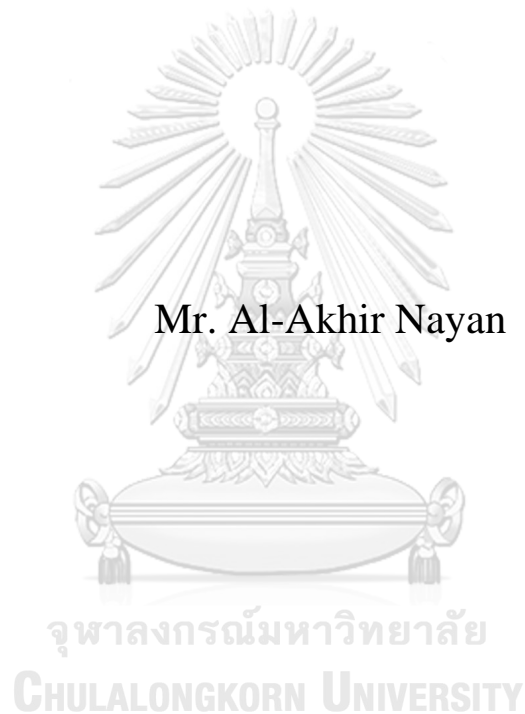


Mediastinal Lymph Node Detection and Segmentation using Deep Learning



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Computer Engineering
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2022
Copyright of Chulalongkorn University



นายอล เอเคอร์ นายาน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	Mediastinal Lymph Node Detection and Segmentation using Deep Learning
By	Mr. Al-Akhir Nayan
Field of Study	Computer Engineering
Thesis Advisor	BOONSERM KIJSIRIKUL
Thesis Co Advisor	Yuji Iwahori

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Engineering

Dean of the FACULTY OF
ENGINEERING
(SUPOT TEACHAVORASINSKUN)

THESIS COMMITTEE

----- (SUKREE SINTHUPINYO)	Chairman
----- (BOONSERM KIJSIRIKUL)	Thesis Advisor
----- (Yuji Iwahori)	Thesis Co-Advisor
----- (Nuanwan Soonthornphisaj)	External Examiner



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อด เอเคอร์ นายน :- (Mediastinal Lymph Node Detection and Segmentation using Deep Learning) อ.ที่ปรึกษาหลัก : บุญเสริม กิจศิริกุล, อ.ที่ปรึกษาร่วม : ยูจิ ไอวาโฮริ

การแบ่งกลุ่มและตรวจจับต่อมน้ำเหลืองสำหรับภาวะมะเร็งของโรคมะเร็งแบบอัตโนมัติเป็นเรื่องที่ซับซ้อน ในทางการแพทย์ (ในระบอบที่ต่อมน้ำเหลืองผู้ป่วย) การถ่ายภาพรังสีโดยใช้คอมพิวเตอร์ (CT) และการถ่ายภาพรังสีด้วยการปล่อยโพซิตรอน (PET) สามารถใช้ในการระบุต่อมน้ำเหลืองที่มีลักษณะผิดปกติได้ ถึงแม้ว่าวิธีการดังกล่าวก็มีข้อจำกัดในเรื่องของการจำแนกความแตกต่างและความหลากหลายในเชิงขนาดและรูปร่างของต่อมน้ำเหลือง การแบ่งกลุ่มต่อมน้ำเหลืองยังคงเป็นงานที่ท้าทาย วิธีการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Deep Convolutional Neural Network) จึงมักเป็นวิธีที่ใช้ในการแบ่งกลุ่มสิ่งต่างๆ บนภาพถ่ายสำหรับทางการแพทย์ เทคนิคที่ทันสมัยต่างๆ จึงมักทำลายความละเอียดของภาพผ่านเทคนิคการกรองข้อมูลของรูปภาพ เช่น การทำ Image Pooling และ Image Convolution ด้วยเหตุนี้เอง รูปแบบที่ถูกสร้างขึ้นเหล่านั้น จึงมักให้ผลลัพธ์ที่ไม่น่าพึงพอใจ เมื่อคิดถึงปัญหาเหล่านี้ต่อไป เทคนิค UNet++ ที่เป็นเทคนิคการเรียนรู้ขั้นสูงที่เป็นที่รู้จักกันอย่างแพร่หลาย จึงถูกปรับปรุงโดยการนำเทคนิคการสุ่มระดับสีจากรูปภาพเดิมอย่างเทคนิค Bilinear Interpolation พร้อมทั้งกลยุทธ์การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่างๆ อย่างกลยุทธ์ Total Generalized Variation-based Upsampling มาใช้ในการแบ่งกลุ่มและตรวจจับกลุ่มต่อมน้ำเหลืองบริเวณทรวงอกด้านหน้า (Mediastinal Lymph Nodes) เทคนิค UNet++ ที่ถูกปรับปรุงนี้ยังคงให้สัมพัทธ์ถึงความที่ไม่ต่อเนื่อง เลือกใช้บริเวณที่มีจุดรบกวนในภาพ และค้นหาจุดสมดุลที่เหมาะสมผ่านเทคนิคการหาโครงข่ายประสาทที่เชื่อมแบบย้อนกลับ (Backpropagation) และสร้างความละเอียดของภาพขึ้นมาใหม่ งานวิจัยนี้ได้รวบรวมข้อมูลภาพถ่ายรังสีจาก TCIA จากกลุ่มผู้ป่วย 5 รายและใช้ชุดข้อมูลสาธารณะอย่าง ELCAP ซึ่งเป็นชุดข้อมูลที่ถูกจัดเตรียมขึ้นโดยผู้เชี่ยวชาญทางการแพทย์หลากหลายท่าน และใช้ข้อมูลเหล่านี้เป็นพื้นฐานในการวิเคราะห์และประมวลผลด้วยเทคนิค UNet++ การผสมผสานข้อมูลที่แตกต่างกันถึง 3 แบบถูกใช้เพื่อทดสอบ ผลปรากฏว่า รูปแบบที่ได้จากเทคนิคดังกล่าว ให้ผลลัพธ์ที่มีความแม่นยำถึง 94.8% และได้ดัชนี Jaccard 91.9% อีกทั้งยังมีอัตราการเรียกคืนข้อมูล (Recall) สูงถึง 94.1% และมีค่า Precision สูงถึง 93.1% จากข้อมูลทั้ง 3 ชุด สมรรถภาพของมันถูกวัดด้วยชุดข้อมูลที่แตกต่างกันถึง 3 ชุดข้อมูล รวมถึงยังมีการเปรียบเทียบสมรรถนะของมันกับเทคนิคทันสมัยอื่นๆ อีกด้วย และผลปรากฏว่า รูปแบบที่ได้จาก UNet++ ที่ผสมผสานการใช้กลยุทธ์ต่างๆ ที่กล่าวมาข้างต้นนั้น ทำงานได้ดีกว่ารูปแบบที่ได้จากเทคนิคอื่นๆ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2565

ลายมือชื่อนิสิต
ลายมือชื่อ อ.ที่ปรึกษาหลัก
ลายมือชื่อ อ.ที่ปรึกษาร่วม

6470306121 : MAJOR COMPUTER ENGINEERING

KEYWORD: Lymph Node, Segmentation, Detection, Deep Learning, Mediastinal Lymph Node, UNet++

AI-Akhir Nayan : Mediastinal Lymph Node Detection and Segmentation using Deep Learning. Advisor: BOONSERM KIJSIRIKUL Co-advisor: Yuji Iwahori

Automatic lymph node (LN) segmentation and detection for cancer staging are critical. In clinical practice, computed tomography (CT) and positron emission tomography (PET) imaging detect abnormal LNs. Despite its low contrast and variety in nodal size and form, LN segmentation remains a challenging task. Deep convolutional neural networks frequently segment items in medical photographs. Most state-of-the-art techniques destroy image's resolution through pooling and convolution. As a result, the models provide unsatisfactory results. Keeping the issues in mind, a well-established deep learning technique UNet++ was modified using bilinear interpolation and total generalized variation (TGV) based upsampling strategy to segment and detect mediastinal lymph nodes. The modified UNet++ maintains texture discontinuities, selects noisy areas, searches appropriate balance points through backpropagation, and recreates image resolution. Collecting CT image data from TCIA, 5-patients, and ELCAP public dataset, a dataset was prepared with the help of experienced medical experts. The UNet++ was trained using those datasets, and three different data combinations were utilized for testing. Utilizing the proposed approach, the model achieved 94.8% accuracy, 91.9% Jaccard, 94.1% recall, and 93.1% precision on COMBO_3. The performance was measured on different datasets and compared with state-of-the-art approaches. The UNet++ model with hybridized strategy performed better than others.



Field of Study: Computer Engineering
Academic Year: 2022

Student's Signature
Advisor's Signature
Co-advisor's Signature

ACKNOWLEDGEMENTS

I would like to extend my gratitude to Professor Dr. Boonserm Kijirikul and Professor Dr. Yuji Iwahori for their invaluable guidance, support, and instruction throughout my study. Furthermore, I wish to thank Chulalongkorn University for awarding me the ASEAN or Non-ASEAN Countries Graduate Scholarship and the Japan Society for the Promotion of Science (JSPS) for funding this thesis. This thesis was supported by the JSPS Grant-in-Aid Scientific Research (C) under Yuji Iwahori's Grant number 20K11873.

Al-Akhir Nayan



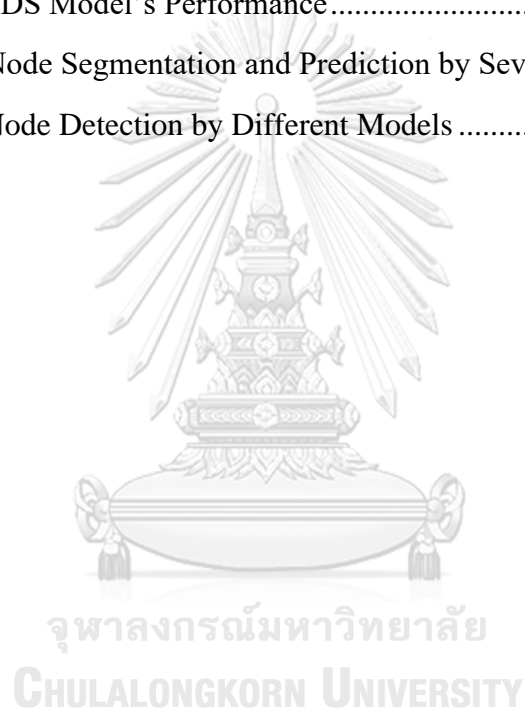
TABLE OF CONTENTS

	Page
.....	iii
ABSTRACT (THAI)	iii
.....	iv
ABSTRACT (ENGLISH)	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: RELATED WORKS	3
CHAPTER 3: RESEARCH METHOD	6
1. Dataset	6
A. Data Processing	6
B. Data Augmentation	6
2. Unet Model Architecture & Upsampling Technique	7
3. Proposed Model Architecture	9
4. Proposed Upsampling Technique	10
5. Backpropagation	12
6. Evaluation Process	12
7. Lymph Node Segmentation Criteria	13
CHAPTER 4: RESULT AND ANALYSIS	14
1. Experimental Environment Setup	14
2. Performance Measurement on TCIA Dataset	14
A. Training Accuracy Measurement	14
B. Testing Accuracy Measurement	15

i.	Experimental Analysis on COMBO_1	15
ii.	Experimental Analysis on COMBO_2	16
iii.	Experimental Analysis on COMBO_3	16
3.	Performance Measurement on ELCAP Dataset.....	18
A.	Training Accuracy Measurement	18
B.	Testing Accuracy Measurement.....	18
i.	Experimental Analysis on COMBO_1	18
ii.	Experimental Analysis on COMBO_2	19
iii.	Experimental Analysis on COMBO_3	19
4.	Performance Measurement on 5-Patients Dataset	19
A.	Training Accuracy Measurement	19
B.	Testing Accuracy Measurement.....	21
i.	Experimental Analysis on COMBO_1	21
ii.	Experimental Analysis on COMBO_2	21
iii.	Experimental Analysis on COMBO_3	21
5.	Measurement Summary	23
6.	Performance Comparison	23
A.	Performance Measurement using Old_U-Net	24
B.	Performance Measurement using AlexNet.....	24
C.	Performance Measurement using SegNet	25
D.	Performance Measurement using Auto-LNDS (Modified Mask RCNN)	25
7.	Detection, Segmentation, and Precision Capability Analysis.....	26
CHAPTER 5: DISCUSSION & LIMITATION		29
CHAPTER 6: CONCLUSION		30
REFERENCES		31
VITA.....		36

LIST OF TABLES

	Page
Table 1. Model Parameters	14
Table 2. Old_U-Net Model's Performance	24
Table 3. AlexNet Model's Performance	25
Table 4. SegNet Model's Performance	25
Table 5. Auto-LNDS Model's Performance	26
Table 6. Lymph Node Segmentation and Prediction by Several Models	27
Table 7. Lymph Node Detection by Different Models	28



LIST OF FIGURES

	Page
Figure 1. Dataset Sample a) Image, b) Mask.....	6
Figure 2. UNet Model Architecture	7
Figure 3. An Example of Transpose Convolution	8
Figure 4. Modified UNet++ Model Architecture.....	9
Figure 5. Artifact issue created by the transposed convolution: (a) checkerboard problem, (b) uneven overlap	10
Figure 6. Example of bilinear interpolation.....	11
Figure 7. Training Score Evaluation a) Train Accuracy vs. Loss, b) Validation Accuracy vs. Loss, c) Intersect of Union (IOU), d) Dice Coefficient, e) Recall.....	15
Figure 8. Testing Accuracy Measurement on Test Dataset COMBO_1, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	16
Figure 9. Testing Accuracy Measurement on Test Dataset COMBO_2, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	17
Figure 10. Testing Accuracy Measurement on Test Dataset COMBO_3, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	17
Figure 11. Training Score Evaluation a) Train Accuracy vs. Loss, b) Validation Accuracy vs. Loss, c) Intersect of Union (IOU), d) Dice Coefficient, e) Recall.....	18
Figure 12. Testing Accuracy Measurement on Test Dataset COMBO_1, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	19
Figure 13. Testing Accuracy Measurement on Test Dataset COMBO_2, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	20
Figure 14. Testing Accuracy Measurement on Test Dataset COMBO_3, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	20
Figure 15. Training Score Evaluation a) Train Accuracy vs. Loss, b) Validation Accuracy vs. Loss, c) Intersect of Union (IOU), d) Dice Coefficient, e) Recall.....	21
Figure 16. Testing Accuracy Measurement on Test Dataset COMBO_1, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	22
Figure 17. Testing Accuracy Measurement on Test Dataset COMBO_2, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1	22

Figure 18. Testing Accuracy Measurement on Test Dataset COMBO_3, a) Accuracy,
b) Jaccard, c) Precision, d) Recall, e) F123



CHAPTER 1: INTRODUCTION

With 59.9 million cases of death each year, cancer holds the record for the highest mortality rate [1]. The percentage of cancer patient's survival rate is minimized with time. UK Cancer Research Centre found that just 5% of patients would survive if they had been ill for ten years, whereas approximately 30% would survive if they had been sick for a year [2]. Therefore, early discovery, diagnosis, and treatment are essential to increase the cancer survival rate. CT scans have benefited cancer studies due to the advancement of radiology in recent years. Lymph nodes (LNs) significantly manage cancer and evaluate therapeutic responses [3]. However, segmenting and detecting LNs on CT images is challenging due to their oblique boundary and low contrast with surrounding tissue and organs [4]. In addition, manually assessing and counting LNs in medical images by human observers is time-consuming and prone to inaccuracy [5].

Recent years have seen a rise in the use of deep learning algorithms for medical image analysis. It performs better than conventional statistical and atlas-based machine learning algorithms and maintains good speed and accuracy [6]. Fully convolutional networks (FCNs) have been prominent in applications for medical image segmentation [7]. Following FCNs, several different convolutional neural network-based segmentation architectures have been developed, including SegNet [8], U-Net [9], and DeepLab [10]. However, there are not many published works on segmenting LNs using deep learning. FCNs were taught to recognize lymph node clusters or contour probabilistic output maps in [11], and conditional random fields were used to segment lymph nodes. A 3D U-Net was utilized to partition mediastinal LNs and other anatomical features such as lungs, airways, aortic arches, and pulmonary arteries to regulate the balance of sizes between the target classes in [12].

Deep learning algorithms in the field of medical image segmentation face significant challenges. The need for radiologists to name the anatomical objects in the images makes it somewhat expensive to construct a vast dataset (like ImageNet [13]). Additionally, when the deep segmental neural network is trained, some insignificant or peculiar observations will cause label imbalance. Due to insufficient datasets and an excellent segmentation approach, previous investigations on mediastinal LN segmentation reveal several problems. Additionally, most research has given minor importance to the small but significant portion of medical images and ignored information loss issues brought on by the max pool and convolution process. Those models' performance is subpar due to such ignorance.

This thesis addresses the shortcomings of earlier strategies by proposing a new strategy for medical image segmentation and detection. A dataset was created with the assistance of experienced medical experts. The proposed method uses the UNet++ model, which introduces a trainable upsampling technique for restoring the image resolution lost during the max pooling and convolution process. This tactic enables an information-loss-free training method for neurons. The hybridized UNet++ model was tested using various combinations of multiple datasets, including TCIA, 5-patients,

and the public dataset ELCAP, and compared to state-of-the-art methods such as Auto LNDS, SegNet, and AlexNet. The best approach was determined using accuracy, precision, dice score, recall, and F1.

The thesis is structured into different sections. Section 2 discusses related research works, their benefits, and their downsides. Section 3 explains the architecture of the improved model, the suggested upsampling method, and the evaluation procedure. Section 4 discusses the model's performance on several datasets and compares it with other cutting-edge techniques. Section 5 explains the limitations of the thesis. Finally, in Section 6, the entire thesis is summarized and concluded.



CHAPTER 2: RELATED WORKS

Previously, many studies on computer-aided detection (CADe) systems have focused on using 3D data from volumetric CT images to detect lymph nodes (LNs). Barbu et al. utilized feature extraction and integration based on boosting to create a reliable binary classifier for specific features. Their method was tested on 131 volumes with 371 LNs to identify axillary LNs, and the results demonstrated an 83.0% detection rate with only 1.0 False Positive (FP) per volume. However, this investigation had limited training specimens available, and 3D CT scans had more dimensions [14].

A technique based on convolutional networks was presented by Oda et al. to detect and segment mediastinal lymph nodes from chest CT volumes. The evaluation indicated that 95% of lymph nodes could be detected with 16.3 False Positives (FP) per CT volume. However, the model may have been overfitting due to its numerous parameters. To address the size discrepancy between classes, additional portions of the chest anatomy should be included in the dataset [12].

Bouget et al. developed a 3D convolutional neural network using either slab-wise or downsampled complete volume methods to segment tissue based on nearby anatomical characteristics, such as similar lymph node attenuation values. The performance of the approach was evaluated using a dataset of 120 contrast-enhanced CT volumes and a 5-fold cross-validation method. For the 1178 lymph nodes with a 10 mm diameter, the technique achieved a patient-wise recall of 92% and a segmentation overlap of 80.5%, with a false positive rate of 5.0 per patient. The best results were obtained by combining both slab-wise and full-volume techniques. However, the most notable findings required more than four organs, highlighting the need for extensive sample size [15].

To identify abdominal lymph nodes (ALNs) and mediastinal lymph nodes (MLNs) in CT images, Tekchandani et al. recommended attention U-Net-based deep learning architecture variants with complete and partial transfer learning. The attention U-Net was chosen because it could focus more on the target structures with trainable parameters. To address the small image dataset, full and partial transfer learning (TL) was employed. The attention U-Net was trained on MLN and ALN CT images and multiple tests were conducted to evaluate the effectiveness of the proposed approach. SegNet, U-Net, and ResUNet were compared with the suggested strategy. Although the model appeared overfitted when segmenting MLNs, the maximum achievable sensitivity and Dice scores for ALNs were 91.69% and 93.08%, respectively [16].

Seff et al. [17] proposed a solution for addressing 3D identification issues in CT scans by dividing the task into 2D detection subtasks. The authors divided a prospect into 27 CT slices to mitigate the curse of dimensionality problems. They used Histogram of Oriented Gradients (HOG) to extract features and performed linear classification. For the mediastinal and abdominal datasets, the sensitivity was 78.0% at 6 False Positives (FP) per volume or 86.1% at 10 FP per volume, and 73.1% at 6 FP per volume or 87.2% at 10 FP per volume, respectively. In addition, they utilized baseline HOG methods and a state-of-the-art deep CNN [18] to identify mediastinal LNs, achieving sensitivities of 78% vs. 70% at 3 FP per scan and

88% vs. 84% at 6 FP per scan, respectively. Nonetheless, a more practical approach was needed to improve accuracy and sensitivity.

Roth et al. introduced a 2.5D detection technique to increase the number of samples. Their approach quickly decomposed each 3D volume of interest (VOI) by resampling 2D reformatted orthogonal views, which were then used to train the CNN classifier. The authors' experiment yielded a sensitivity of 70% / 83% at 3 False Positives (FP) per volume and a sensitivity of 84% / 90% at 6 FP per volume [19].

Zhu et al. proposed a multi-branch detection by segmentation network for detecting and segmenting cancerous lymph nodes. Their framework included an efficient distance-based gating method that mimicked the protocols used by oncologists in daily practice. However, since a PET/CT scan is usually obtained later in the diagnostic process, optimal processing of the initial CT volume is necessary before considering the PET/CT modality [20].

Liu et al. tackled the mediastinal lymph node detection and station mapping issue using a two-stage pipeline. They also segmented eight anatomical structures to aid in the station mapping process. Each lymph node candidate was allocated a station based on its centroid location in relation to surrounding structures, following IASLC guidelines. The final segmentation and instance detection results took over half an hour to produce and required five stages. However, the use of hand-crafted features and multiple steps restrict the method's generalization and practical applicability. Fully end-to-end approaches may be better suited to address both limitations simultaneously [21] [22].

Xu et al. put forward a method to perform pixel-wise semantic segmentation using the DeepLabv3+ architecture, which utilizes atrous spatial pyramid pooling at various grid scales to improve boundary segmentation. They studied the use of focal loss, which helps the network focus on voxels that are difficult to segment while preventing overfitting on non-challenging voxels of the other category. This approach is particularly effective for simple pixel-wise semantic segmentation tasks [23].

Lai et al. [24] utilized deep Convolutional Neural Networks (CNNs) to identify and amplify lymph nodes (LNs), leading to a significant improvement in the efficiency of training set amplification and feature extraction from three-dimensional (3D) images. They also incorporated a novel method of using three continuous parallel images, in addition to the standard approach of decomposing 3D images into three orthogonal two-dimensional (2D) images. Furthermore, they devised a vote algorithm to integrate various possibilities and their respective values. During the initial experiment, they trained on 90% of candidate slices from each patient and tested on the remaining samples, achieving a sensitivity of up to 99%. However, to account for identical samples, they trained on 80% of patient slices and tested on the remaining 20%, resulting in a sensitivity of 98%. Nonetheless, the model appeared to be overfitted since it produced a linear line in the ROC curve.

In their study, Paing et al. [25] suggested a pipeline that combines traditional methods of extracting image features, such as thresholding, watershed, and Hessian eigenvalues, with a 3D neural network for classifying lymph node candidates and false lesions. Meanwhile, Nogues et al. focused on segmenting

lymph node clusters in mediastinal and abdominal contexts [11]. They addressed the challenge of reduced intensity and texture contrast among neighbouring lymph nodes by proposing the use of 2D holistically nested neural networks for embedded edge detection. To refine the segmentation, structured optimization techniques such as conditional random fields and graph cuts were recommended.

Singh et al. [26] employed U-Net in the candidate generation stage to achieve a 100% sensitivity rate, albeit at the cost of high levels of false positives, for generating volumes of interest (VOI). They comprehensively analyzed different input representations to train Convolutional Neural Networks (CNN) and 3D CNN classifiers and Support Vector Machines (SVMs) trained on features extracted from the aforementioned models to reduce false positives and detect lymph nodes. They evaluated their approaches on a dataset of 90 CT volumes containing 388 mediastinal lymph nodes. Their best approach achieved 84% sensitivity at 2.88 false positives per volume in the mediastinum of chest CT volumes.

Ono et al. [27] proposed a novel approach to detect abdominal lymph nodes in contrast-enhanced CT images by utilizing R2U-Net to suppress over-detection, detect and integrate lymph nodes. The authors evaluated three methods, each with different parameters. In method 1, the entire contrast-enhanced CT image (512×512) was used for learning and detection by R2U-Net. In method 2, the entire image was strided by 64 pixels (the same size as the patch image) with a kernel of 64×64 , and detection was performed. Finally, method 3's kernel was set to a stride of 32 pixels each (half the size of the patch image) of method 2, and detection was performed. During the patch-based detection of the lymph node region, the image was cropped out with a stride of 1/2 pixel for the size of the image. The proposed method 3 achieved a recall of 0.76, outperforming the other methods.

Numerous investigations on LN segmentation have shown several issues and resulted in suboptimal detection accuracy. These problems emphasize the need for an efficient approach to improve the sensitivity and accuracy of LN segmentation. Our goal was to develop a method that surpasses other methods in performance.

CHAPTER 3: RESEARCH METHOD

1. DATASET

CT images from TCIA [28], 5-patients [29], and ELCAP [30] public datasets were acquired, prepared, and used for training the proposed network. Data from those three datasets were combined in a ratio, and test datasets were built. Figure 1 displays some sample images and their masks.

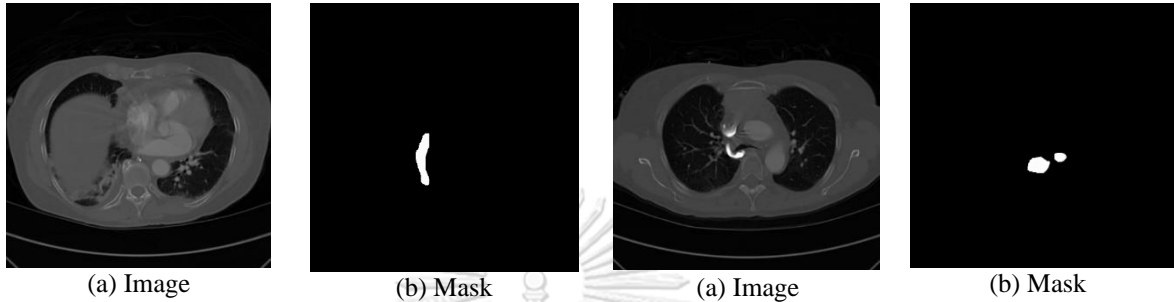


Figure 1. Dataset Sample a) Image, b) Mask

Three radiologists with different experience levels (21, 20, and 30 years, respectively) suggested the mediastinal LN datasets for this thesis. The VGG Image Annotation Toolkit was used to mask all LNs by two experienced radiologists with 21 and 20 years of experience. If there was a discrepancy, a third senior radiologist (30 years of expertise) was called in to decide about the presence of LN. In total, 28830 LNs were masked in the training dataset, which was used to train the model. Three additional testing datasets, each with an average of 8500 images, were segmented for testing and assessing the model's performance.

A. DATA PROCESSING

The experiment used high-resolution CT images because they produced the most significant diffusion effects. The background tissue signal is effectively suppressed on high-resolution CT scans, allowing the high signal intensity LNs to be seen and recognized. Images were manually cropped in a 512x512 matrix. The training dataset included 28830 processed images, and the testing dataset included three additional datasets, each with 8500 images. The model was trained separately using the TCIA, 5-patients, and ELCAP datasets to investigate its performance on different dataset images. There were three combinations for testing. The first combination (COMBO 1) was created by gathering 80% of images from TCIA, 10% from the 5-patients, and 10% from the ELCAP public dataset. The second combination (COMBO 2) was built with 60% images from TCIA, 20% from the 5-patients, and 20% from the ELCAP dataset. The third combination (COMBO 3) was created by collecting 50% of images from TCIA, 25% from the 5-patients, and 25% from the ELCAP dataset. The training dataset was divided into training and validation using the Python data split library.

B. DATA AUGMENTATION

Artificial data augmentation is a standard method used by CNN to provide enough training data. When the dataset is insufficient, augmentation teaches the network the needed invariances and

resilience characteristics [31]. The Python data augmentation program improved the training dataset in this thesis. The following operations were carried out on the images: image cropping, affine adjustments, vertical and horizontal flipping, noise and blur reduction, and contrast and brightness management. This study's testing data was left un-augmented, whereas the training dataset was augmented.

2. UNET MODEL ARCHITECTURE & UPSAMPLING TECHNIQUE

The UNet model is a convolutional neural network architecture that was originally proposed for medical image segmentation tasks in 2015 by Olaf Ronneberger, Philipp Fischer, and Thomas Brox [9]. It is named after its U-shaped architecture, which consists of a contracting path (left side) and an expanding path (right side) with a bottleneck in between. The architecture is shown in Figure 2.

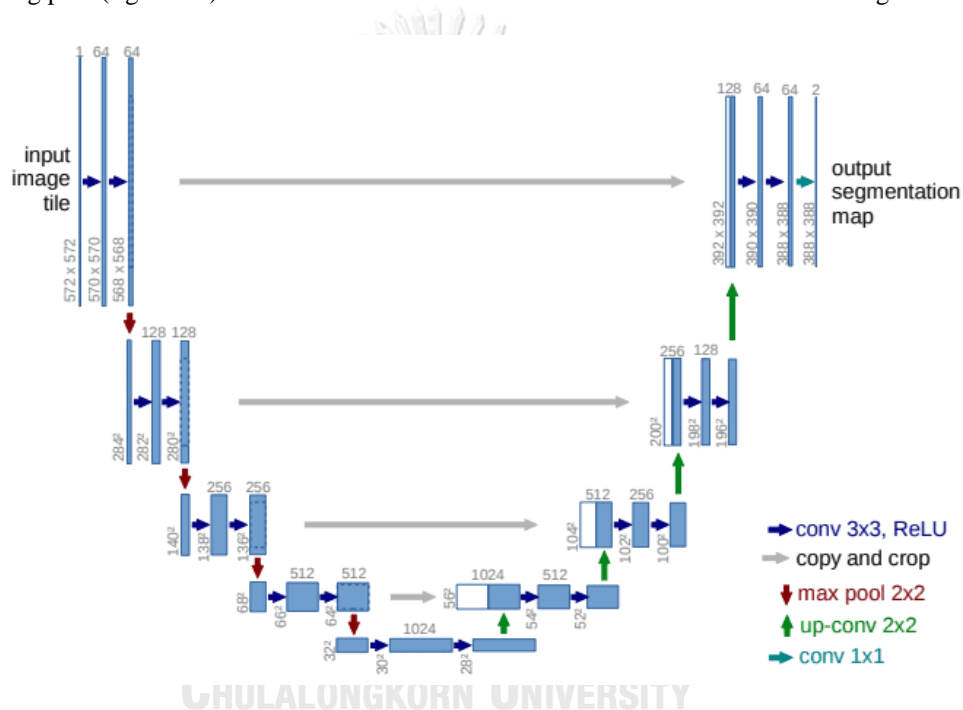


Figure 2. UNet Model Architecture

The UNet architecture is designed to handle the challenges of medical image segmentation, such as dealing with small training datasets and capturing fine-grained details in images. The contracting path is on the left side of the U and consists of several convolutional and pooling layers. The convolutional layers extract features from the input image and increase the number of channels (feature maps). The pooling layers reduce the feature maps' spatial size while keeping the essential features. The contracting path is designed to learn high-level features that capture the global structure of the image. As the spatial size of the feature maps decreases, the network can focus on the most relevant features while reducing the computational cost. The bottleneck of the U shape is where the contracting path ends and the expanding path begins. The bottleneck layer has a minimal spatial size, which limits the receptive field and encourages the network to learn more local features.

The expanding path is on the right side of the U and consists of several convolutional and upsampling layers. The convolutional layers recover the spatial resolution of the feature maps and reduce the number of channels. The upsampling layers increase the spatial resolution by a factor of two (or more) and allow the network to produce a segmentation map that has the same size as the original image. The expanding path is designed to refine the segmentation mask by adding fine-grained details. The skip connections that connect the corresponding layers between the contracting and expanding paths allow the network to combine both coarse and fine-grained information. The skip connections concatenate feature maps from the contracting path with the corresponding feature maps in the expanding path. This allows the network to learn to localize objects and retain fine-grained information.

The final layer of the UNet network is a pixel-wise classification layer that produces a binary segmentation map of the input image. This map assigns a class label to each pixel, indicating whether it belongs to the object of interest or not.

One of the critical features of the UNet architecture is its use of transpose convolutions, also known as deconvolutions or upsampling convolutions [32]. Transpose convolution is an operation that can upsample an input image, meaning that it increases its spatial resolution while keeping the number of channels constant. The transpose convolution operation can be considered the opposite of a regular convolution operation. While a regular convolution operation takes a small filter (kernel) and applies it to each pixel of the input image to produce an output feature map, a transpose convolution operation takes a larger filter and applies it to each pixel of the output feature map to produce an upsampled input image (figure 3).

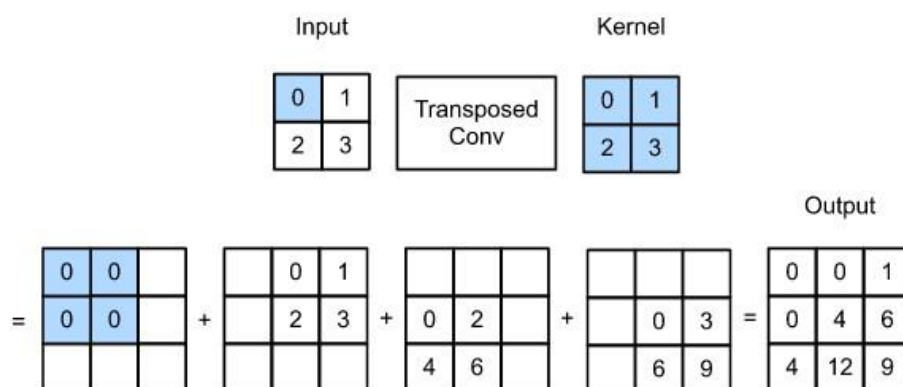


Figure 3. An Example of Transpose Convolution

In the UNet architecture, transpose convolutions are used in the decoder part of the network to increase the spatial resolution of the feature maps. The transpose convolutional layers in the UNet architecture are typically implemented using the so-called "convolution transpose" operation, which is also sometimes called "fractional stride convolution." This operation works by taking an input feature map with dimensions (H, W, C), where H and W are the height and width of the feature map, and C is the number of feature maps, and applying a learned kernel of size (K, K, C', C), where K is the kernel

size, C' is the number of output feature maps, and C is the number of input feature maps. The convolution transpose operation is similar to a regular one but with the spatial dimensions reversed. Specifically, instead of sliding the kernel over the input feature map, as in a regular convolution, the kernel is slid over the output feature map, and the values in the kernel are multiplied by the corresponding values in the input feature map.

3. PROPOSED MODEL ARCHITECTURE

A modified UNet++ network was used to identify potential lymph node candidates from mediastinal CT data. The UNet++ is a convolutional network based on U-Net for segmenting biological images quickly and precisely. The network can be divided into a contracting-expansive path or an encoder-decoder path. Two 3×3 convolutions were used repeatedly to create the encoder. After each convolution, there was a batch normalizing and ReLU. Each layer could learn more independently of the others because of batch normalization's reduced covariate shift. 2×2 max-pooling was used to minimize the spatial dimensions. Before max-pooling, a dropout layer acts as a regularizer and prevents overfitting. At each step of the downsampling process, the spatial dimensions were reduced by half while the number of feature channels was doubled. Every step in the expanding route started with an upsampling of the feature map. Total generalized variation (TGV) and bilinear interpolation were used to restore image resolution, preserve texture discontinuities, pick out noisy regions, and find the right balance points using backpropagation. Before concatenation, nested and dense skip connections [33] were used to fill the semantic gap between the encoder and decoder's feature maps. The modified UNet++ model architecture is shown in Figure 4, where thick convolution blocks on the skip paths are green and blue. We combined a 3×3 convolutional feature map from the contracting path with an equal feature map. Each 16-component feature vector was mapped using a 1×1 convolution, guaranteeing accurate image segmentation at the top layer.

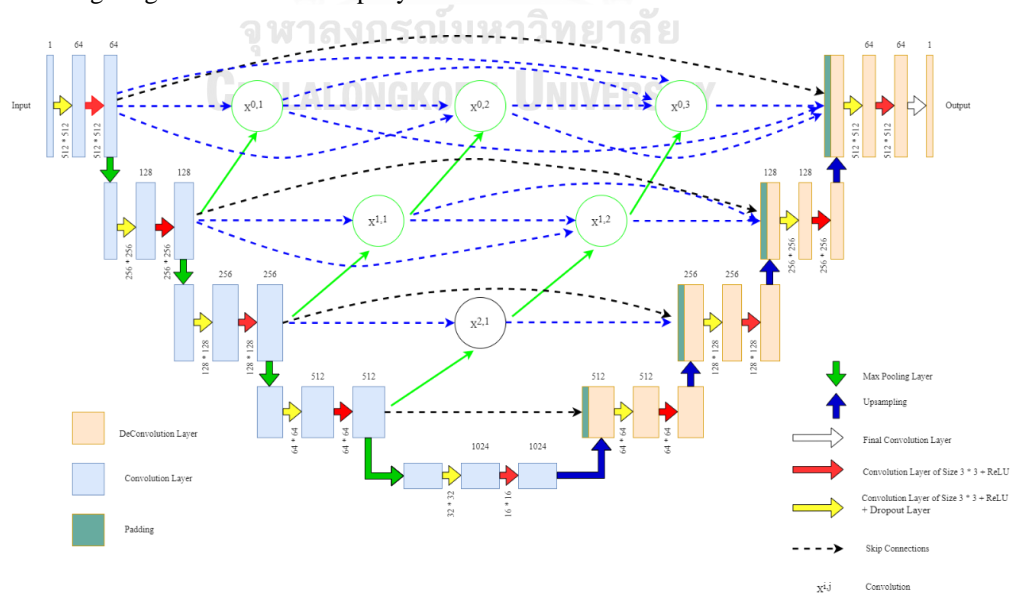


Figure 4. Modified UNet++ Model Architecture

4. PROPOSED UPSAMPLING TECHNIQUE

The feature map is an essential component of the UNet++ design since it generates the spatial information of the feature map by upsampling. Transpose convolution is used by default in the upsampling section of the network. Nevertheless, this execution strategy is quite time-consuming because of the large number of learnable parameters and the requirement to train new weights by kernels. Additionally, as demonstrated in Figure 5, it may lead to uneven overlap and artifacts at different scales.

An alternative to the traditional deconvolution, bilinear interpolation was employed in the decoder section of the model to avoid artifacts. This approach has no artifacts because of its normal behavior. The image's $g(x_1, x_2)$ nearest neighbors are the image coordinates $f(x_{10}, x_{20})$, $f(x_{11}, x_{21})$, $f(x_{12}, x_{22})$ and $f(x_{13}, x_{23})$. The interpolated image's $g(x_1, x_2)$ intensity values are evaluated as follows:

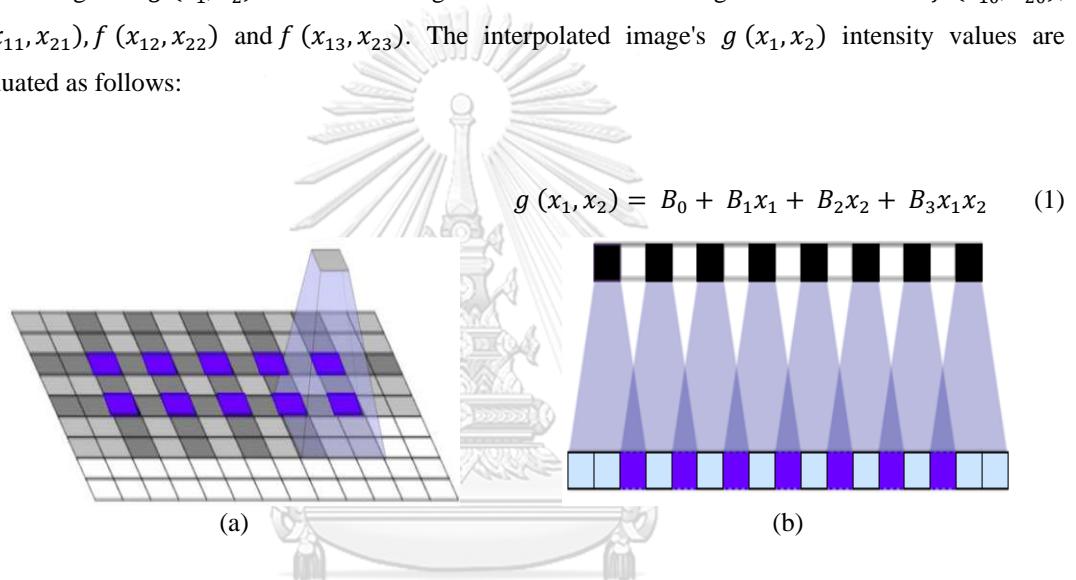


Figure 5. Artifact issue created by the transposed convolution: (a) checkerboard problem, (b) uneven overlap

Equation (1) denotes a bilinear function that has coordinates (x_1, x_2) . Here, g refers to an intensity value of interpolated image matrix in (x_1, x_2) coordinates, B_s are the bilinear weights, and f refers intensity value given at (x_{10}, x_{20}) , (x_{11}, x_{21}) , (x_{12}, x_{22}) and (x_{13}, x_{23}) pixel locations before performing interpolation. The bilinear weights, B_0 , B_1 , B_2 and B_3 are calculated by evaluating the matrix mentioned in Equation (2):

$$\begin{bmatrix} B_0 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} 1 & x_{10} & x_{20} & x_{10}x_{20} \\ 1 & x_{11} & x_{21} & x_{11}x_{21} \\ 1 & x_{12} & x_{22} & x_{12}x_{22} \\ 1 & x_{13} & x_{23} & x_{13}x_{23} \end{bmatrix} \quad (2)$$

As an outcome, $g(x_1, x_2)$ is determined as a linear combination of its four closest neighbor's gray levels. According to Equation (1), when the perfect least-squares planar fit is made to those four neighbors, the value is assigned to $g(x_1, x_2)$. This optimal averaging methodology provides smoother

outcomes considering the low-and high-resolution map's size difference. The bilinear interpolation technique is depicted in Figure 6.

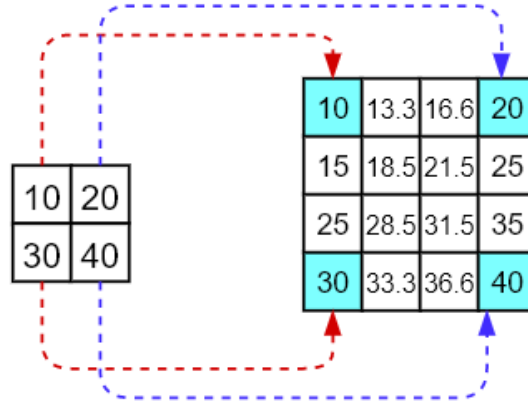


Figure 6. Example of bilinear interpolation

The result generated by bilinear interpolation contains the equal size as processed maps in the encoding stage. These maps refer to the entered image and are seen as noisy areas. The TGV treats the processed maps for denoising resolution. The mathematical term used for the model is mentioned in Equation (3):

$$\min_u \{ \gamma \text{loss}(u, x) + TGV_\alpha^K(v) \} \quad (3)$$

Here, image fidelity is represented by $\text{loss}(u, x)$, γ is utilized for global optimization, and the term $TGV_\alpha^K(v)$ is the regularization term.

Convolutional computations may effectively extract and generalize features, and their outputs can be visualized as a set of features. To better comprehend the semantics of an image, the smoothness of textures and border edges are essentially used for the feature maps. The mentioned technique maintains texture discontinuities by incorporating different channel information.

For our work, a second-order TGV-based technique was enough. That is why the utilized Equation (4) was:

$$TGV_\alpha^2(v) = \min_w \left\{ p_1 \int_\Omega |\nabla v - w| + p_2 \int_\Omega |\varepsilon(w)| \right\} \quad (4)$$

Here, v indicates the minimum TGV over complex vector fields w in the bounded domain. The symmetric derivative is represented by $\varepsilon(w)$. p_1 and p_2 are responsible for balancing the first and second derivatives.

Combining loss function and Equation (3), the term TGV can be defined as

$$\min_w \left\{ \frac{\lambda}{2} \left\| u(C_{input}) - \overline{C_{input}} \right\|_2^2 + p_1 \int_{\Omega} |\nabla v - w| + p_2 \int_{\Omega} |\varepsilon(w)| \right\} \quad (5)$$

Taking weights p_1 and p_2 into backpropagation, an appropriate balance point was decided by the network.

5. BACKPROPAGATION

Backpropagation is a supervised algorithm used for adjusting the weights p_1 , p_2 and the network's deviation. As a result, the output and the targeted vectors can be closer. It compares the error sum, and the training is finished if the network's output layer's error is evaluated as lower than the specified error. The network's deviation and weight are stored. The process is performed following the steps:

1. Setting nonzero coefficient and initial weight p_{ik} but $p_{i,n+1} = -threshold$
2. Entering sample $A = (A_1, A_2, \dots)$ and getting the desired output $B = (B_1, B_2, \dots)$.
3. Calculating output of all layers. For output A_i^j there is:

$$X_i^j = \sum_{k=1}^{m+1} p_{ik} A_k^{j-1} \quad (6)$$

$$A_i^j = f(X_i^j) \quad (7)$$

4. Finding learning errors for each layer. For the output layer, $j=n$:

$$e_i^n = A_i^n (1 - A_i^n) (A_i^n - B_i) \quad (8)$$

5. Modify weight and threshold.

$$\Delta p_{ik}(h+1) = \Delta p_{ik}(h) - \eta \cdot e_i^n \cdot A_k^{j-1} \quad (9)$$

6. Ensure the quality of weights according to the requirement. If the requirement is not filled up, return to step 3. The process will be repeated until the requirement is satisfied.

6. EVALUATION PROCESS

The parameter's value was determined using the following method: True Positives (TP) are accurately predicted positive values that show both the projected and actual class values to be true. True Negatives (TN) are precisely anticipated negative values, indicating that the actual and projected class values are zero. When the actual and the projected classes are different, false positives and negatives occur. False Positives (FP) happen when the projected class is yes, but the actual class is no. False Negatives (FN) occur when the projected class is negative but the actual class is positive. The model's performance was measured using precision, recall, F1, and accuracy. A higher value indicates

better performance. The ratio of accurately predicted observations to all observations is known as accuracy. Equation (10) was used to calculate the model's accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (10)$$

The ratio of accurately predicted positive observations and total expected positive observations is precision. The precision of the model was calculated following Equation (11).

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

The ratio of accurately predicted positive observations and all actual yes class observations are known as recall. Precision and recall are weighted into the F1. As a result, false positives and negatives are considered while calculating this score. When there is an unequal distribution of classes, F1 is usually more helpful. Recall and F1 are calculated following Equations (12) and (13) accordingly.

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \quad (13)$$

7. LYMPH NODE SEGMENTATION CRITERIA

The Dice similarity coefficient (DSC) measures the similarity between the segmentation results and the ground truth. The DSC is a number from 0 to 1, with a more excellent value indicating better segmentation accuracy. The DSC was established as follows:

$$DSC(R, T) = \frac{2P(R \cap T)}{P(R) + P(T)} \quad (14)$$

Here R is the segmentation result of the segmentation method, T is the ground truth, and P is the number of pixels in the corresponding set.

CHAPTER 4: RESULT AND ANALYSIS

1. EXPERIMENTAL ENVIRONMENT SETUP

An Ubuntu 18.04 workstation with 32 Intel CPUs, a Tesla 16 Gb GPU, and 64 GB of RAM was used to train the UNet++ model. The layers were trained using an Adam Optimizer with the default value ($b1 = 0.9$, $b2 = 0.999$) and 32 images per batch. Table 1 illustrates the training hyperparameters. 10% of the 28830 images in the training dataset were utilized to evaluate the deep-learning model's learning effect (as a validation set), while the rest were employed to train the model.

Table 1. Model Parameters

Parameter Name	Value
Epochs	100
Learning Rate	1. e - 4
Batch Size	32
Optimizer	Adam
Input Image	512 * 512

The study used a 10-fold cross-validation approach to validate CT data. The entire dataset was split into two subgroups for a cross-validation run: the training and the validation sets. Several cross-validations with different random partitions were frequently executed. The model performance was calculated using the average validation data acquired from numerous cross-validation rounds. The model's loss changes were examined on the validation set during the training phase to consider alternate options for the learning rate and other parameters. If the loss of the validation set did not minimize for ten epochs, the learning rate would be cut in half from its current level. If the loss of the validation set does not decrease for 20 epochs, the training procedure must be terminated early to prevent over-fitting. The final product of the network was a probability map for the foreground and background. The train set received primary data augmentation, whereas the test set remained unaltered. The final prediction results were assessed by averaging the model's outputs over the enriched data.

2. PERFORMANCE MEASUREMENT ON TCIA DATASET

A. TRAINING ACCURACY MEASUREMENT

The updated UNet++ model was trained on 34 mediastinal LN slices from the TCIA dataset, each with an average of 600 pictures, after being reviewed by three radiologists with varying levels of experience (21, 20, and 30 years) in mediastinal LNs imaging diagnosis. 10% of the training set's images were used to fine-tune the model's learning capabilities and trained for 100 epochs.

Figure 7(a) displays the model's training accuracy and loss on the TCIA dataset, while 7(b), 7(c), 7(d), and 7(e) show the validation accuracy and loss, intersect of union (IOU) score, dice coefficient score, and recall, respectively. The model achieved a validation accuracy of 93.4% and a training accuracy of 97.1%. Although there was a significant loss at the start of the training process, it was quickly reduced as the number of epochs increased, indicating that the model could learn quickly. The IOU, dice coefficient, and recall scores were 90.8%, 92.5%, and 94.4%, respectively.

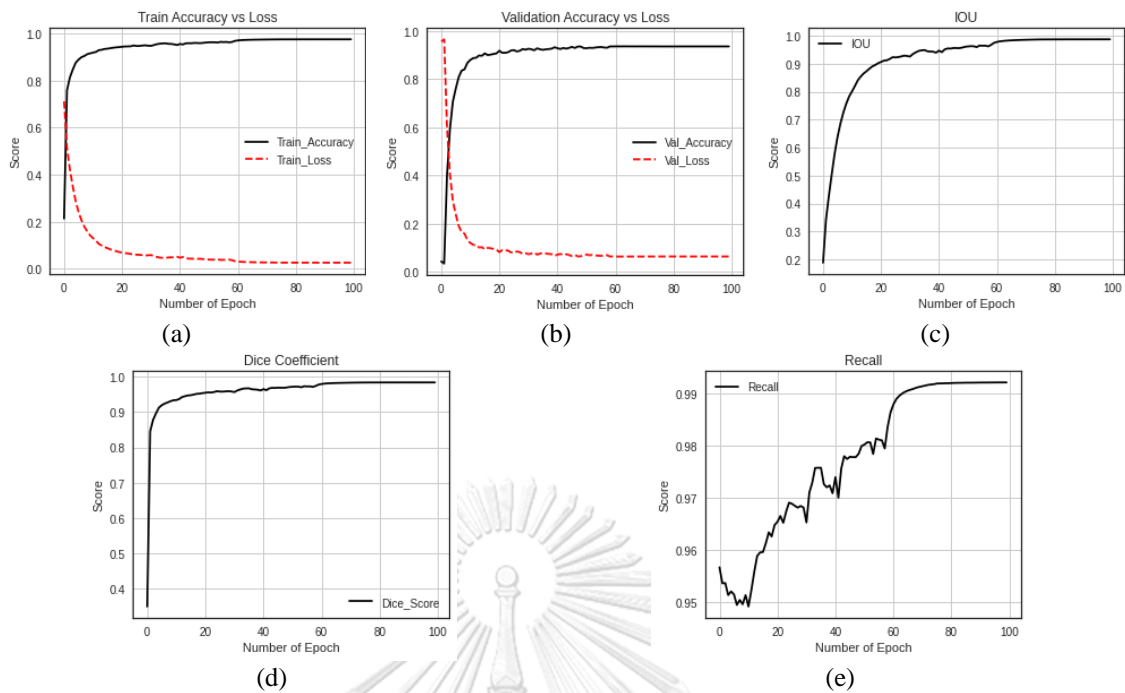


Figure 7. Training Score Evaluation a) Train Accuracy vs. Loss, b) Validation Accuracy vs. Loss, c) Intersect of Union (IOU), d) Dice Coefficient, e) Recall

B. TESTING ACCURACY MEASUREMENT

To evaluate the accuracy of our model on diverse datasets, we generated three distinct testing datasets by combining CT images from the TCIA, ELCAP, and 5-patient datasets. Each testing dataset (COMBO 1, COMBO 2, and COMBO 3) contained 8500 mediastinal lymph node CT scans and their corresponding masks. Expert radiologists with years of experience in this field created and validated the testing datasets.

I. EXPERIMENTAL ANALYSIS ON COMBO_1

We began by evaluating the model's performance on COMBO 1, a combination of CT images from TCIA (80%), the 5-patient dataset (10%), and the ELCAP public dataset (10%). We assessed the model's performance using five criteria - accuracy, Jaccard, precision, recall, and F1.

The performance of the model on the COMBO 1 dataset is presented in Figure 8, where the accuracy, Jaccard score, precision, recall, and F1 parameter values are shown in 8(a), 8(b), 8(c), 8(d), and 8(e), respectively. The model achieved remarkable results, with 91.1% accuracy, 89.0% Jaccard score, 87.4% precision, 89.7% recall, and 88.5% F1 score.

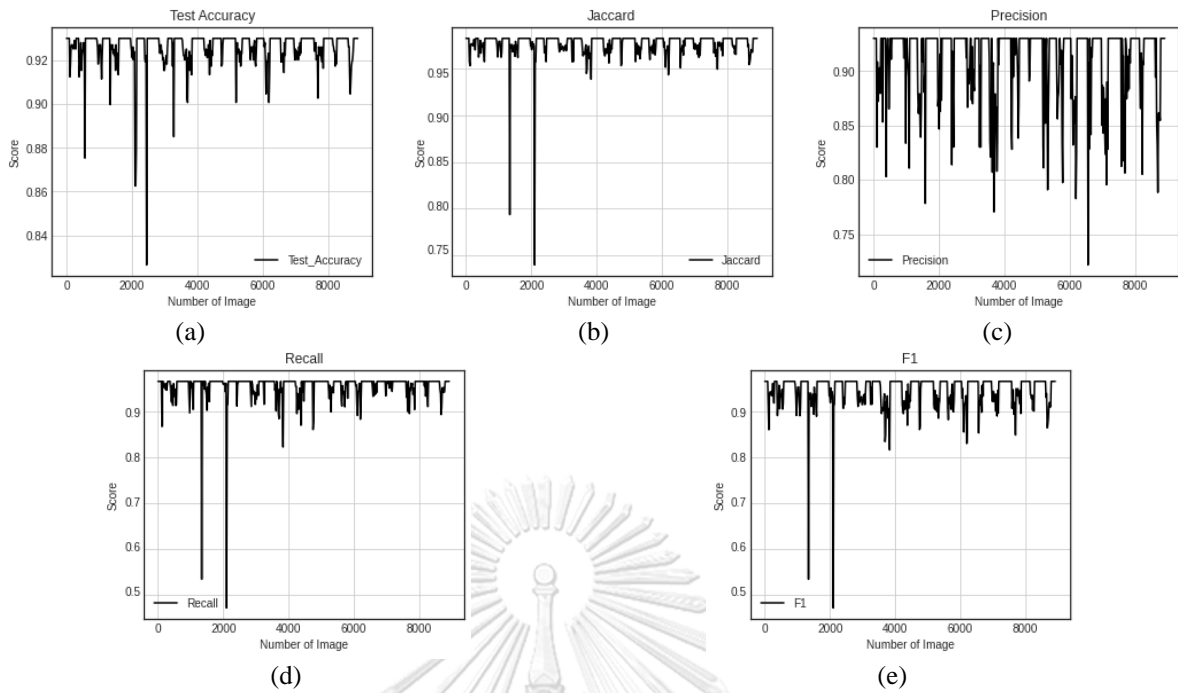


Figure 8. Testing Accuracy Measurement on Test Dataset COMBO_1, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

II. EXPERIMENTAL ANALYSIS ON COMBO_2

The COMBO_2 dataset was created by combining 60% of images from TCIA, 20% from the 5-patients, and 20% from the ELCAP public dataset. The model's performance on this dataset is presented in Figure 9, where (a) shows the accuracy parameter, (b) displays the Jaccard score, (c) illustrates precision, (d) exhibits recall, and (e) presents the F1 parameter value. The model achieved an accuracy of 92.7%, a Jaccard score of 90.5%, a precision of 89.3%, a recall of 91.4%, and an F1 of 90.3%.

III. EXPERIMENTAL ANALYSIS ON COMBO_3

The COMBO_3 dataset was created by combining 50% of the images from TCIA, 25% from the 5-patients dataset, and 25% from the ELCAP public dataset. Figure 10 presents the model's performance on the COMBO_3 dataset, with (a) displaying the accuracy parameter, (b) showing the Jaccard score, (c) depicting the precision parameter, (d) showing the recall parameter, and (e) displaying the F1 parameter. The model achieved an accuracy of 93.8%, a Jaccard score of 91.71%, a precision of 89.81%, a recall of 93.21%, and an F1 of 91.51%. The COMBO_3 dataset produced better results.

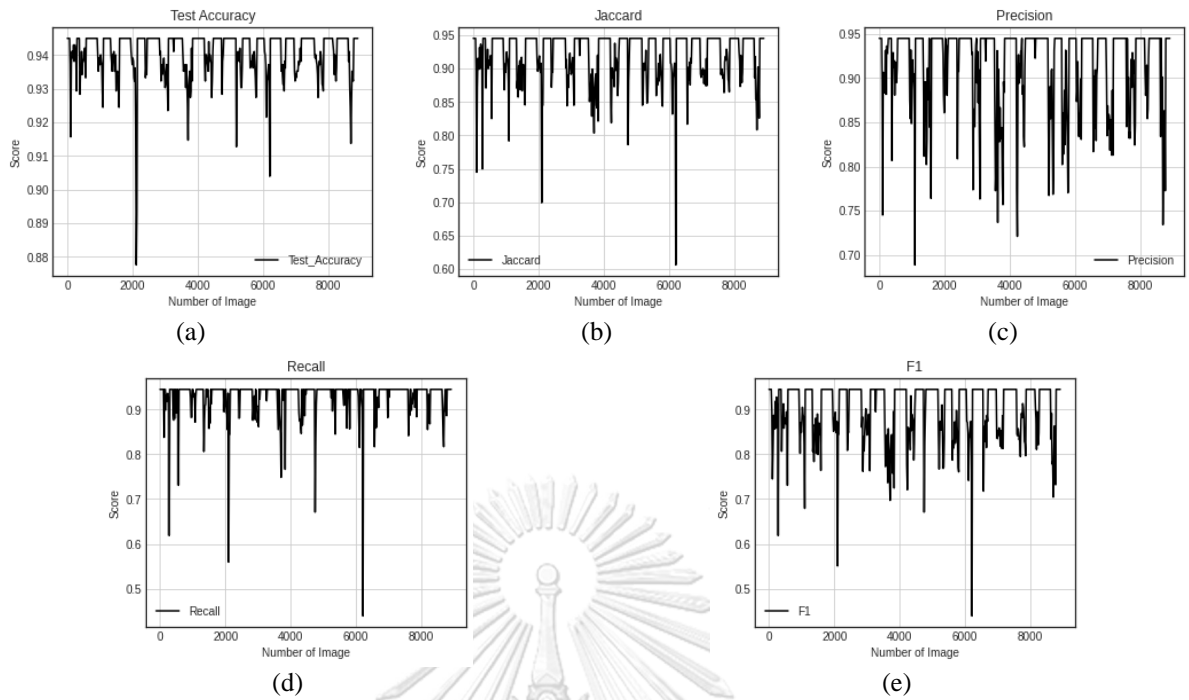


Figure 9. Testing Accuracy Measurement on Test Dataset COMBO_2, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

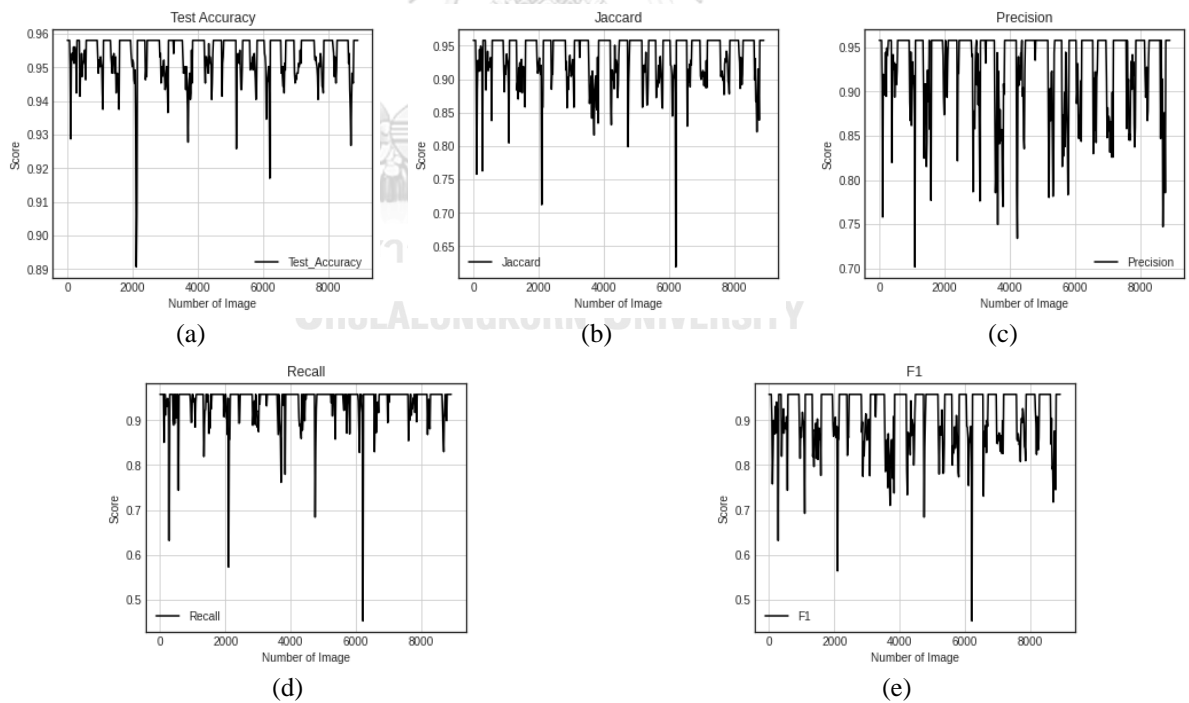


Figure 10. Testing Accuracy Measurement on Test Dataset COMBO_3, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

3. PERFORMANCE MEASUREMENT ON ELCAP DATASET

A. TRAINING ACCURACY MEASUREMENT

A modified UNet++ model was trained on 20 mediastinal LN slices from the ELCAP dataset, each containing an average of 250 images. To fine-tune the model's learning capabilities, 10% of the training set's images were used. The model was trained for 100 epochs, and Figure 11 presents the computed parameter values.

Figure 11(a) shows the accuracy and loss of the model during training on the ELCAP dataset, while Figure 11(b) displays the accuracy and loss during validation. The IOU score is shown in Figure 11(c), the dice coefficient score in Figure 11(d), and the recall score in Figure 11(e). With 95.04% validation accuracy and 97.1% training accuracy, the model initially showed a significant loss in both training and validation, which was quickly reduced, indicating faster learning. The IOU, dice coefficient, and recall scores were 89.8%, 91.58%, and 93.88%, respectively, demonstrating the model's excellent performance.

B. TESTING ACCURACY MEASUREMENT

I. EXPERIMENTAL ANALYSIS ON COMBO_1

In Figure 12, the accuracy parameter value is depicted in (a), the Jaccard score in (b), precision in (c), recall in (d), and the F1 parameter score in (e). The model achieved an accuracy of 92.1%, a Jaccard score of 89.6%, a precision of 88.2%, a recall of 91.8%, and an F1 score of 89.9%.

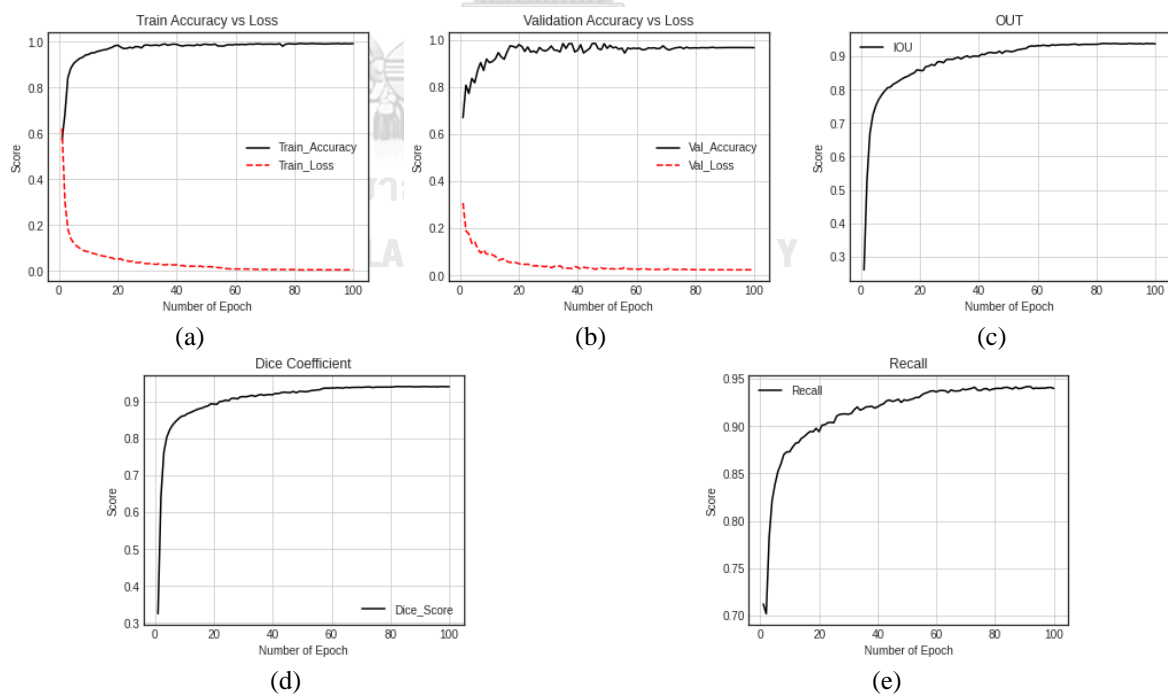


Figure 11. Training Score Evaluation a) Train Accuracy vs. Loss, b) Validation Accuracy vs. Loss, c) Intersect of Union (IOU), d) Dice Coefficient, e) Recall

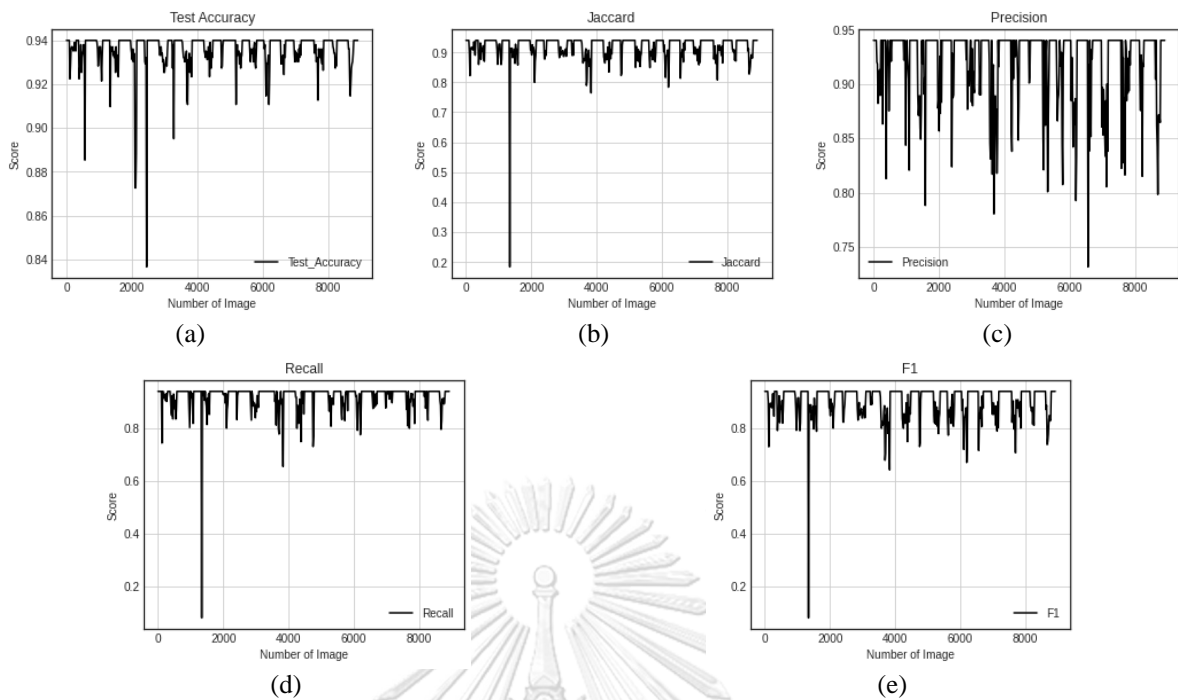


Figure 12. Testing Accuracy Measurement on Test Dataset COMBO_1, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

II. EXPERIMENTAL ANALYSIS ON COMBO_2

The model's performance on COMBO_2 was satisfactory, with an accuracy of 94.1%, Jaccard score of 93.6%, precision of 90.2%, recall of 93.8%, and F1 score of 91.9%, as demonstrated in Figure 13.

III. EXPERIMENTAL ANALYSIS ON COMBO_3

Figure 14 displays the performance metrics of the model, with (a) showing the accuracy, (b) the Jaccard score, (c) the precision, (d) the recall, and (e) the F1 score. The model achieved an accuracy of 93.6%, a Jaccard score of 91.7%, a precision of 90.1%, a recall of 92.9%, and an F1 score of 91.5%.

4. PERFORMANCE MEASUREMENT ON 5-PATIENTS DATASET

A. TRAINING ACCURACY MEASUREMENT

The modified UNet++ model was trained using five mediastinal LN slices from the 5-patient dataset, with an average of 436 images per slice. To improve the model's learning capabilities, 10% of the training set's images were used for fine-tuning. The model underwent 100 epochs, and Figure 15 displays the estimated parameter values.

Figure 15 presents a comprehensive evaluation of the model's performance on the 5-patient dataset. Specifically, 15(a) shows the training accuracy and loss, 15(b) illustrates the validation accuracy and loss, 15(c) depicts the intersection over union (IOU) score, 15(d) displays the dice coefficient score, and 15(e) indicates the recall score. The model performed remarkably well, maintaining its high training accuracy of 97.9% while achieving a validation accuracy of 94.1%.

Additionally, the IOU score, dice coefficient score, and recall score were also impressive, at 92.8%, 91.6%, and 95.7%, respectively.

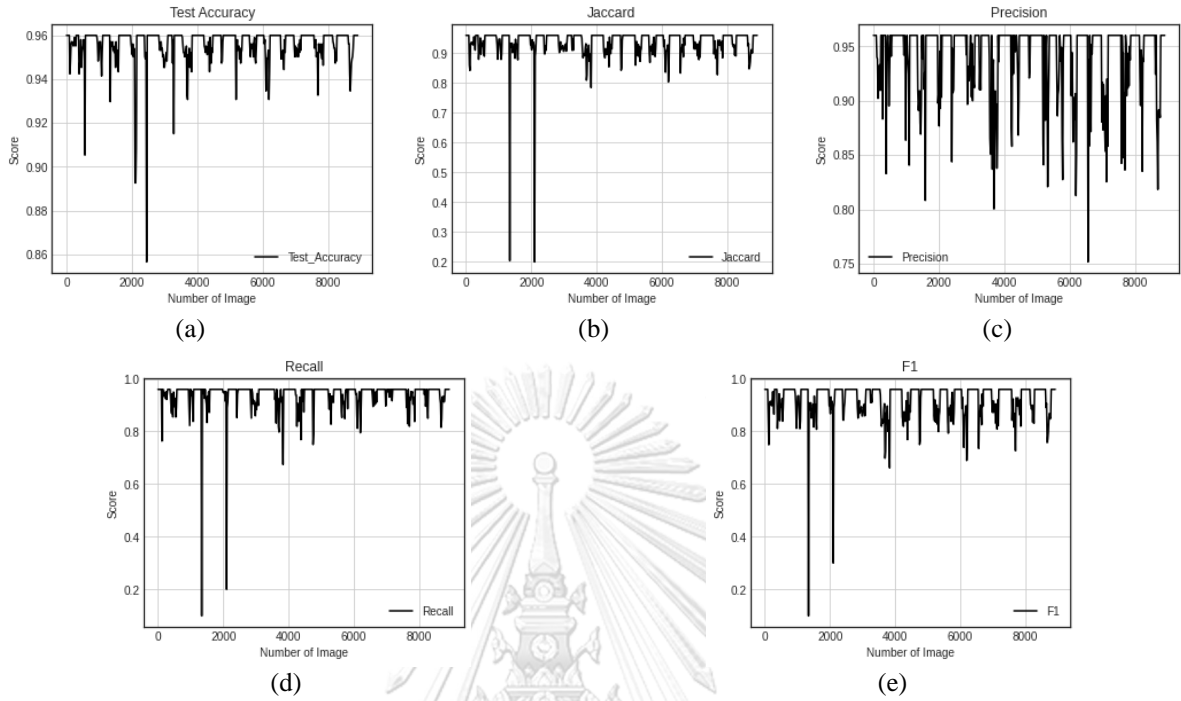


Figure 13. Testing Accuracy Measurement on Test Dataset COMBO_2, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

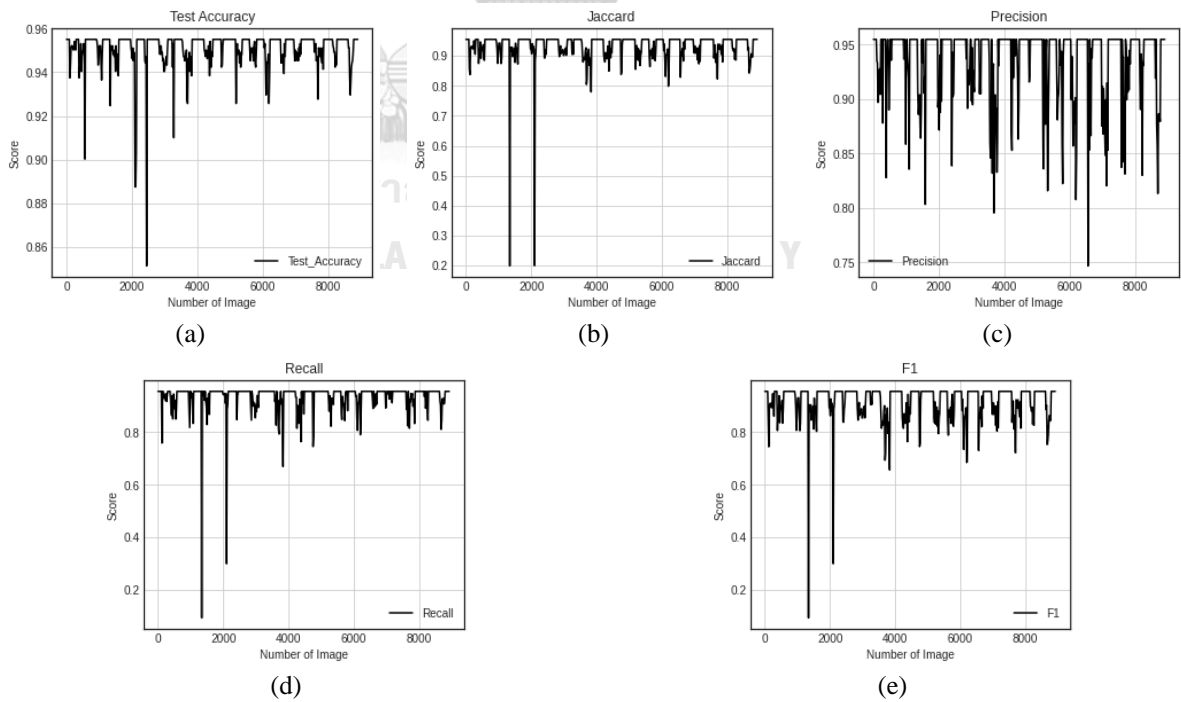


Figure 14. Testing Accuracy Measurement on Test Dataset COMBO_3, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

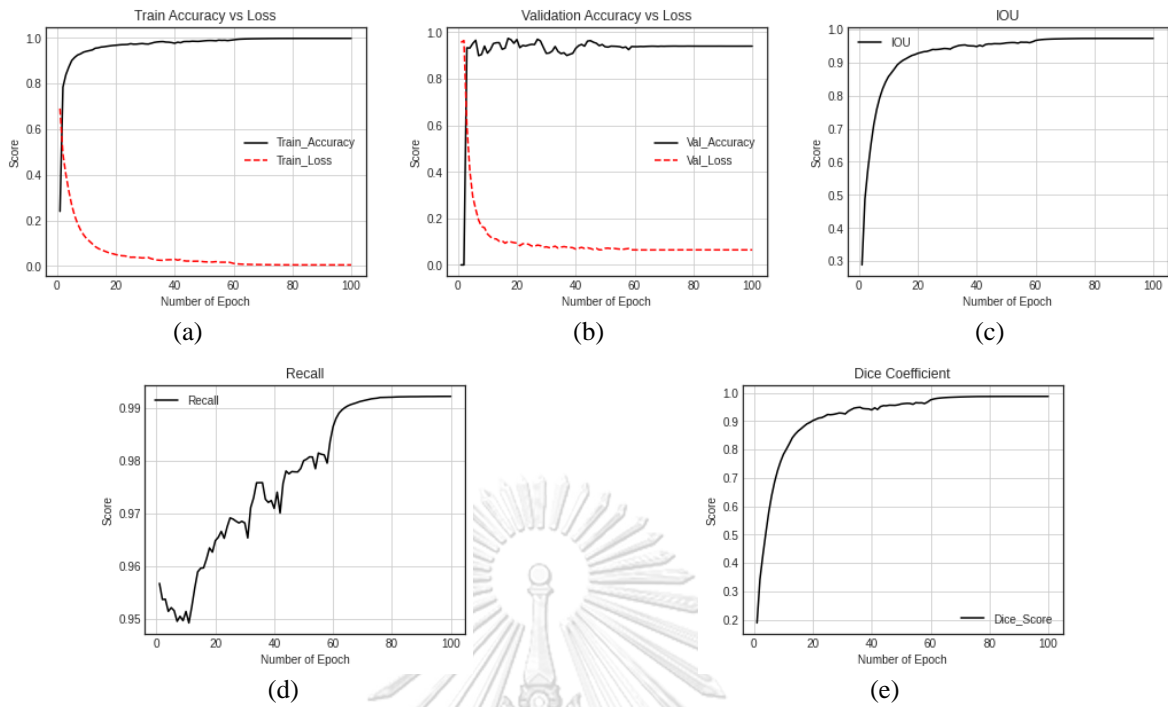


Figure 15. Training Score Evaluation a) Train Accuracy vs. Loss, b) Validation Accuracy vs. Loss, c) Intersect of Union (IOU), d) Dice Coefficient, e) Recall

B. TESTING ACCURACY MEASUREMENT

I. EXPERIMENTAL ANALYSIS ON COMBO_1

In Figure 16, the accuracy parameter is shown in (a), the Jaccard score in (b), precision in (c), recall in (d), and the F1 score in (e). The model achieved 93.2% accuracy, 89.4% Jaccard, 91.3% precision, 92.9% recall, and 92.0% F1.

II. EXPERIMENTAL ANALYSIS ON COMBO_2

In Figure 17, (a) shows the accuracy parameter, (b) shows the Jaccard score, (c) represents precision, (d) indicates recall, and (e) illustrates the F1 parameter's score. The model obtained 94.1% accuracy, 90.6% Jaccard, 91.6% precision, 93.4% recall, and 92.5% F1.

III. EXPERIMENTAL ANALYSIS ON COMBO_3

Figure 18 presents the evaluation metrics for the model, including accuracy (a), Jaccard score (b), precision (c), recall (d), and F1 score (e). The model performed exceptionally well on this dataset, achieving 94.8% accuracy, 91.9% Jaccard score, 93.1% precision, 94.1% recall, and 93.5% F1 score, making it the best-performing combination.

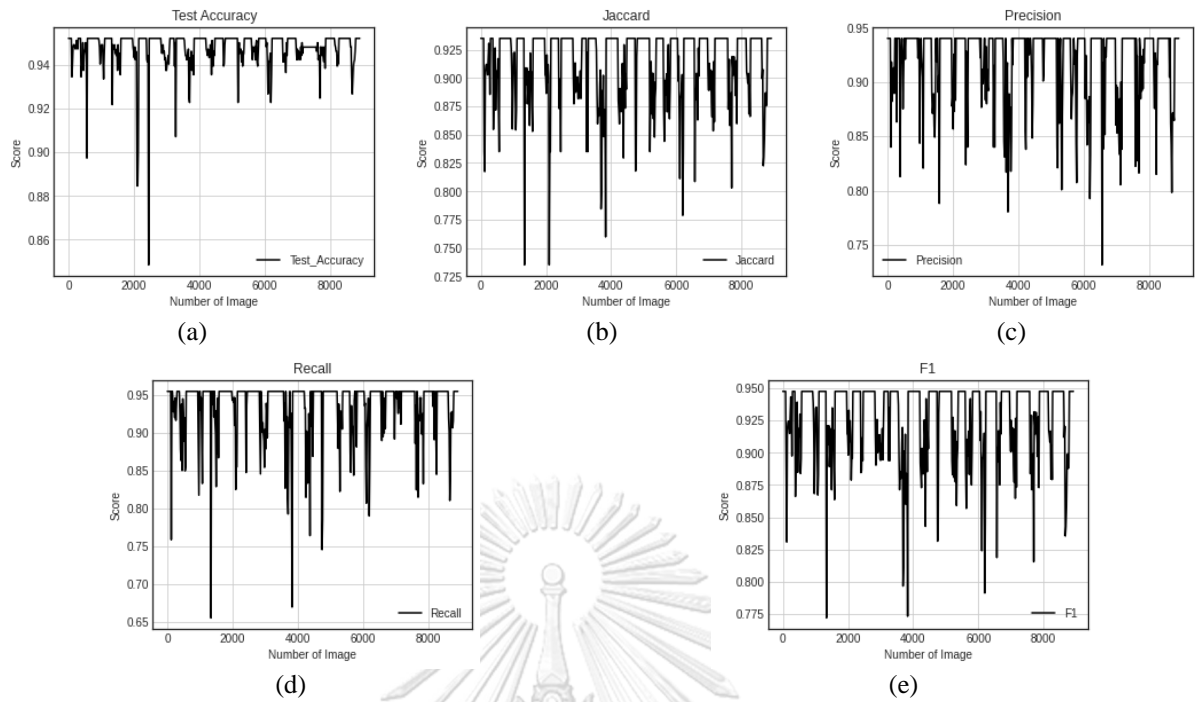


Figure 16. Testing Accuracy Measurement on Test Dataset COMBO_1, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

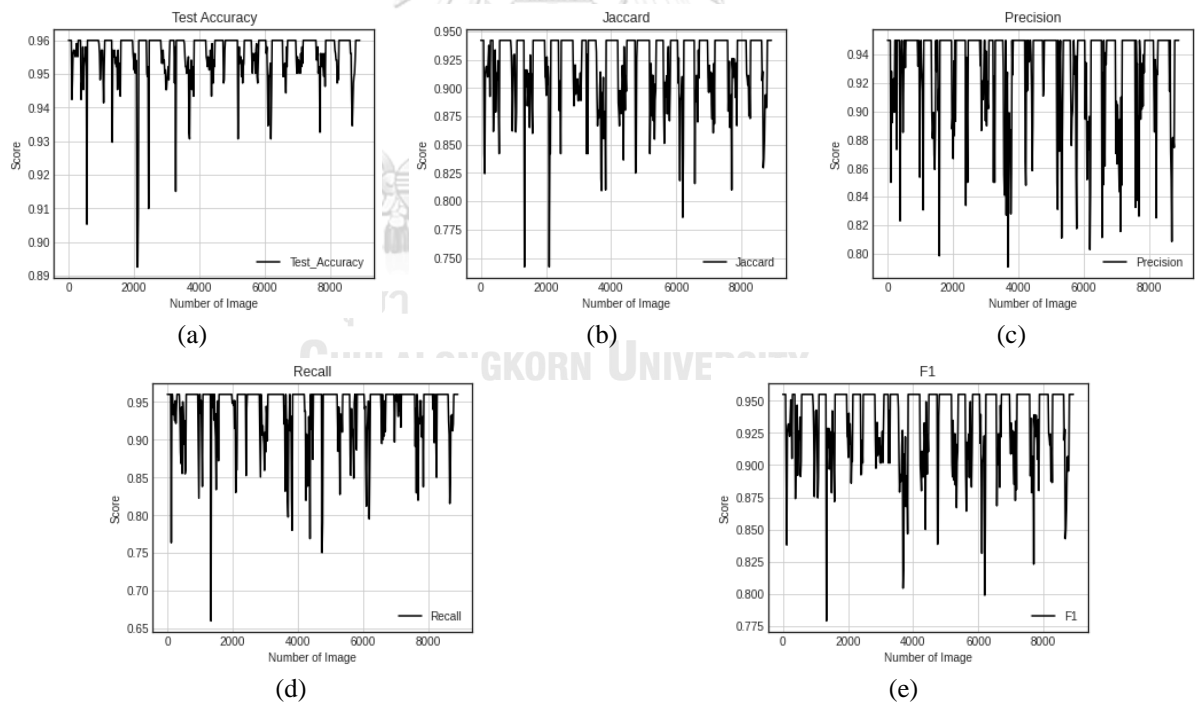


Figure 17. Testing Accuracy Measurement on Test Dataset COMBO_2, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

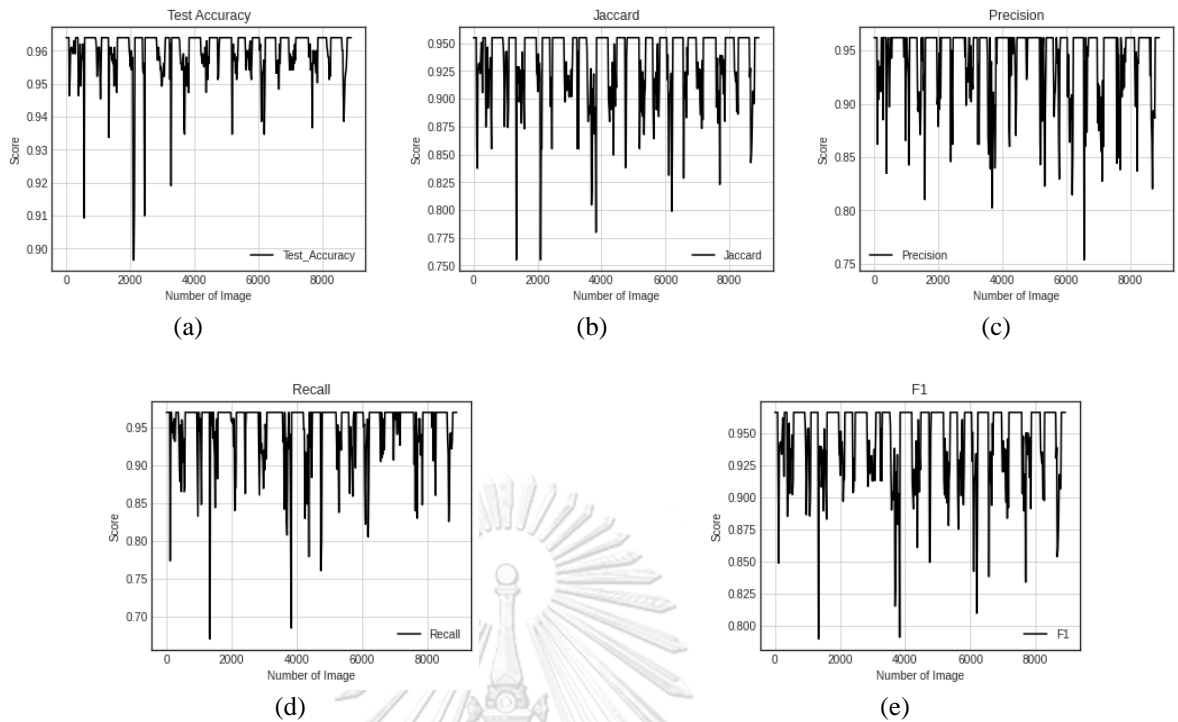


Figure 18. Testing Accuracy Measurement on Test Dataset COMBO_3, a) Accuracy, b) Jaccard, c) Precision, d) Recall, e) F1

5. MEASUREMENT SUMMARY

The proposed model performed better on the COMBO 3 test dataset and the 5-patient dataset. Those datasets' mean and standard deviation were examined to determine the actual cause. We found 0.1881, 2085, and 0.2272 standard deviation scores and 0.2207, 0.4070, and 0.2226 mean scores for TCIA, ELCAP, and 5-patient. The standard deviation measures observed values or data deviation from the mean. The measured values converge toward the mean when the standard deviation is near zero. The values or data are distributed far from the mean when the standard deviation is high. When the standard deviation is 0, we can state that all the data are equal. A higher standard value indicates an image with a sharp contrast, and a sharp contrast image emphasizes the comparison of various pixel colors. The model can learn different pixel values more accurately on high-contrast image datasets that directly impact the model's performance. The suggested model performs better because the 5-patient dataset has a more considerable standard deviation than previous datasets. The proposed architecture performed better on the other two datasets even though the 5-patient dataset gave it the best accuracy. The model has maintained 91+% accuracy and 90+% precision score on both TCIA and ELCAP, which demonstrates the model's capability in segmenting and detecting LNs from different dataset images.

6. PERFORMANCE COMPARISON

The model's performance was compared to other cutting-edge methods, including U-Net (unmodified), AlexNet, SegNet, and Auto-LNDS (Mask RCNN-based Auto Lymph Node Detection

and Segmentation, a proposed model by X. Zhao et al. [34]). The TCIA, ELCAP, and 5-patient datasets were used to train those models, and the COMBO 1, COMBO 2, and COMBO 3 test datasets were used to test them. The parameter values for the models were kept constant, as shown in Table 1.

A. PERFORMANCE MEASUREMENT USING OLD_U-NET

This study used new upsampling methods to modify the UNet++ model; the preceding section covers the changed approach. A second experimental analysis was completed using the default transpose convolution for upsampling to compare the performance of the modified UNet++ and the original U-Net. No dense layer and no TGV-based technique were used. The model's performance was estimated on three datasets and three test combinations, as in Table 2.

Table 2. Old_U-Net Model's Performance

Dataset	Combo	FP/Vol	Accuracy	Dice	F1	Recall	Precision
TCIA	1	3.7	82.4%	81.7%	74.7	74.6%	74.9%
	2	3.09	82.2%	75.6%	78.1	74.5%	82.1%
	3	2.7	84.9%	76.7%	72.6	66.6%	79.7%
ELCAP	1	3.0	86.1%	79.5%	81.7	79.8%	83.7%
	2	2.4	85.9%	80.7%	77.5	69.1%	88.3%
	3	2.9	86.7%	79.1%	78.4	72.5%	85.4%
5-Patients	1	2.5	87.02%	81.7%	75.3	75.5%	75.1%
	2	2.48	87.9%	77.5%	79.1	72.6%	86.9%
	3	2.36	89.01%	79.2%	84.3	80.8%	88.2%

According to Table 2, Old_U-Net acquired the highest accuracy on the 5-patients training dataset and COMBO_3 test dataset. Still, its highest precision was developed on the ELCAP training dataset and COMBO_2 test dataset. Compared with the AlexNet, SegNet, and Auto-LNDS model's performance, the Old_U-Net scored the highest precision, but its accuracy was lower than that of AlexNet and SegNet. The model's overall performance maintains its second position among the four models.

B. PERFORMANCE MEASUREMENT USING ALEXNET

AlexNet has eight layers, three are entirely connected, and five are convolutional layers. Two crucial components of the AlexNet are overlapping pooling and rectified linear units (ReLU) [35]. Compared to the standard tanh function, ReLU nonlinearity enables substantially faster training. The overlapping pooling produces nearby clusters of neurons without overlapping, lowering training error [36]. The TCIA, ELCAP, and 5-patient datasets were used independently to train the network, and the test dataset combinations were used to evaluate it. In Table 3, the performance is mentioned.

Table 3 shows AlexNet obtained the best accuracy and precision on the COMBO 3 test dataset and 5-patient training dataset. The model's maximum accuracy exceeded SegNet, Auto-LNDS, and Old U-Net, but its precision lagged behind Auto-LNDS and Old U-Net. Overall, the model outperformed SegNet, and among all models trained on the TCIA dataset, AlexNet performed with higher precision and accuracy.

Table 3. AlexNet Model's Performance

Dataset	Combo	FP/Vol	Accuracy	Dice	F1	Recall	Precision
TCIA	1	2.37	82.1%	72.6%	74.1%	73.5%	74.7%
	2	2.77	83.9%	71.3%	78.3%	78.7%	77.9%
	3	2.39	89.7%	82.2%	81.5%	80.7%	82.3%
ELCAP	1	2.41	87.3%	80.06%	81.3%	83.7%	79.03%
	2	2.38	89.01%	83.03%	79.9%	81.9%	78.01%
	3	2.25	88.5%	80.2%	81.05%	82.1%	80.04%
5-Patients	1	2.43	88.4%	84.04%	82.1%	82.2%	82.05%
	2	2.20	89.03%	83.08	81.05%	79.3%	82.9%
	3	2.32	90.01%	84.2%	84.02%	84.0%	84.03%

C. PERFORMANCE MEASUREMENT USING SEGNET

SegNet comprises an encoder and decoder network and a pixel-wise classification layer. The encoder network has 13 convolutional layers that match the first 13 convolutional layers of the VGG16 [37]. We deleted the eventually linked layers from the SegNet encoder network to reduce the number of parameters and keep higher resolution feature maps at the deepest encoder output. There are 13 layers in the decoder network, with one for each encoder layer. A multi-class softmax classifier receives the output of the final decoder and generates class probabilities for each pixel. Each encoder in the encoder network performs convolution with a filter bank to produce a set of feature maps. Then, each element is given a rectified-linear nonlinearity (ReLU) [38]. The TCIA, ELCAP, and 5-patient datasets were used independently to train the network, and the test dataset combinations were used to evaluate it. In Table 4, the performance is mentioned.

Table 4. SegNet Model's Performance

Dataset	Combo	FP/Vol	Accuracy	Dice	F1	Recall	Precision
TCIA	1	2.43	82.8%	72.4%	75.7%	75.7%	75.7%
	2	2.50	86.5%	74.9%	78.3%	87.1%	71.1%
	3	2.71	87.6%	77.6%	76.4%	73.6%	79.4%
ELCAP	1	3.02	86.01%	74.1%	79.01%	77.1%	81.06%
	2	2.78	88.03%	75.02%	77.04%	73.4%	81.05%
	3	3.00	88.5%	76.1%	78.02%	74.4%	82.03%
5-Patients	1	2.56	87.05%	75.1%	78.03%	75.2%	81.09%
	2	2.39	89.01%	77.05%	77.02%	71.8%	83.08%
	3	2.28	89.8%	78.02%	80.01%	76.3%	84.00%

SegNet achieved the highest accuracy and precision on the 5-patient training and COMBO_3 test dataset shown in Table 4. The model performed poorly than AlexNet, Old_U-Net, and Auto-LNDS, but the model's precision on the ELCAP training dataset was more reasonable than Auto-LNDS and AlexNet.

D. PERFORMANCE MEASUREMENT USING AUTO-LNDS (MODIFIED MASK RCNN)

Auto-LNDS employs the Mask RCNN [39] as its primary network, which is a system capable of efficiently identifying objects in an image while producing a high-quality segmentation mask for each instance. The Auto-LNDS consists of the Feature Pyramid Network (FPN) [40], the Region Proposal Network (RPN) [41], and the head network Mask R-CNN. To overcome the degradation problem and

enable training of the deeper network, Resnet-101 was selected as the backbone of Mask R-CNN [42]. Since the quantity and size of LNs vary from patient to patient, Mask R-CNN was chosen for nodal recognition and segmentation. Table 5 summarizes the performance of the Auto-LNDS model, which was trained on TCIA, ELCAP, and the 5-patients dataset separately and tested on various combinations of test datasets.

Table 5. Auto-LNDS Model's Performance

Dataset	Combo	FP/Vol	Accuracy	Dice	F1	Recall	Precision
TCIA	1	3.01	89.4%	85.7%	81.0%	81.5%	80.5%
	2	3.75	85.9%	80.1%	82.6%	84.9%	80.4%
	3	2.62	87.4%	84.7%	80.1%	77.5%	82.9%
ELCAP	1	3.11	87.03%	75.08%	76.01%	78.1%	74.03%
	2	2.90	88.02%	77.3%	79.01%	78.9%	79.1%
	3	2.73	87.1%	74.7%	77.9%	77.7%	78.06%
5-Patients	1	2.33	87.5%	76.09%	78.02%	76.1%	80.07%
	2	2.89	87.9%	78.03%	78.05%	74.4%	82.02%
	3	2.40	89.02%	78.9%	81.03%	77.4%	85.05%

The model outperformed itself in terms of accuracy when using the TCIA dataset and COMBO_1 test dataset. Notably, the model trained on the 5-patients dataset and tested on COMBO_3 achieved a higher precision score than those trained on TCIA and ELCAP datasets. The Auto-LNDS model achieved a precision score of 85.05%, which was higher than that of AlexNet and SegNet but lower than that of Old_U-Net. However, regarding the accuracy, the Auto-LNDS model ranked third among the four models.

7. DETECTION, SEGMENTATION, AND PRECISION CAPABILITY ANALYSIS

Table 6 showcases several instances of how effectively different models segment lymph nodes. Upon examining the segmented images, medical experts have unanimously concluded that each model accurately divided the lymph node parts in the first row. However, some models, including Old U-Net, AlexNet, and Auto-LNDS, missed out on a lymph node part in the second row. SegNet and the proposed model, on the other hand, performed better by correctly segmenting the data. Factors such as the small size of the section, poor image quality, or the proximity of two challenging-to-disentangle portions could have contributed to this issue. Row three also presented similar challenges, but the Auto-LNDS and suggested model outperformed the rest. Apart from Old U-Net and SegNet, all models correctly segmented the lymph node in row four. Only the proposed model could detect the tiny lymph node section in row five, which other models missed due to its small size, iso-intensity, or partial volume effect. The models accurately identified the lymph node parts in rows six and seven. However, row eight presented a similar issue to row five, but the proposed model was able to segment the lymph node part appropriately. Medical professionals have commended the proposed approach for successfully segmenting small or closely spaced lymph node sections.

Table 6. Lymph Node Segmentation and Prediction by Several Models

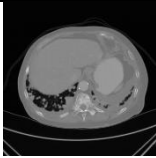
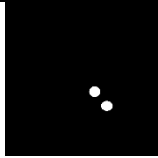
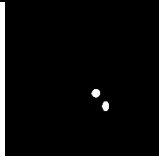
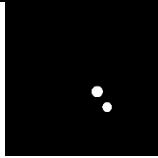
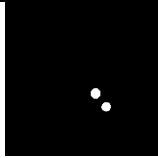
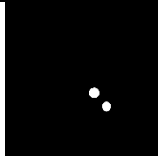
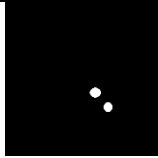
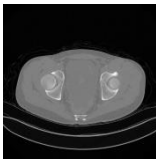
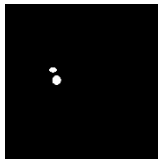
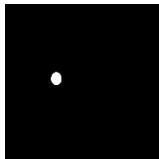
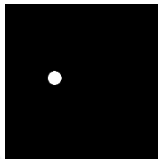
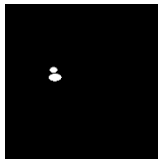
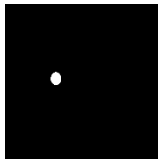
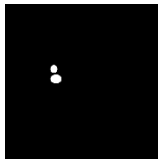
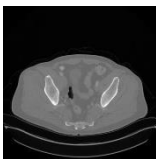
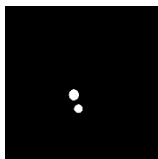
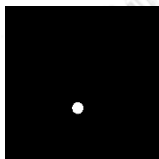
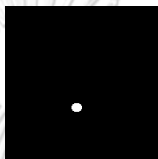
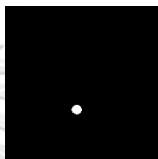
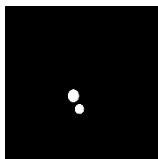
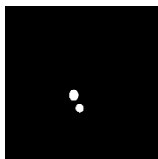
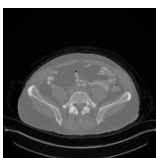


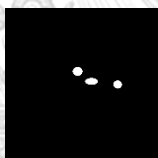
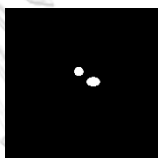


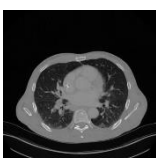
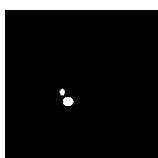
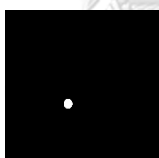
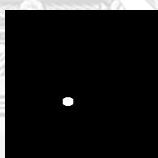
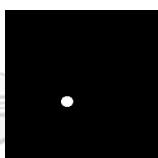
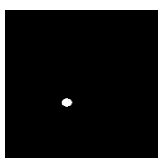
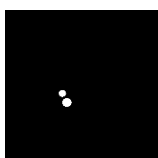
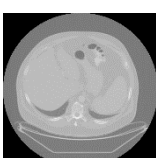
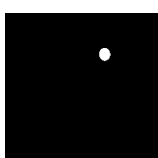
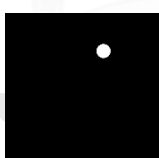
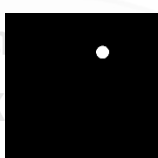
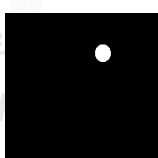
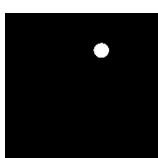
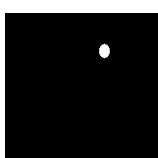
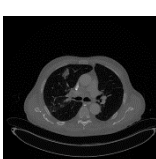
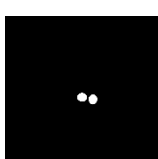
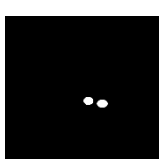
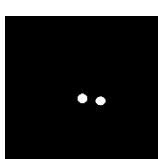
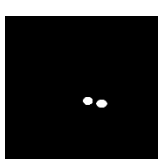
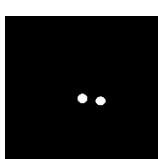
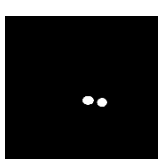

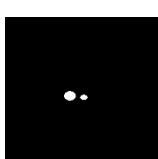
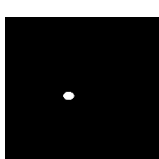
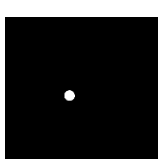
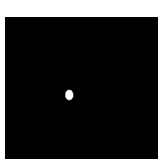
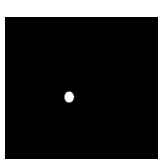
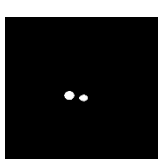






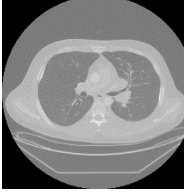

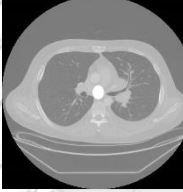
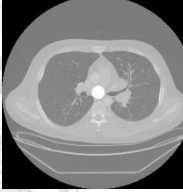
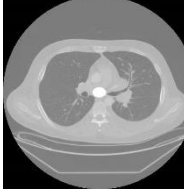
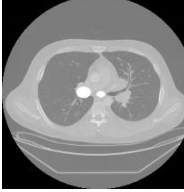
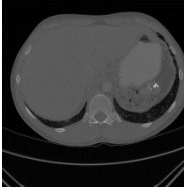
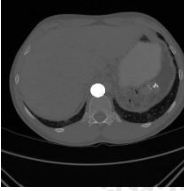
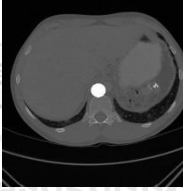
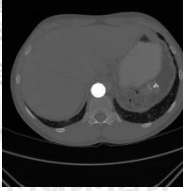
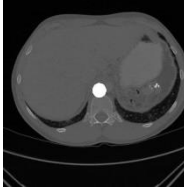
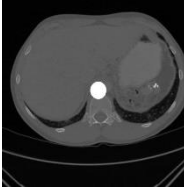
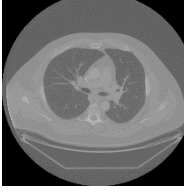






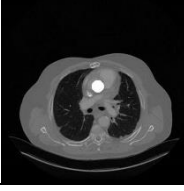
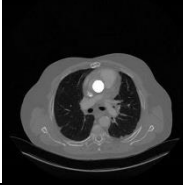
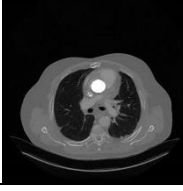
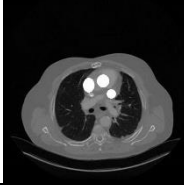
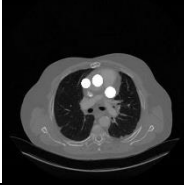
CT Image	Actual	Precision				Proposed Model
		Old_U-Net	AlexNet	SegNet	Auto-LNDS	
						
						
						
						
						
						
						
						

Table 7 provides examples of how various models have been utilized to find lymph nodes. Upon examining the detected photographs, medical experts have concluded that the proposed model correctly identified lymph nodes in both large and small sizes, starting from the fifth column. Due to the thin

dimension, some versions of models, including Auto-LNDS, Old U-Net, SegNet, and AlexNet, skipped one lymph node in the second row. In row four, AlexNet and SegNet missed one lymph node due to low image resolution and an unobtrusive display on the lymph node image. Two lymph nodes in row five were disregarded by Old U-Net, AlexNet, and SegNet due to poor picture resolution or partial volume effects. Medical experts have asserted that the proposed model's performance is superior to others in recognizing lymph node portions affected by the minute shape, size, insufficient resolution, or partial volume effect.

Table 7. Lymph Node Detection by Different Models

CT Image	LN Detection				
	Old_U-Net	AlexNet	SegNet	Auto-LNDS	Proposed Model
					
					
					
					
					

CHAPTER 5: DISCUSSION & LIMITATION

The initial step in treatment involves identifying every LN, which can be a time-consuming and laborious task. However, we have developed an innovative deep-learning approach that enables quick and accurate detection and segmentation of mediastinal LNs. With the help of this technology, radiologists can now obtain an LN map quickly, which can aid in diagnostic interpretation. This is the first attempt to detect and segment LNs using CT data automatically. Previous efforts were focused on contrast-enhanced or small-diameter LNs, utilizing various methods. Moreover, earlier methods struggled with CT images that have tiny LN sections. Several methods overlapped or missed the small LN component during segmenting. To address this, we prioritized the minor component to reduce knowledge loss during training.

Our approach involved achieving information lossless upsampling through a hybridized bilinear interpolation method, while the TGV-based methodology reduced noise and smoothed the image. The vivid contrast and quality of the processed datasets significantly contributed to accuracy, allowing the model to outperform cutting-edge methods with a maximum accuracy score of 94.8%, a Jaccard score of 91.9%, a recall score of 94.1%, and a precision score of 93.1%.

However, there are several limitations to this thesis. Our primary objective was to identify and separate mediastinal lymph nodes from CT scans. We used a modified deep-learning algorithm called UNet++ to address slight concerns with chunk detection in medical photos. The proposed approach aimed to preserve texture discontinuities, pick noisy regions, find the right balance points by backpropagation, and restore image resolution. Due to time constraints, we utilized high-resolution CT images instead of contrast-enhanced CT pictures for training and testing. While we examined three datasets and compared the model's accuracy with four different methodologies, more open datasets and methods are available. Additionally, the model may not perform well in real-time LN detection since detection and segmentation quality were given more weight in this thesis.

CHAPTER 6: CONCLUSION

The identification of lymph nodes is crucial in detecting cancer and providing patients with the necessary treatment. This thesis presents an innovative approach to segment and detect mediastinal lymph nodes from CT images using a modified UNet++ model. The model's efficiency was enhanced using a combination of bilinear interpolation and a total generalized variation (TGV)-based upsampling technique. With the aid of medical experts, a dataset containing three different combinations was processed. The performance metrics for the modified UNet++ model were evaluated against state-of-the-art tactics and outperformed other methods with an impressive 94.8% accuracy and 93.1% precision accuracy, with a minimal 4.7 false-positive rate per volume. The model's performance was calculated across multiple datasets, with a satisfactory outcome achieved. It can be confidently inferred that the suggested technique provides precise detection and segmentation of mediastinal lymph nodes using the UNet++ model. Future plans involve optimizing the filter settings to reduce the image's noise and further enhance the model's performance.



REFERENCES

- [1] L. A. Torre, R. L. Siegel, and A. Jemal, "Lung cancer statistics," *Adv. Exp. Med. Biol.*, vol. 893, pp. 1-19, 2016 2016, doi: 10.1007/978-3-319-24223-1_1.
- [2] W.-H. L. P. Yates, C. Dixon, "Taking action to improve outcomes for men living with and beyond prostate cancer," *Austral. Nurs. Midwifery J. Big Data*, vol. 24, no. 6, p. 35, 2016.
- [3] D. Sarrut *et al.*, "Learning directional relative positions between mediastinal lymph node stations and organs: Segmentation of mediastinal lymph nodes stations," *Med. Phys.*, vol. 41, no. 6, p. 061905, 2014 2014, doi: 10.1118/1.4873677.
- [4] G. Xu *et al.*, "Thoracic lymph node station recognition on CT images based on automatic anatomy recognition with an optimal parent strategy," *Proc. SPIE*, vol. 10574, 2018 2018, doi: 10.1117/12.2293258.
- [5] M. Feuerstein, B. Glocker, T. Kitasaka, Y. Nakamura, S. Iwano, and K. Mori, "Mediastinal atlas creation from 3-D chest computed tomography images: application to automated detection and station mapping of lymph nodes," *Med. Image Anal.*, vol. 16, no. 1, pp. 63-74, 2012 2012, doi: 10.1016/j.media.2011.05.005.
- [6] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60-88, 2017 2017, doi: 10.1016/j.media.2017.07.005.
- [7] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640-651, 2017 2017, doi: 10.1109/TPAMI.2016.2572683.
- [8] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481-2495, 2017 2017, doi: 10.1109/TPAMI.2016.2644615.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2015, pp. 234-241.
- [10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834-848, 2018 2018, doi: 10.1109/TPAMI.2017.2699184.
- [11] I. Noguees *et al.*, "Automatic lymph node cluster segmentation using holistically-nested neural networks and structured optimization in CT images," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer International Publishing, 2016,

pp. 388-397.

- [12] H. Oda *et al.*, "Dense volumetric detection and segmentation of mediastinal lymph nodes in chest CT images," presented at the Medical Imaging 2018: Computer-Aided Diagnosis, 2018, 2018.
- [13] Image-net.org. "ImageNet." <https://image-net.org/index.php> (accessed Jun. 25, 2022).
- [14] A. Barbu, M. Suehling, X. Xu, D. Liu, S. K. Zhou, and D. Comaniciu, "Automatic detection and segmentation of lymph nodes from CT data," *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 240-250, 2012 2012, doi: 10.1109/TMI.2011.2168234.
- [15] D. Bouget, A. Pedersen, J. Vanel, H. O. Leira, and T. Langø, "Mediastinal lymph nodes segmentation using 3D convolutional neural network ensembles and anatomical priors guiding," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, pp. 1-15, 2022 2022, doi: 10.1080/21681163.2022.2043778.
- [16] H. Tekchandani, S. Verma, and N. D. Londhe, "Improving the detection of abdominal and mediastinal lymph nodes in CT images using attention U-net based deep learning model," in *Artificial Intelligence Applications for Health Care*. Boca Raton: CRC Press, 2022, pp. 181-202.
- [17] A. Seff *et al.*, "2D view aggregation for lymph node detection using a shallow hierarchy of linear classifiers," *Med. Image Comput. Comput. Assist. Interv.*, vol. 17, no. Pt 1, pp. 544-552, 2014 2014, doi: 10.1007/978-3-319-10404-1_68.
- [18] A. Seff, L. Lu, A. Barbu, H. Roth, H.-C. Shin, and R. M. Summers, "Leveraging mid-level semantic boundary cues for automated lymph node detection," in *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2015, pp. 53-61.
- [19] H. R. Roth *et al.*, "A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations," *Med. Image Comput. Comput. Assist. Interv.*, vol. 17, no. Pt 1, pp. 520-527, 2014 2014, doi: 10.1007/978-3-319-10404-1_65.
- [20] Z. Zhu *et al.*, "Lymph node gross tumor volume detection and segmentation via distance-based gating using 3D CT/PET imaging in radiotherapy," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Cham: Springer International Publishing, 2020, pp. 753-762.
- [21] J. Liu *et al.*, "Mediastinal lymph node detection on thoracic CT scans using spatial prior from multi-atlas label fusion," presented at the Medical Imaging 2014: Computer-Aided Diagnosis, 2014, 2014.
- [22] J. Liu *et al.*, "Mediastinal lymph node detection and station mapping on chest CT using spatial priors and random forest: Mediastinal lymph node detection and station mapping," *Med. Phys.*, vol. 43, no. 7, p. 4362, 2016 2016, doi: 10.1118/1.4954009.
- [23] G. Xu, H. Cao, Y. Dong, C. Yue, K. Li, and Y. Tong, "Focal loss function based DeepLabv3+

- for pathological lymph node segmentation on PET/CT," presented at the Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing, 2020, 2020.
- [24] H. Lai, W. Chen, and H. Fu, "A new double-sampling method for mediastinal lymph nodes detection by deep conventional neural network," presented at the 2018 Chinese Control And Decision Conference (CCDC), 2018, 2018.
- [25] M. P. Paing, C. Pintavirooj, S. Tungjikusolmun, K. P. Win, and K. Hamamoto, "Automatic detection of mediastinal lymph nodes using 3D convolutional neural network," presented at the Proceedings of the 2019 4th International Conference on Biomedical Imaging, Signal Processing, 2019, 2019.
- [26] J. Singh, Y. Iwahori, M. Bhuyan, H. Usami, T. Oshiro, and Y. Shimizu, "Mediastinal lymph node detection using deep learning," presented at the Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods, 2020, 2020.
- [27] T. Ono *et al.*, "Detection of lymph nodes using CNN from contrast-enhanced CT images," presented at the 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), 2020, 2020.
- [28] Wiki.cancerimagingarchive.net. "A large-scale CT and PET/CT dataset for lung cancer diagnosis (lung-PET-CT-Dx) The cancer imaging archive (TCIA) public access-Cancer imaging archive Wiki." <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70224216> (accessed Jun. 25, 2022., 2022).
- [29] D. Data. <https://www.creatis.insa-lyon.fr/lymph-stations-atlas/data/14a2770a2e.php?token=QUwtQUtISVIgTkFZQU4tdGltZS1TYXQsIDE0IEF1ZyAyMDIx> (accessed Jun. 25, 2022., 2022).
- [30] "Public Lung Image Databases," in *Computer-Aided Detection and Diagnosis in Medical Imaging*: CRC Press, 2015, pp. 218-229.
- [31] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, 2019 2019, doi: 10.1186/s40537-019-0197-0.
- [32] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer International Publishing, 2016, pp. 424-432.
- [33] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018)*, vol. 11045, pp. 3-11, 2018 2018, doi: 10.1007/978-3-030-00889-5_1.
- [34] X. Zhao *et al.*, "Deep learning-based fully automated detection and segmentation of lymph nodes

- on multiparametric-mri for rectal cancer: A multicentre study," *EBioMedicine*, vol. 56, no. 102780, p. 102780, 2020 2020, doi: 10.1016/j.ebiom.2020.102780.
- [35] H. B. Alwan and K. R. Ku-Mahamud, "Cancellable Face Biometrics Template Using AlexNet," in *Communications in Computer and Information Science*. Cham: Springer International Publishing, 2020, pp. 336-348.
- [36] T. Shanthi and R. S. Sabeenian, "Modified Alexnet architecture for classification of diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 76, pp. 56-64, 2019 2019, doi: 10.1016/j.compeleceng.2019.03.004.
- [37] J. Akther, M. Harun-Or-Roshid, A.-A. Nayan, and M. G. Kibria, "Transfer learning on VGG16 for the classification of potato leaves infected by blight diseases," presented at the 2021 Emerging Technology in Computing, Communication and Electronics (ETCCE), 2021, 2021.
- [38] G. R. Padalkar and M. B. Khambete, "Analysis of basic-SegNet architecture with variations in training options," in *Advances in Intelligent Systems and Computing*. Cham: Springer International Publishing, 2020, pp. 727-735.
- [39] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," presented at the 2017 IEEE International Conference on Computer Vision (ICCV), 2017, 2017.
- [40] Q. Wang, L. Zhou, Y. Yao, Y. Wang, J. Li, and W. Yang, "An Interconnected Feature Pyramid Networks for object detection," *J. Vis. Commun. Image Represent.*, vol. 79, no. 103260, p. 103260, 2021 2021, doi: 10.1016/j.jvcir.2021.103260.
- [41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, 2017 2017, doi: 10.1109/TPAMI.2016.2577031.
- [42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 2016.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

NAME Al-Akhir Nayan

DATE OF BIRTH 28 November 1995

PLACE OF BIRTH Satkhira, Bangladesh

INSTITUTIONS ATTENDED Chulalongkorn University

HOME ADDRESS Badaghata, Nokipur, Shyamnagar, Satkhira, Bangladesh

PUBLICATION A. A. Nayan, B. Kijirikul, Y. Iwahori, "Mediastinal Lymph Node Detection and Segmentation using Deep Learning," IEEE Access, vol. 10, pp. 89289-89307, 2022, doi: 10.1109/ACCESS.2022.3198996.

AWARD RECEIVED Chulalongkorn University's Graduate Scholarship Program for ASEAN or Non-ASEAN Countries.

