applied to merge the encoder features with the decoder features, enhancing the reconstruction results by emphasizing the relevant foreground regions (104).

In our model inspired by FocusNet, following the two times downsampling and application of dense ASPP, the resulting feature is up-sampled to restore its original resolution and then concatenated with the original feature. As downsampling in the

encoder layers was twice, the up-sampling step would be performed twice as well. Finally, a sigmoid activation function would be applied to the concatenated feature, which maps the values between 0 and 1. The sigmoid output is then used to predict a segmentation mask.

Overall, the process involves downsampling, applying ASPP, up-sampling the feature, concatenating it with the original feature, applying a sigmoid function, and finally predicting the segmentation mask. Figure 3.6 describes the FocusNet architecture.

We fed our DL with raw data, and it produced the output; we did not have to do any extra feature extraction, illustrated in Figure 3.6.

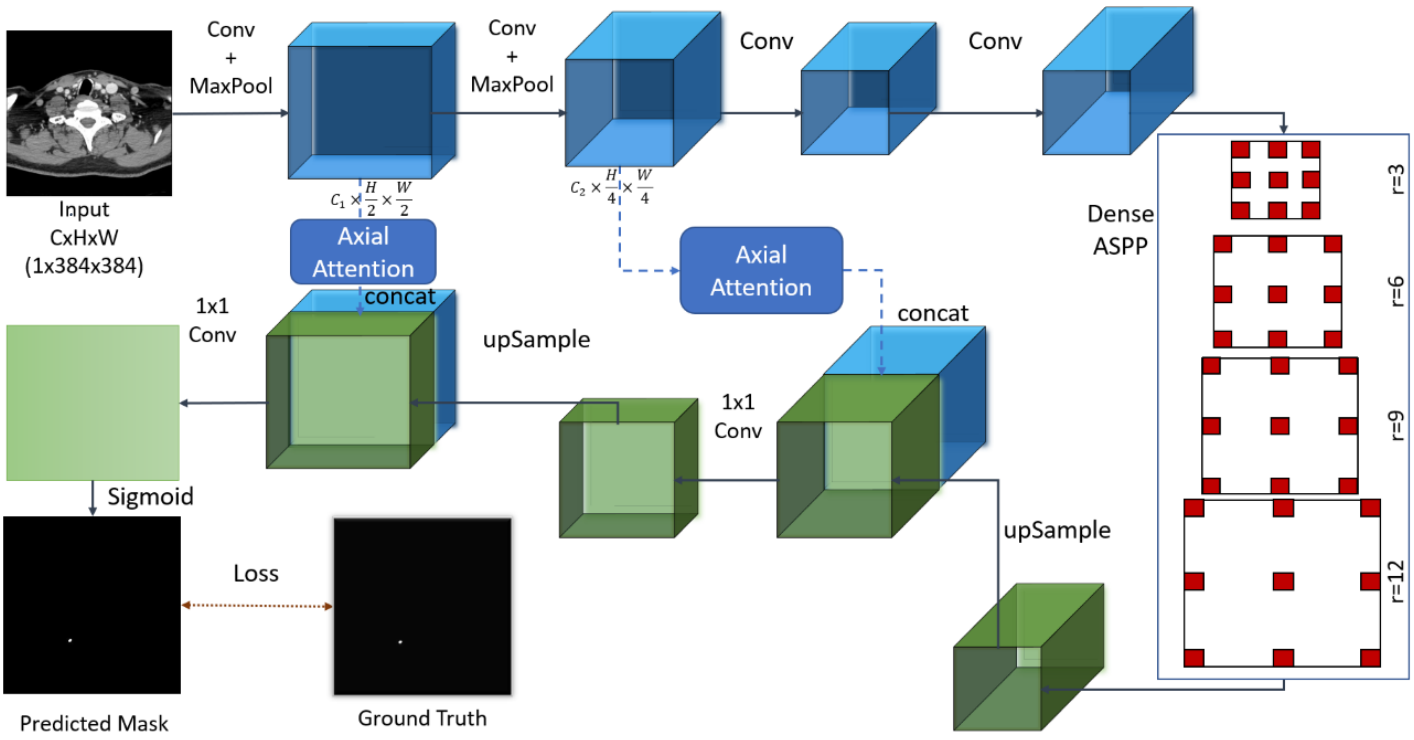

Figure 3.6. The architecture of the FocusNet

Due to the mentioned limitation for the traditional loss function, we used Binary Tversky loss to evaluate our model. The Tversky index is a similarity measure that compares two sets and is defined as:

$$Tversky\ Index = TP / (TP + \alpha * FP + \beta * FN)$$

where TP represents the number of pixels correctly identified as positive, FP indicates the number of pixels incorrectly identified as positive, and FN represents the number of pixels incorrectly classified as negative. α and β are weighting parameters that regulate the balance between false positives and false negatives.

Binary Tversky loss extends the Tversky index and serves to calculate the dissimilarity between ground truth and predicted output segmentation masks.

$$Binary\ Tversky\ Loss = 1 - Tversky\ Index$$

The Binary Tversky loss function aims to minimize the dissimilarity between the predicted and ground truth segmentation masks. By optimizing this loss function, the model learns to accurately segment and classify regions of interest in the binary segmentation task.

3.4 Experimental Setup

3.4.1 Data Split

Data splitting was performed to avoid overfitting. The images and their corresponding contours were used as inputs for the training phase of our DL-assisted models. The 221 CT exams were split into three sets: training, validation, and testing.

The first 160 CT scans were used as the training set to train the model. The subsequent 40 CT scans were used to validate the algorithm by tuning different hyperparameters. Finally, the last 21 CT scans were used as the test set to assess the model's performance and generalizability. It should be noticed that the data split was at the patient level rather than the image level.

Overall, we had 18054 CT slices for training, 4463 slices for validation, and 2602 slices for testing the model.

During the training phase, we encountered a significant imbalance in the distribution of LN within the dataset. Of the total 18,054 slices, only 4,644 CT slices, which account for 25.7% of the total slices, contained LNs, resulting in a class imbalance at the dataset level. This dataset-level class imbalance posed challenges during the learning phase, and we also had to address the intra-sample (pixel-level) class imbalance to prevent bias in the model's performance.

We implemented a uniform sampling strategy at the class level to alleviate this issue. In each training epoch, we randomly selected an equal number of negative samples,

defined as samples without LN, along with the positive training samples. This method mitigates the problem and allows us to focus on improving the pixel-level imbalance.

3.4.2 Training

We used the S-Net architecture from FocusNet (101) as the backbone of our method and incorporated reverse axial attention with the network. We also observed the effect of training and finetuning the network with classification and localization, respectively.

3.4.3 Computational Resources and Hyperparameters

We used the PyTorch DL library. The procedures were performed on machines running the Unix system (Fedora).

During the training process, a weighted Adam optimizer with a learning rate of 5e-5 was utilized. We decided to use a relatively small learning rate based on the limited number of available LN slices. The U-Net model was comprised of a total of 34,877,421 parameters. The Spatial Context Network with reverse axial attention, had a precise count of 27,134,416 parameters. Additionally, the Spatial Context Network without attention model was constructed with 821,613 parameters.

We used a batch size of 8 for each experiment with images of 384x384 (height x width). We utilized two NVIDIA GeForce RTX 2080 graphics cards, each of them equipped with a memory capacity of 12 GB.

3.4.4 Evaluation metrics

We used the DSC and Jaccard index, also known as the Jaccard similarity coefficient, as both are among the most popular metrics for evaluating segmentation tasks in medical imaging.

Dice similarity coefficient

We assessed the performance of our DL models for cervical LN segmentation by calculating the DSC, referred to as the Dice score, of the generated contours by the model against the original contours.

$$DSC = \frac{2 * intersection}{Total\ Predicted + Total\ Ground\ Truth}$$

where intersection refers to the number of pixels correctly identified as positive in both the ground truth and the predicted masks, Total predicted represents the overall number of positive pixels in the predicted mask, and Total ground truth represents the total count of positive pixels in the ground truth mask.

Jaccard Index

We also used the Jaccard index, commonly known as IOU, which is another similarity measure between two sets and is calculated using the following formula:

Jaccard Index = Intersection of the Sets / Union of the Sets

Both of these metrics range from 0 to 1 to quantify the segmentation accuracy. The main difference between them lies in how their dominators are calculated; the DSCs use the sum of the sizes of each set, while the Jaccard index uses the size of the union of the sets. DSC tends to have more sensitivity to small differences. Conversely, the Jaccard index is more commonly used for evaluating the overall similarity between sets. Based on the emphasis on small variations or balanced similarity, each of them can be chosen. The following formulas are written to provide an easier comparison between these two metrics:

$$\textit{Dice Score} = DSC = (2 * |A \cap B|) / (|A| + |B|)$$

$$\textit{Jaccard Index} = J = |A \cap B| / |A \cup B|$$

Where A and B are the two sets being compared, |A| and |B| represent the number of elements of sets A and B, respectively, ∩ represents the intersection of sets A and B, and ∪ represents the union of sets A and B (105-107).

Chapter 4

Results and Analysis

In this chapter, the results of the experiments mentioned in the previous chapter are described in consecutive order.

We leverage the produced annotated data to establish the LN segmentation tools. The DL pipeline developed in our laboratory by adapting existing pipelines is presented in the following.

4.1 Lymph node Segmentation

The dataset used for the study consisted of 221 contrast-enhanced head and neck CT images. All LNs with a long axis diameter of $\geq$ 5mm on axial planes were detected, manually contoured, and then reviewed by an expert neuroradiologist (Figures 4.1 and 4.2).

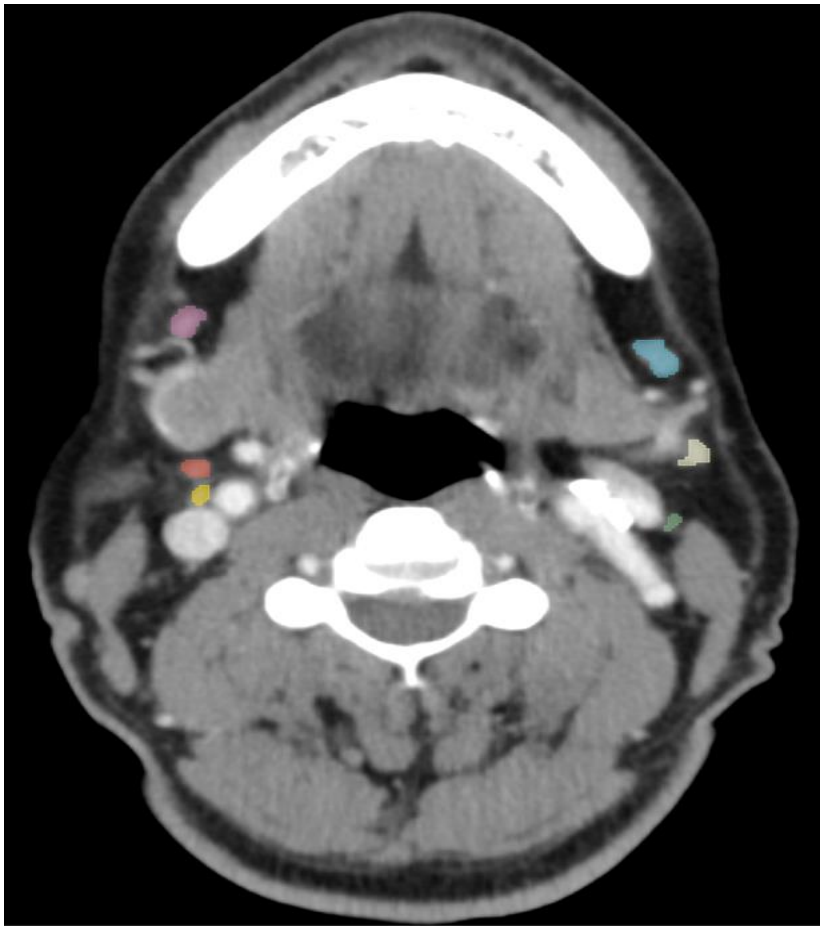

Figure 4.1. Sample image of the dataset. Axial view of CT scan. Manually segmented Right level IB LN (pink), Right level II LNs (red and yellow), Left level IB LN (blue) and Left level II LNs (green and white). This person had multiple other LNs with long axis of ≥ 5 mm, which can be seen on the subsequent slices on the CT scan.

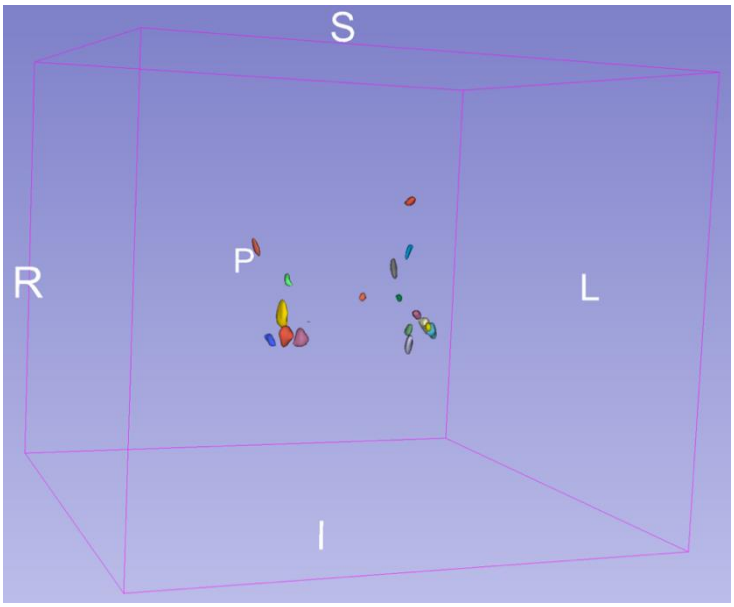

Figure 4.2. Whole 3D of LNs of the person in Figure 4.1. LNs in other levels, including level IA, level III (right side), level III (left side) and level IV (left side), are presented as well.

4.2 Training

The annotated LNs were used to train our model. We trained both attentional U-Net and Spatial Attention networks for our analysis. Comparing the predicted segmentation mask with the original mask, we computed the DSC and Jaccard index to assess the performance of the cervical LN segmentation.

Figures 4.3 and 4.4 visualize the training progression of the S-Net model, the increasing trend of accuracy and the decreasing trend of the loss values as the model progress through the epochs, respectively. The X-axis indicates the epoch number; each epoch

indicates a complete pass of the entire training dataset through the model n mini-batch format, with batch size 8. The Y-axis in Figures 4.3 (a) and 4.4 (a) represents the corresponding matrix value.

At the end of each epoch, the model is saved and evaluated using the validation set. The corresponding loss and dice-coefficient values of images from the validation set are saved. The average values of loss and dice-coefficient achieved from the validation set using the saved model after each epoch training are plotted in the graph. The training process continues iteratively, using the saved model from the previous epoch as the

starting point for the subsequent epoch.

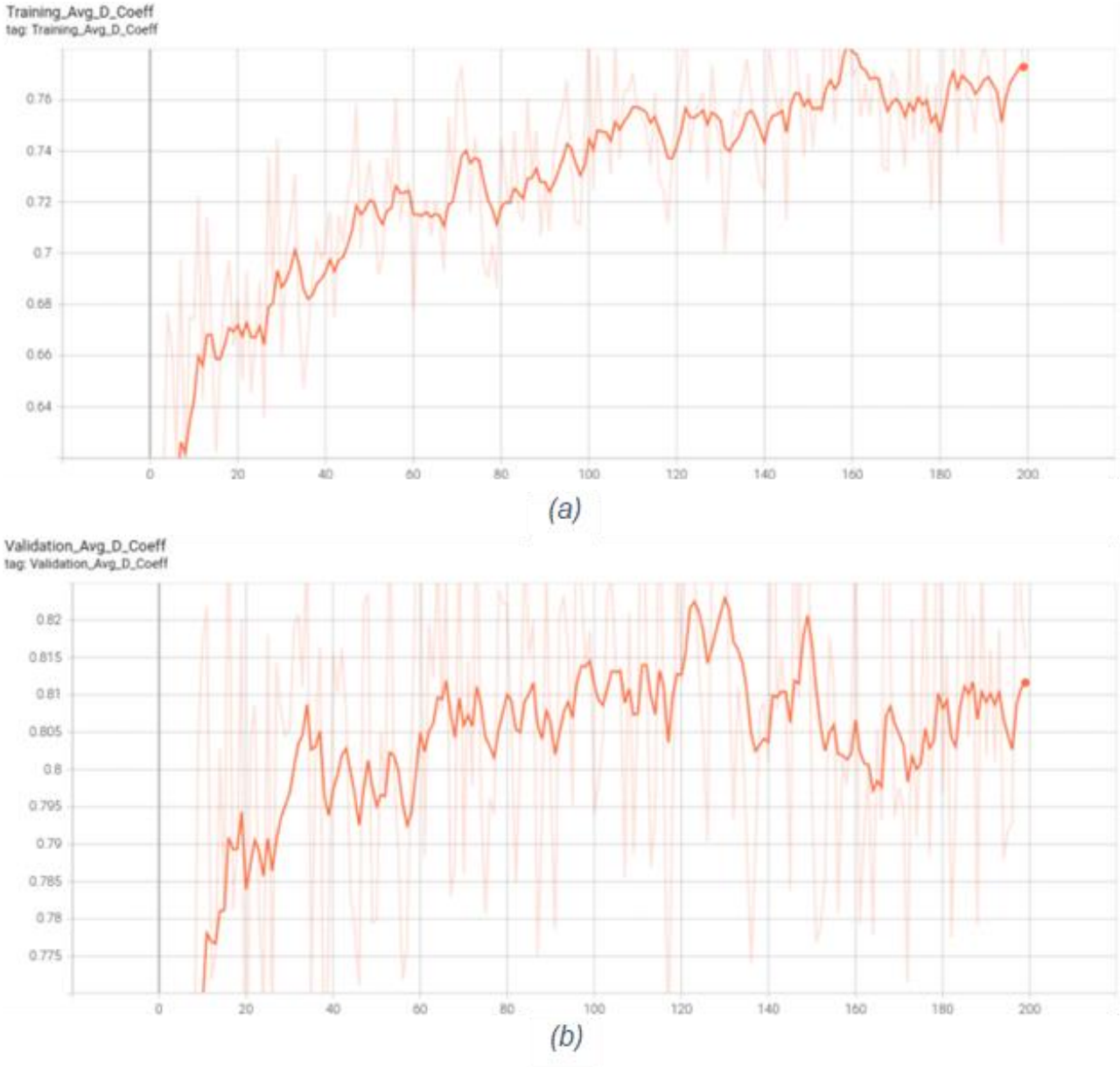

Figure 4.3. Part of the learning curve of the S-Net. (a) on the training dataset (b) on the validation dataset. The horizontal axis indicates the number of epochs. The vertical axis represents the performance of the learning model, shown and calculated as the Dice coefficient.

(a)

(b)

Figure 4.4. Plot of Loss error of the S-Net. (a) over the training Epochs (b) over the validation Epochs. The horizontal axis indicates the number of epochs. The vertical axis represents the Loss error.

There is considerable variance in the matrix attributed to the output sensitivity of the object size. Small lymph nodes result in a substantial inter-class variance in pixel levels, as the number of background pixels greatly outnumbers that of foreground pixels, referred to as lymph node pixels. Also, although we sampled negative samples likewise positive sample numbers, this approach does not entirely address the class imbalance problem at the image level. To tackle these problems, we adopted a focal Tversky loss function that put emphasis on the losses by foreground pixels. This emphasis considerably alleviates the impact of losses by background pixels during the training phase, leading to better model performance.

Leveraging the focal Tversky loss helps the model to improve in the learning process compared to training with traditional binary cross entropy (BCE) or focal loss.

The issue of inter and intra-sample level class imbalance remains an unsolved and challenging problem in the computer vision field. Future research is required to address this multifaceted imbalance in this complex setting where both inter-sample (within a sample) and intra-sample (within samples of the entire dataset) class imbalances are intense.

The variance in the matrix of validation graphs also hints toward the model overfitting, as overfitting occurs due to the dominance of the background loss during the training process. Despite using focal Tversky loss to mitigate the class imbalance, the extremely small size of the target organs prevents this problem from being completely resolved. In

other words, the imbalance problem cannot be fully addressed by only relying on advanced loss functions.

Based on the explanation above, we carefully chose the model from a certain epoch during its evaluation on the test set. We opted for a model from an earlier Epoch where the loss function and DSC had consistent downward and upward trends, respectively, specifically from Epoch 34. We extended the training process to Epoch 200 only with the aim of observing the behaviour of the model and its learning procedure.

4.3 Inference and Performance

We evaluated the performance of our models by calculating the DSC and Jaccard Index. Table 4.1. shows the LN segmentation performance of the Attention U-Net model and S-Net models, with and without attention, on the test set. The S-Net adopted from FocusNet outperforms the Attention U-Net even without spatial reverse attention (0.7828 versus 0.7513). Our model also demonstrated better performance than the CaraNet network proposed by Lou et al. (104), with a DSC value of 0.8014 versus 0.7707, respectively. CaraNet stands for Context Axial Reverse Attention Network, an attention-based deep neural network aimed at improving the performance of small object segmentation in medical imaging. It is worth considering that the backbone of CaraNet is pre-trained on ImageNet, a dataset consisting of natural images that differ considerably from medicalimages.

As mentioned before, fewer downsampling leads to better pixel-level prediction and, therefore, better segmentation performance and a higher DSC, as shown 0.8014 for Spatial Context Network with reverse axial attention. Also, by incorporating the reverse axial attention between the encoder and its corresponding decoder layers, the model achieved a higher DSC compared to the baseline model without attention (0.8014 versus 0.7828).

Network	DSC	Jaccard Index
Attention U-Net	0.7513	0.7394
CaraNet	0.7707	0.7602
Spatial Context Network without attention	0.7828	0.7740
Spatial Context Network with reverse axial attention	0.8014	0.78

Table 4.1. Comparing the performance of the Attention U-Net model, CaraNet model and S-Net models.

The Spatial Context Network with reverse axial attention showed enhanced performance even in CT images with multiple LNs. Figure 4.5 indicates an example of the multiple LN segmentation of our model. However, there were instances where the model failed to detect the LNs or generated false positive LN detections. Figures 4.6 and 4.7 represent various segmentation outcomes achieved by our S-Net model.

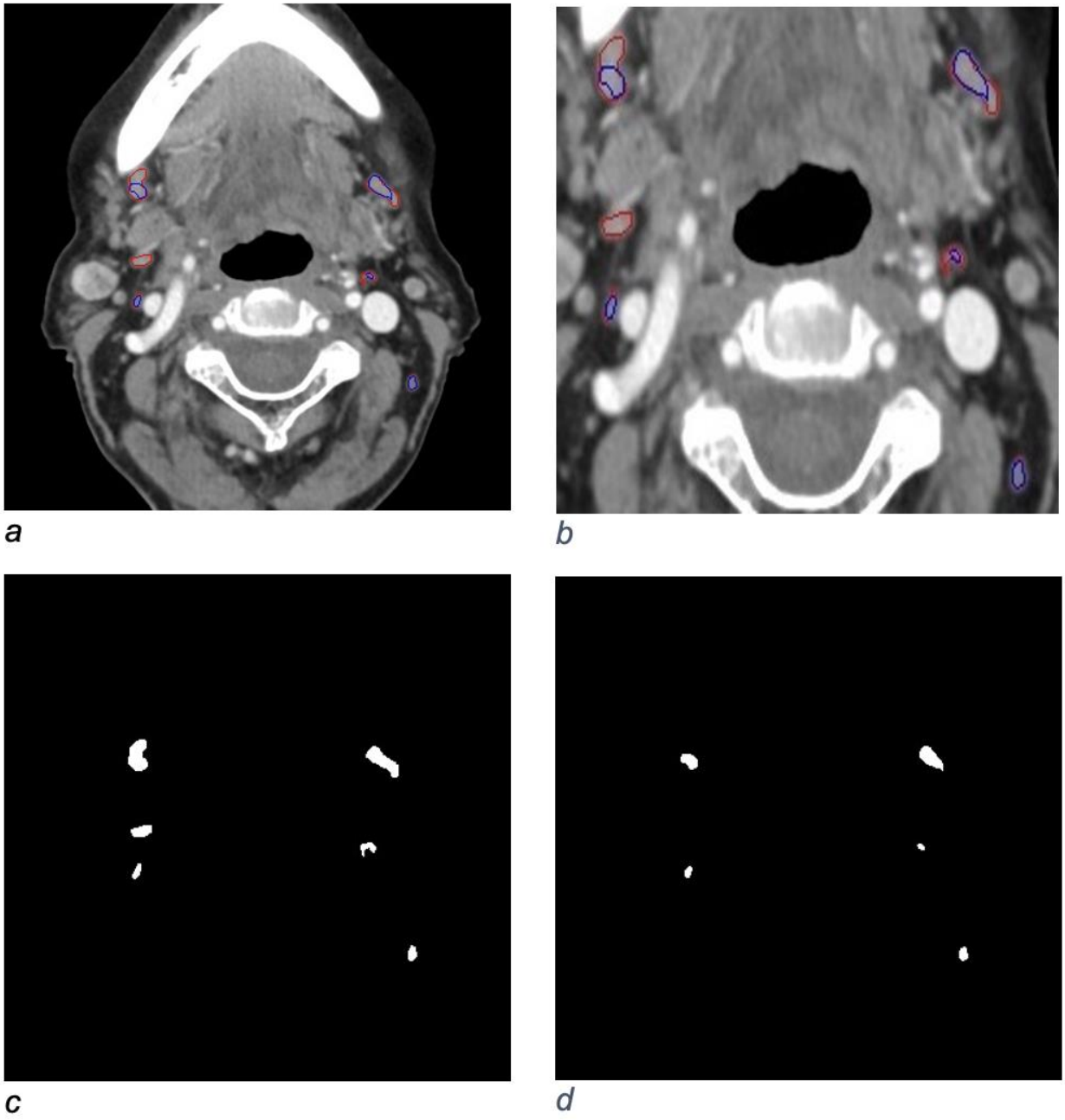

Figure 4.5. Multiple LNs and their segmentation results by spatial context network. (a) Image with segmented LNs. Expert annotations are shown in red and model annotations are shown in blue. (b) The nodal area is zoomed in for visual inspection. (c) Mask image. (d) The output of the model.

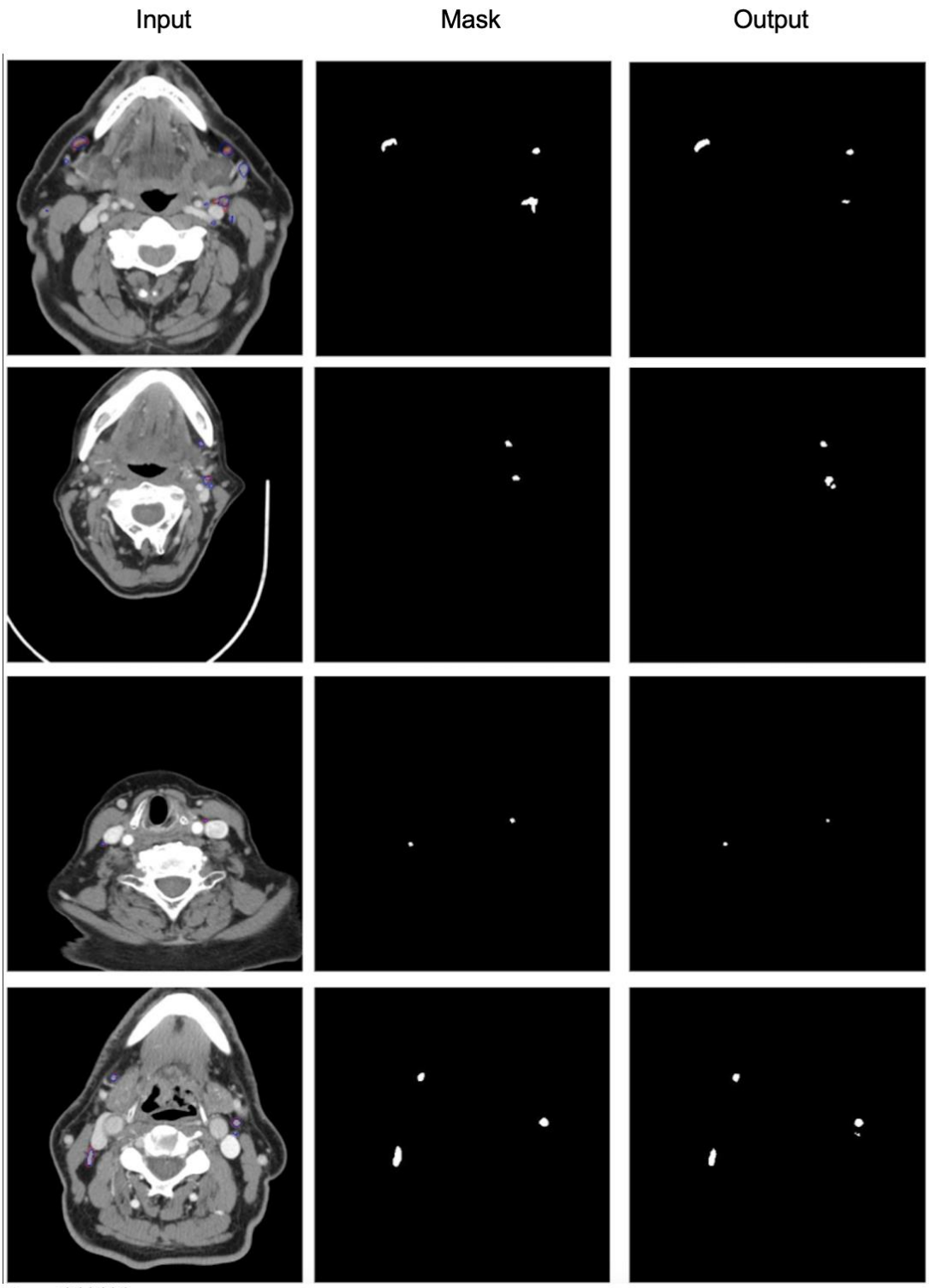

Figure 4.6. Examples of segmentation performance of the S-Net model for Multiple LNs. Mask is the expert annotations and output is the model annotations.

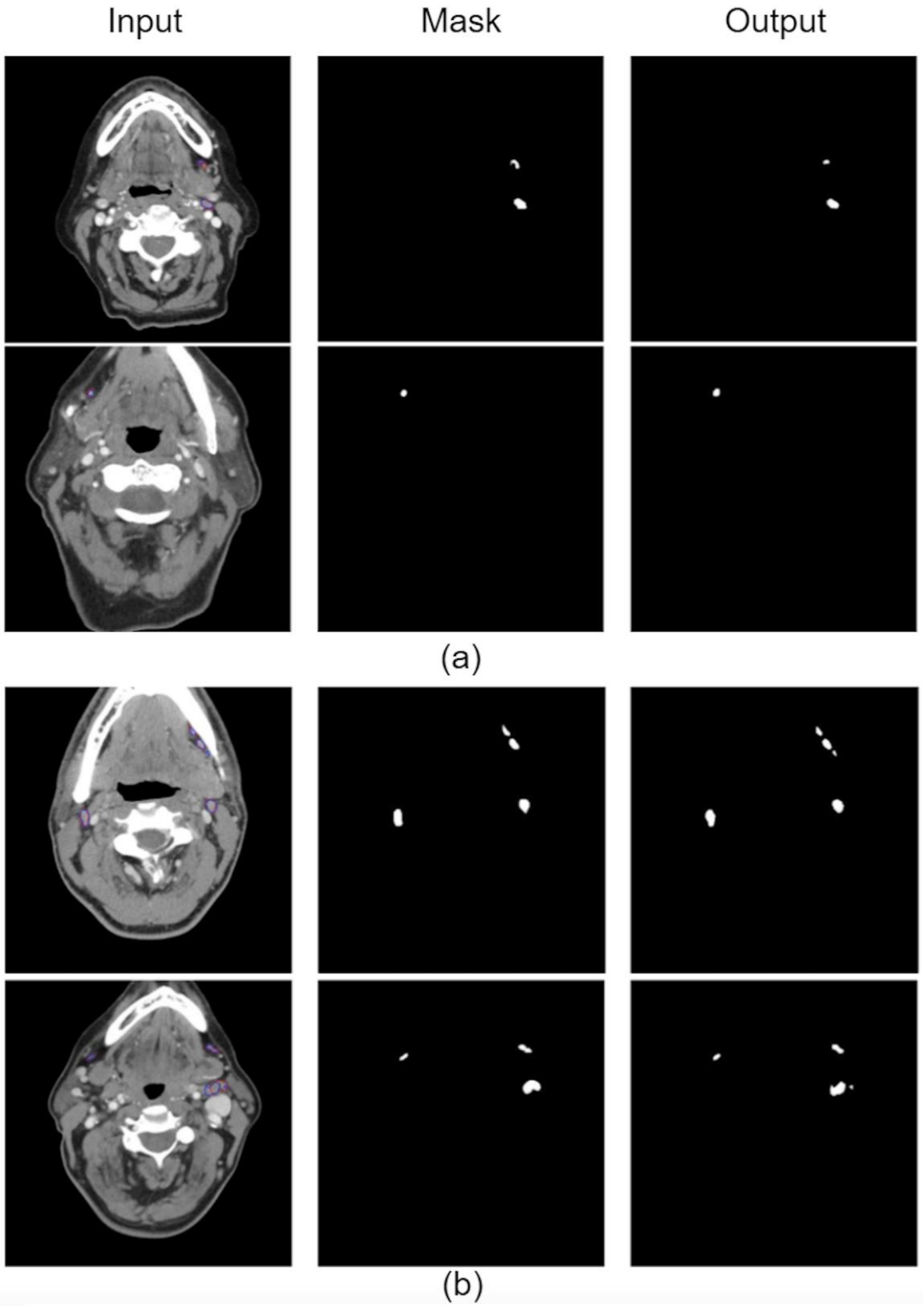

Figure 4.7. Spatial context network prediction for Lymph Node Segmentation: (a) accurate prediction, (b) multiple lymph nodes prediction with low false positive. Continued on next page.

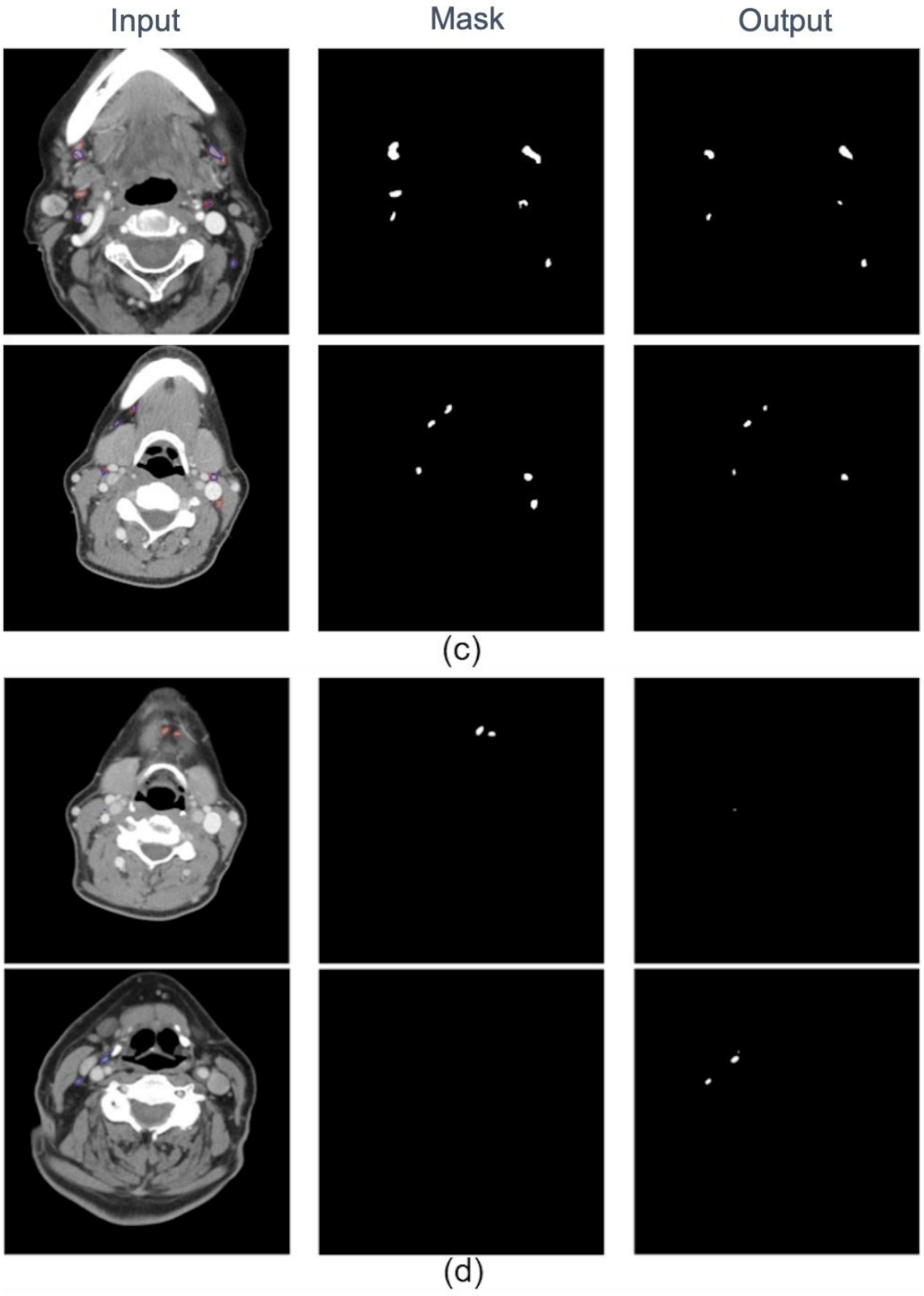

Figure 4.7. (c) multiple lymph nodes prediction with low false positive, and (d) failed cases.

Chapter 5

Discussion

The management of patients with HNSCC presenting clinically negative (N0) necks, where there are no apparent signs of cervical lymph node metastasis upon examination, remains a major challenge. The difficulty arises from the potential existence of occult lymph node metastases that can be detected through neither radiological nor clinical means. Neck dissections are commonly performed on patients with clinically N0 necks and high-risk tumors to tackle this issue, but this approach might result in overtreatment and potential complications. In addition, the accurate analysis of small lymph nodes for disease staging is essential as they may contain micro-metastases. Still, the manual detection and segmentation of cervical lymph nodes can be time-consuming, prone to errors, and reliant on the observer's expertise.

Our research demonstrates the power and potential of DL algorithms in segmenting cervical LNs, a traditionally challenging task considering their small size. This research developed and evaluated a novel DCNN algorithm capable of accurately and efficiently segmenting cervical LNs with a long-axis diameter of ≥ 5mm in levels I to IV from contrast-enhanced CT datasets.

Our research demonstrates the power and potential of DL algorithms in segmenting cervical LNs, a traditionally challenging task considering their small size. This research developed andevaluated a novel DCNN algorithm capable of accurately and efficiently segmenting cervical LNs with a long-axis diameter of ≥ 5mm in levels I to IV from contrast-enhanced CT datasets.

Automatically segmenting LNs using DL algorithms has several potential benefits and impactful applications. It can dramatically improve time efficiency by reducing the labour-intensive process of manual segmentation, allowing clinicians to focus more on diagnosis and treatment planning. It provides consistency in results, eliminating the variability inherent to different human interpreters and thus increasing the reliability of outcomes. Furthermore, it offers scalability that can handle large volumes of data - an essential feature for large-scale studies or busy clinical settings. Importantly, this automated approach lays crucial groundwork for subsequent studies focusing on the development of classification algorithms aimed at facilitating the early detection of subtle LN abnormalities that may not be visible to the human eye. This is particularly pertinent when evaluating small LNs, where the detection of metastasis is traditionally reliant on size. Consequently, our method is poised to bring transformative changes in the assessment of these smaller nodes, providing a more accurate analysis that extends beyond the conventional size-based evaluation. Ultimately, accurate

segmentation and classification could reduce the necessity of invasive procedures like elective neck dissections, decreasing associated complications and morbidity.

A key differentiator of our approach, compared to prior studies like Tomita et al. (90), is the meticulous manual segmentation of regions of interest across all slices of the targeted LNs, rather than relying on a single, largest dimension. Additionally, our technique incorporates texture feature extraction to enhance the precision of our analysis further. These methodological enhancements increase the accuracy and robustness of our model, thus potentially improving LN evaluation in patients.

To ensure high-quality data input, our study concentrated on segmenting the borders of LNs without including surrounding soft tissues, in contrast to the use of arbitrary-sized squares employed in some previous methods. This selective segmentation approach enhances the precision of our DL model by minimizing interference from surrounding structures. A large amount of high-quality data enables the model to learn generalizable patterns and achieve high performance. Poor manual segmentations would inevitably lead to inaccuracy in the model's output. To ensure the best possible output, our LN annotations were meticulously reviewed by two expert neuroradiologists separately and served as the definitive ground truth. Our S-Net imitates how human physicians delineate medical images.

We needed a quantitative evaluation of our algorithm and used the DSC, the most commonly used statistic in the literature, to measure the similarity between the two samples. In a study conducted by Li et al. (22), the U-Net model was employed, achieving an overall DSC value of 0.6586 for LN segmentation in patients diagnosed with nasopharyngeal cancer. In contrast, our S-Net model achieved a significantly higher DSC of 0.8014. This improvement was partly due to the incorporation of the ASPP module into our model, which enlarges the receptive field and facilitates the capture of multi-scale contextual information from the input image. DenseASPP demonstrates to achieve high-performance levels, even when paired witha weak baseline model, leading to considerable improvement of the segmentation performance of the base model (101, 103).

By incorporating reverse axial attention between the encoder and corresponding decoder layers, our S-Net model surpasses the baseline model, achieving superior outcomes with DSCs values of 0.8014 and 0.7828, respectively.

The attention U-Net we employed achieved a DSC of 0.7513 for LN segmentation. This lower performance compared to our S-Net model can be attributed to too much down-sampling in the U-Net and its limitations in representing complex features. The 3D variants of U-Net have contributed to remarkable advancements in medical image segmentation, but they still encounter challenges that cause suboptimal performance

when it comes to small organ segmentation in the head and neck region due to the issues mentioned earlier (101).

We must acknowledge the limitations of our study. These include the use of data from a single source and occasional image quality issues, which may introduce potential bias or impact model performance. Therefore, subtle changes may occur in the image contrast while using other imaging protocols, but if the algorithm parameters are reoptimized, the proposed model can be adapted to these changes. Also, by including these lower-quality images, we mimic the realities of clinical practice and encourage model generalization. Going forward, we intend to evaluate the impact of incorporating data augmentation and expanding our dataset to include scans from other geographical regions and diverse imaging protocols.

Conclusion and Future Work

In this study, we developed an innovative and non-invasive DL-based algorithm for the segmentation of cervical LNs in levels I to IV, demonstrating its potential as a valuable tool in medical diagnostics. Beyond its immediate applicability to LN analysis, this framework could provide a cornerstone for future algorithms designed to classify head and neck LNs by extracting radiomic features, thereby helping to differentiate metastatic from non-metastatic LNs. It could potentially detect small, early-stage nodal metastases that are difficult to discern with the naked eye. This advancement could revolutionize cervical LN assessment in CT scan imaging, a task traditionally challenging for both clinicians and radiologists.

It’s worth noting that our current model relies on 2D CT slices as the input data for analysis. Given the potential benefits of a more comprehensive contextual understanding, one of our future objectives is to enhance our model by incorporating Three-dimensional (3D) voxel data, thereby capturing volumetric information and improving the network's decision-making capabilities.

In summary, this study developed a novel algorithm with strong potential to improve cervical LN segmentation, thus saving clinicians valuable time and potentially improving patient care; in particular, the management of HNSCC patients with clinical N0 necks will be improved by enhancing disease identification and therefore, minimizing the frequency or extent of elective neck dissections.

References

1. Verma OP, Roy S, Pandey SC, Mittal M. Advancement of machine intelligence in interactive medical image analysis: Springer; 2019.
2. Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention; 2015: Springer.
3. Chow LQ. Head and neck cancer. New England Journal of Medicine. 2020;382(1):60-72.
4. Siddique N, Paheding S, Elkin CP, Devabhaktuni V. U-net and its variants for medical image segmentation: A review of theory and applications. Ieee Access. 2021;9:82031-57.
5. Forghani R, Yu E, Levental M, Som PM, Curtin HD. Imaging evaluation of lymphadenopathy and patterns of lymph node spread in head and neck cancer. Expert Review of Anticancer Therapy. 2015;15(2):207-24.
6. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 2017;39(12):2481-95.
7. Commandeur F, Goeller M, Betancur J, Cadet S, Doris M, Chen X, et al. Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast CT. IEEE transactions on medical imaging. 2018;37(8):1835-46.
8. de Souza Figueiredo P, Leite A, Barra F, Dos Anjos R, Freitas A, Nascimento L, et al. Contrast-enhanced CT and MRI for detecting neck metastasis of oral cancer: comparison between analyses performed by oral and medical radiologists. Dentomaxillofacial Radiology. 2012;41(5):396-404.
9. Pfister DG, Spencer S, Adelstein D, Adkins D, Anzai Y, Brizel DM, et al. Head and neck cancers, version 2.2020, NCCN clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network. 2020;18(7):873-98.
10. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-E86.
11. Aupérin A. Epidemiology of head and neck cancers: an update. Current opinion in oncology. 2020;32(3):178-86.
12. Gupta B, Johnson NW, Kumar N. Global epidemiology of head and neck cancers: a continuing challenge. Oncology. 2016;91(1):13-23.
13. Barbu A, Suehling M, Xu X, Liu D, Zhou SK, Comaniciu D. Automatic detection and segmentation of lymph nodes from CT data. IEEE Transactions on Medical Imaging. 2011;31(2):240-50.

14. Forghani R, Chatterjee A, Reinhold C, Pérez-Lara A, Romero-Sanchez G, Ueno Y, et al. Head and neck squamous cell carcinoma: prediction of cervical lymph node metastasis by dual-energy CT texture analysis with machine learning. European radiology. 2019;29:6172-81.
15. Hoang JK, Vanka J, Ludwig BJ, Glastonbury CM. Evaluation of cervical lymph nodes in head and neck cancer with CT and MRI: tips, traps, and a systematic approach. American Journal of Roentgenology. 2013;200(1):W17-W25.
16. Bardosi ZR, Dejaco D, Santer M, Kloppenburg M, Mangesius S, Widmann G, et al. Benchmarking Eliminative Radiomic Feature Selection for Head and Neck Lymph Node Classification. Cancers. 2022;14(3):477.
17. De Bree R, Takes RP, Castelijns JA, Medina JE, Stoeckli SJ, Mancuso AA, et al. Advances in diagnostic modalities to detect occult lymph node metastases in head and neck squamous cell carcinoma. Head & neck. 2015;37(12):1829-39.
18. Paleri V, Urbano T, Mehanna H, Repanos C, Lancaster J, Roques T, et al. Management of neck metastases in head and neck cancer: United Kingdom National Multidisciplinary Guidelines. The Journal of Laryngology & Otology. 2016;130(S2):S161-S9.
19. Kelly HR, Curtin HD, editors. Squamous cell carcinoma of the head and neck—imaging evaluation of regional lymph nodes and implications for management. Seminars in Ultrasound, CT and MRI; 2017: Elsevier.
20. Ho T-Y, Chao C-H, Chin S-C, Ng S-H, Kang C-J, Tsang N-M. Classifying neck lymph nodes of head and neck squamous cell carcinoma in MRI images with radiomic features. Journal of Digital Imaging. 2020;33(3):613-8.
21. Li Z, Xia Y. Deep reinforcement learning for weakly-supervised lymph node segmentation in CT images. IEEE Journal of Biomedical and Health Informatics. 2020;25(3):774-83.
22. Li S, Xiao J, He L, Peng X, Yuan X. The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods. Technology in cancer research & treatment. 2019;18:1533033819884561.
23. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, Van Ginneken B, Madabhushi A, et al. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE. 2021;109(5):820-38.
24. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA: a cancer journal for clinicians. 2019;69(2):127-57.
25. Kann BH, Aneja S, Loganadane GV, Kelly JR, Smith SM, Decker RH, et al. Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. Scientific reports. 2018;8(1):1-11.
26. Reticker-Flynn NE, Zhang W, Belk JA, Basto PA, Escalante NK, Pilarowski GO, et al. Lymph node colonization induces tumor-immune tolerance to promote distant metastasis. Cell. 2022;185(11):1924-42. e23.

27. Rassy E, Nicolai P, Pavlidis N. Comprehensive management of HPV-related squamous cell carcinoma of the head and neck of unknown primary. Head & Neck. 2019;41(10):3700-11.
28. Jones T, De M, Foran B, Harrington K, Mortimore S. Laryngeal cancer: United Kingdom national multidisciplinary guidelines. The Journal of Laryngology & Otology. 2016;130(S2):S75-S82.
29. Arain AA, Rajput MSA, Ansari SA, Mahmood Z, Ahmad AN, Dogar MR, et al. Occult nodal metastasis in oral cavity cancers. Cureus. 2020;12(11).
30. Ferlito A, Silver CE, Rinaldo A. Elective management of the neck in oral cavity squamous carcinoma: current concepts supported by prospective studies. British Journal of Oral and Maxillofacial Surgery. 2009;47(1):5-9.
31. Stodulski D, Mikaszewski B, Majewska H, Wiśniewski P, Stankiewicz C. Probability and pattern of occult cervical lymph node metastases in primary parotid carcinoma. European Archives of Oto-Rhino-Laryngology. 2017;274(3):1659-64.
32. Gontarz M, Bargiel J, Gąsiorowski K, Marecik T, Szczurowski P, Zapała J, et al. Epidemiology of Primary Epithelial Salivary Gland Tumors in Southern Poland—A 26-Year, Clinicopathologic, Retrospective Analysis. Journal of Clinical Medicine. 2021;10(8):1663.
33. Iuga A-I, Carolus H, Höink AJ, Brosch T, Klinder T, Maintz D, et al. Automated detection and segmentation of thoracic lymph nodes from CT using 3D foveal fully convolutional neural networks. BMC Medical Imaging. 2021;21(1):1-12.
34. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical image analysis. 2017;42:60-88.
35. Melonakos J. Geodesic tractography segmentation for directional medical image analysis: Georgia Institute of Technology; 2008.
36. Bankman I. Handbook of medical image processing and analysis: Elsevier; 2008.
37. Ahamed KU, Islam M, Uddin A, Akhter A, Paul BK, Yousuf MA, et al. A deep learning approach using effective preprocessing techniques to detect COVID-19 from chest CT-scan and X-ray images. Computers in biology and medicine. 2021;139:105014.
38. Allen G. Understanding AI technology. Joint Artificial Intelligence Center (JAIC) The Pentagon United States, 2020.
39. Deng L, Yu D. Deep learning: methods and applications. Foundations and trends® in signal processing. 2014;7(3–4):197-387.
40. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. Radiographics. 2017;37(7):2113-31.
41. Bengio Y. Learning deep architectures for AI. Foundations and trends® in Machine Learning. 2009;2(1):1-127.
42. Kooi T, Litjens G, Van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. Medical image analysis. 2017;35:303-12.

43. Kann BH, Hicks DF, Payabvash S, Mahajan A, Du J, Gupta V, et al. Multi-institutional validation of deep learning for pretreatment identification of extranodal extension in head and neck squamous cell carcinoma. Journal of Clinical Oncology. 2020;38(12):1304-11.
44. Ji S, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. IEEE transactions on pattern analysis and machine intelligence. 2012;35(1):221-31.
45. Baumgartner CF, Oktay O, Rueckert D. Fully convolutional networks in medical imaging: Applications to image enhancement and recognition. Deep Learning and Convolutional Neural Networks for Medical Image Computing: Springer; 2017. p. 159-79.
46. Lu L, Zheng Y, Carneiro G, Yang L. Deep learning and convolutional neural networks for medical image computing. Advances in computer vision and pattern recognition. 2017;10:978-3.
47. Le WT, Maleki F, Romero FP, Forghani R, Kadoury S. Overview of machine learning: part 2: deep learning for medical image analysis. Neuroimaging Clinics. 2020;30(4):417-31.
48. Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. Medical image analysis. 2017;41:40-54.
49. Belal SL, Sadik M, Kaboteh R, Enqvist O, Ulén J, Poulsen MH, et al. Deep learning for segmentation of 49 selected bones in CT scans: first step in automated PET/CT-based 3D quantification of skeletal metastases. European journal of radiology. 2019;113:89-95.
50. Guo Y, Gao Y, Shen D. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. IEEE transactions on medical imaging. 2015;35(4):1077-89.
51. Xu X, Zhou F, Liu B. Automatic bladder segmentation from CT images using deep CNN and 3D fully connected CRF-RNN. International journal of computer assisted radiology and surgery. 2018;13(7):967-75.
52. Cancer Facts & Figures 2022: American Cancer Society; 2022. Available from: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2022.html.
53. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. Sustainability. 2021;13(3):1224.
54. López-Linares Román K, García Ocaña MI, Lete Urzelai N, González Ballester MÁ, Macía Oliver I. Medical image segmentation using deep learning. Deep Learning in Healthcare: Springer; 2020. p. 17-31.
55. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging. 2019;32(4):582-96.
56. Pham DL, Xu C, Prince JL. Current methods in medical image segmentation. Annual review of biomedical engineering. 2000;2(1):315-37.
57. Munir K, Elahi H, Ayub A, Frezza F, Rizzi A. Cancer diagnosis using deep learning: a bibliographic review. Cancers. 2019;11(9):1235.
58. Samber DD, Ramachandran S, Sahota A, Naidu S, Pruzan A, Fayad ZA, et al. Segmentation of carotid arterial walls using neural networks. World Journal of Radiology. 2020;12(1):1.

59. Badrigilan S, Nabavi S, Abin AA, Rostampour N, Abedi I, Shirvani A, et al. Deep learning approaches for automated classification and segmentation of head and neck cancers and brain tumors in magnetic resonance images: a meta-analysis study. International journal of computer assisted radiology and surgery. 2021;16(4):529-42.
60. Haque IRI, Neubert J. Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked. 2020;18:100297.
61. Hamwood J, Schmutz B, Collins MJ, Allenby MC, Alonso-Caneiro D. A deep learning method for automatic segmentation of the bony orbit in MRI and CT images. Scientific Reports. 2021;11(1):1-12.
62. Long J, Shelhamer E, Darrell T, editors. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.
63. Myronenko A, editor 3D MRI brain tumor segmentation using autoencoder regularization. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4; 2019: Springer.
64. Edupuganti VG, Chawla A, Kale A, editors. Automatic optic disk and cup segmentation of fundus images using deep learning. 2018 25th IEEE international conference on image processing (ICIP); 2018: IEEE.
65. Anthimopoulos M, Christodoulidis S, Ebner L, Geiser T, Christe A, Mougiakakou S. Semantic segmentation of pathological lung tissue with dilated fully convolutional networks. IEEE journal of biomedical and health informatics. 2018;23(2):714-22.
66. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:170406857. 2017.
67. Alqazzaz S, Sun X, Yang X, Nokes L. Automated brain tumor segmentation on multi-modal MR image using SegNet. Computational Visual Media. 2019;5:209-19.
68. Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. BMC Medical Imaging. 2021;21(1):1-10.
69. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, editors. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International conference on medical image computing and computer-assisted intervention; 2016: Springer.
70. Chen H, Qi X, Yu L, Heng P-A, editors. DCAN: deep contour-aware networks for accurate gland segmentation. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition; 2016.
71. Zhang D, Lin Y, Chen H, Tian Z, Yang X, Tang J, et al. Deep learning for medical image segmentation: tricks, challenges and future directions. arXiv preprint arXiv:220910307. 2022.
72. Kayalibay B, Jensen G, van der Smagt P. CNN-based segmentation of medical imaging data. arXiv preprint arXiv:170103056. 2017.

73. Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. IEEE Transactions on Medical Imaging. 2020;39(11):3679-90.
74. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging. 2019;32:582-96.
75. Golan R, Jacob C, Denzinger J, editors. Lung nodule detection in CT images using deep convolutional neural networks. 2016 international joint conference on neural networks (IJCNN); 2016: IEEE.
76. Milletari F, Ahmadi S-A, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. Computer Vision and Image Understanding. 2017;164:92-102.
77. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, editors. 3D U-Net: learning dense volumetric segmentation from sparse annotation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19; 2016: Springer.
78. Christ PF, Ettlinger F, Grün F, Elshaera MEA, Lipkova J, Schlecht S, et al. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv preprint arXiv:170205970. 2017.
79. Merkow J, Marsden A, Kriegman D, Tu Z, editors. Dense volume-to-volume vascular boundary detection. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III 19; 2016: Springer.
80. Ciresan D, Giusti A, Gambardella L, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. Advances in neural information processing systems. 2012;25.
81. Justus D, Brennan J, Bonner S, McGough AS, editors. Predicting the computational cost of deep learning models. 2018 IEEE international conference on big data (Big Data); 2018: IEEE.
82. Baumgartner CF, Koch LM, Pollefeys M, Konukoglu E, editors. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. Statistical Atlases and Computational Models of the Heart ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8; 2018: Springer.
83. Zhao Y, Li H, Wan S, Sekuboyina A, Hu X, Tetteh G, et al. Knowledge-aided convolutional neural network for small organ segmentation. IEEE journal of biomedical and health informatics. 2019;23(4):1363-73.
84. Ariji Y, Fukuda M, Kise Y, Nozawa M, Yanashita Y, Fujita H, et al. Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence. Oral surgery, oral medicine, oral pathology and oral radiology. 2019;127(5):458-63.

85. Zhou Z, Chen L, Sher D, Zhang Q, Shah J, Pham N-L, et al., editors. Predicting lymph node metastasis in head and neck cancer by combining many-objective radiomics and 3-dimensioal convolutional neural network through evidential reasoning. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018: IEEE.
86. Lee JH, Ha EJ, Kim JH. Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. European radiology. 2019;29(10):5452-7.
87. Chen L, Zhou Z, Sher D, Zhang Q, Shah J, Pham N-L, et al. Combining many-objective radiomics and 3D convolutional neural network through evidential reasoning to predict lymph node metastasis in head and neck cancer. Physics in Medicine & Biology. 2019;64(7):075011.
88. Ariji Y, Fukuda M, Nozawa M, Kuwada C, Goto M, Ishibashi K, et al. Automatic detection of cervical lymph nodes in patients with oral squamous cell carcinoma using a deep learning technique: a preliminary study. Oral Radiology. 2021;37(2):290-6.
89. Ariji Y, Kise Y, Fukuda M, Kuwada C, Ariji E. Segmentation of metastatic cervical lymph nodes from CT images of oral cancers using deep-learning technology. Dentomaxillofacial Radiology. 2022;51(4):20210515.
90. Tomita H, Yamashiro T, Heianna J, Nakasone T, Kobayashi T, Mishiro S, et al. Deep learning for the preoperative diagnosis of metastatic cervical lymph nodes on contrast-enhanced computed tomography in patients with oral squamous cell carcinoma. Cancers. 2021;13(4):600.
91. Schwartz L, Bogaerts J, Ford R, Shankar L, Therasse P, Gwyther S, et al. Evaluation of lymph nodes with RECIST 1.1. European journal of cancer. 2009;45(2):261-7.
92. Manjunatha Y, Sharma V, Iwahori Y, Bhuyan M, Wang A, Ouchi A, et al. Lymph node detection in CT scans using modified U-Net with residual learning and 3D deep network. International Journal of Computer Assisted Radiology and Surgery. 2023:1-10.
93. Cai J, Tang Y, Lu L, Harrison AP, Yan K, Xiao J, et al., editors. Accurate weakly-supervised deep lesion segmentation using large-scale clinical annotations: Slice-propagated 3d mask generation from 2d recist. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11; 2018: Springer.
94. Sartor H, Minarik D, Enqvist O, Ulén J, Wittrup A, Bjurberg M, et al. Auto-segmentations by convolutional neural network in cervical and anorectal cancer with clinical structure sets as the ground truth. Clinical and Translational Radiation Oncology. 2020;25:37-45.
95. Zhou X, Takayama R, Wang S, Hara T, Fujita H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. Medical physics. 2017;44(10):5221-33.
96. Murphy A. Windowing (CT): Radiopaedia; 2023. Available from: https://radiopaedia.org/articles/windowing-ct.
97. Ranjbarzadeh R, Bagherian Kasgari A, Jafarzadeh Ghoushchi S, Anari S, Naseri M, Bendechache M. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. Scientific Reports. 2021;11(1):1-17.

98. Islam M, Vibashan V, Jose VJM, Wijethilake N, Utkarsh U, Ren H, editors. Brain tumor segmentation and survival prediction using 3D attention UNet. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5; 2020: Springer.
99. Noori M, Bahri A, Mohammadi K, editors. Attention-guided version of 2D UNet for automatic brain tumor segmentation. 2019 9th international conference on computer and knowledge engineering (ICCKE); 2019: IEEE.
100. Jadon S, editor A survey of loss functions for semantic segmentation. 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB); 2020: IEEE.
101. Gao Y, Huang R, Yang Y, Zhang J, Shao K, Tao C, et al. FocusNetv2: Imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images. Medical Image Analysis. 2021;67:101831.
102. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence. 2017;40(4):834-48.
103. Yang M, Yu K, Zhang C, Li Z, Yang K, editors. Denseaspp for semantic segmentation in street scenes. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.
104. Lou A, Guan S, Ko H, Loew MH, editors. CaraNet: context axial reverse attention network for segmentation of small medical objects. Medical Imaging 2022: Image Processing; 2022: SPIE.
105. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, et al., editors. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22; 2019: Springer.
106. Thada V, Jaglan V. Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. International Journal of Innovations in Engineering and Technology. 2013;2(4):202-5.
107. Pribadi FS, Adji TB, Permanasari AE, editors. Automated short answer scoring using weighted cosine coefficient. 2016 IEEE Conference on e-Learning, e-Management and e-Services (IC3e); 2016: IEEE.