

**COMPARISON BETWEEN MANUAL AND AUTOMATED SEGMENTATION  
ALGORITHMS IN COMPUTED TOMOGRAPHY IMAGING OF THE LIVER.**



**BY**

**IYEMIFOKHAE CALEB EWOGBAI**

**(BMS2001299)**

**PRESENTED TO**

**THE DEPARTMENT OF RADIOGRAPHY,  
SCHOOL OF BASIC MEDICAL SCIENCES,**

**UNIVERSITY OF BENIN**

**Supervisor Name: Akpobasahan E.A.(Dr.)**

**OCTOBER, 2025.**

**CERTIFICATION**

This is to certify that this project was carried out by Iyemifokhae Caleb Ewogbai with matriculation number (BMS2001299) of the Department of Radiography, School of Basic Medical Science, University of Benin, Benin City, Edo State, Nigeria.

---

**AKPOBASAHAN E.A.(DR.)**  
**(PROJECT SUPERVISOR)**

---

**DATE**

---

**PROF. MRS. F.O. IGBENEDION**  
**(HEAD OF DEPARTMENT)**

---

**DATE**

---

**EXTERNAL EXAMINER.**

---

**DATE**

## **DEDICATION**

This work is dedicated to God Almighty for his goodness and mercies throughout my stay in school. My parents Mr. and Mrs. Iyemifokhae for their love, care and support from the moment I was born till this point in my life.

## ACKNOWLEDGEMENT

My sincere appreciation goes to God Almighty for his Goodness and mercies towards my life and for seeing me throughout my studies. I also want to appreciate my project supervisor Dr. E.A. Akpobasahan for her encouragement and constructive criticism and corrections throughout the duration of this project. I would also like to acknowledge the Head of Department Radiography, Prof. Mrs. F.O. Igbenedion and other lecturers of the department of radiography.

I want to also specially thank Dr. G.N. Okungbowa and Chief Victor Chimezie whose help was vital in the completion of this project, I say God bless you.

I wish to acknowledge my wonderful parents Mr. and Mrs. Iyemifokhae , my siblings Iyemifokhae Joshua, Iyemifokhae Shalom, Iyemifokhae Esther and Iyemifokhae Favour for their loving kindness and support throughout my stay in school. Also, for being a source of joy and blessing in my life.

I also wish to appreciate my pastors. Reverend Amos and Grace Ehizua, Pastor Nosa and Mercy Asemota, Pastor Gilbert Ononyume for their spiritual guidance, support and prayers.

To my friends Adeyemi Fadekemi Esther, Iwu Akachi Ann and Oludare Ayomide Samuel who has made my stay in UNIBEN a joyful one, I say thank you and I pray that God bless you bountifully in Jesus name, Amen.

## TABLE OF CONTENTS

Title page .....	i
Table of Contents .....	ii
<b>Chapter one: Introduction</b>	
1.1 Background of Study .....	1
1.2 Statement of Problem.....	3
1.3 Research Questions .....	4
1.4 Hypothesis.....	4
1.5 Aims and Objectives of Study .....	5
1.6 Significance of Study .....	6
1.7 Scope of the Study .....	8
1.8 Operational Definition of Terms.....	8
<b>Chapter Two: Literature Review</b>	
2.1 Conceptual Review .....	11
2.2 Empirical Review.....	23
<b>Chapter Three: Research Methodology</b>	
3.1 Research Setting .....	27
3.2 Study Design .....	27

3.3 Target Population .....	28
3.4 Sampling Techniques / Sample Size .....	28
3.5 Instruments for Data Collection .....	28
3.6 Validity of Instruments .....	29
3.7 Reliability of Instruments .....	30
3.8 Methods of Data Collection .....	31
3.9 Methods of Data Analysis .....	31
3.10 Ethical Considerations .....	32
 <b>Chapter Four: Result and Discussion</b>	
4.1 Presentation (Data Presentation) .....	33
4.2 Discussion .....	40
 <b>Chapter Five: Conclusion and Recommendation</b>	
5.1 Conclusion .....	42
5.2 Discussion .....	43
5.3 Limitations of study .....	43
5.4 Suggestions for further studies.....	44
<b>References</b> .....	<b>45</b>
Appendix I: Data Collection Sheet .....	53
Appendix II: Ethical Approval .....	54
Appendix III: Python code used to calculate the DSC .....	55

## ABSTRACT

In this study, we compared manual and automated segmentation algorithms in Computed Tomography (CT) imaging of the liver. Segmentation is the delineation of anatomical structures on a CT/MRI image for treatment and dose planning purposes. Accuracy in the segmentation of organs in CT imaging is a critical step in diagnosis and therapy planning. Normally, manual segmentation performed by trained practitioners is the gold standard for the delineation of these anatomical structures. However, manual segmentation is time-consuming, subject to inter-observer differences, and also needs a level of expertise (Lee et al., 2024). Automated segmentation on the other hand, is a potential solution to the problems faced due to manual segmentation, despite this development, there is still an ongoing debate questioning the accuracy and reliability of automated segmentation due to little knowledge about its capabilities (Raudaschl et al., 2017). The main objectives of this study are to compare the accuracy and segmentation time of manual and automated segmentation, also to assess the inter-observer variability between 10 different observers. This study was carried out at the University of Benin Teaching Hospital (UBTH) using an anonymized CT dataset segmented manually using 3D-slicer and automatically using TotalSegmentator. A quantitative analysis of 100 segmentation masks produced by the segmentation of 10 anonymized CT images of patients by 10 observers (radiographers and radiotherapists) was conducted to evaluate the accuracy, time efficiency of automated algorithms and inter-observer variability of manual segmentation. The accuracy of automated segmentation was measured using the mean Dice Similarity Coefficient value, speed of both types of segmentation compared was measured using time and the inter-observer variability was measured using the Inter Class Correlation and Fleiss' Kappa. The results obtained indicates that automated segmentation is as accurate as manual segmentation with a mean DSC value of  $0.85 \pm 0.03$  and also

more time efficient with a 65% decrease in the segmentation time per image. This study also shows that the manual segmentations done by different observers are not far off from each other proving how reproducible it is. The Inter Class Correlation and Fleiss' Kappa was used to determine this with their score being 0.720 and 0.683 respectively. This indicates a moderate to strong agreement between segmentation done by different observers and if there is a strong agreement, the variability will be minimal. The DSC gotten is also greater than 0.7, giving a confirmation of how accurate automated segmentation is. From the results above, it was concluded that automated segmentation is as accurate and faster than manual segmentation. It was also concluded that there are minimal differences between segmentations done by different observers. A wide clinical adoption, Training of radiographers, educational inclusion and further research is recommended to increase the awareness and integration of automated segmentation in various clinical systems.

**Keywords:** Computed tomography, Segmentation, Dice Similarity Coefficient, Time efficiency, Inter-observer variability, Observers, Accuracy.

## **CHAPTER ONE**

### **INTRODUCTION**

#### **1.1 BACKGROUND OF STUDY**

The liver is a very vital organ in the human body. It performs crucial tasks and therefore identifying liver pathologies and making good therapeutic plans for these pathologies are critical. Nearly 25,000 Americans are killed by liver diseases annually (Zareei & Karimi, 2016). About 600,000 to 900,000 global deaths is caused by liver cancer annually. Liver related death is eleventh on the leaderboard of causes of death and liver cirrhosis is number ten on the leading causes of death in Africa today (Devarbhavi et al., 2023).

Computed-Tomography is an imaging modality with uses ionizing radiation to acquire images of the internal structures of the body in form of slices. These slices are then reconstructed to give a 3-dimensional representation of these structures. A major component of Ct imaging is segmentation and it involves marking out specific Regions of Interest (ROI) such as visceras and pathological regions. Traditionally, segmentations are done by radiologist manually, offering great precision it can be time consuming and stressful (Lenchik et al., 2019).

A variety of organ segmentation techniques can be used to identify the livers main position and also possible liver tumors (Zareei & Karimi, 2016). These techniques are: Manual segmentation,

semi-automated segmentation and automated segmentation. Our main focus is on manual and automated segmentation.

Segmentation in general means the process of dividing, but in medical imaging it involves dividing an image into parts so that organs within these different parts can be identified and isolated for individual analysis (Shen, 2022).

Manual segmentation being the “gold standard” is used to identify and isolate region of interests in CT images. Though it is done by trained personnel, there are inter-observer variabilities (Lee et al., 2024). It is also time consuming, hence making it inefficient (Fel et al., 2023).

Automatic segmentation on the other hand is way easy and efficient. It is also time efficient with little to no variability. It is a segmentation technique created by training algorithms with already manually segmented images (Milara et al., 2024). This training is known as deep learning. Semi-automated segmentation combines both manual and automated techniques to help identify and isolate these structures.

Some studies comparing the efficiency of manual and automated segmentation methods has been done. Some of which are: manual versus artificial intelligence-based segmentations as a pre-processing step in whole-body PET dosimetry calculations which indicated that AI-based methods could achieve effective dose estimations within a 2% margin of manual methods (Van Sluis et al., 2022).

Despite these developments, there are still some challenges. Automated segmentation may still have some difficulties with complex anatomical variations or pathologies, which gives rise to the necessity of refinement and validation. Even with these struggles, integration of automated segmentation into the healthcare systems and clinical workflow can help boost efficiency and reduce the work load on trained health-care practitioners, fast-tracking patient care and improving their outcome.

## **1.2 STATEMENT OF PROBLEM**

Accuracy in the segmentation of organs in CT imaging is a critical step in diagnosis and therapy planning. Normally, manual segmentation performed by trained practitioners is the gold standard for the delineation of these anatomical structures. However, manual segmentation is time-consuming, subject to inter-observer differences, and also needs a level of expertise (Lee et al., 2024). Automated segmentation on the other hand, is a potential solution to the problems faced due to manual segmentation. These solutions include: enhanced efficiency, reduced human error and reproducibility (exact or almost an exact replica) of segmented images for example, a deep-learning algorithm for CT liver and hepatic vessel segmentation (Li et al., 2024) achieved a DSC of 0.98 for the liver and 0.86–0.94 for vessels. Similarly, an MRI-based liver segmentation study done by Han et al., 2022 achieved a DSC of approximately 0.92 and reduced segmentation time from 26 minutes to 8 seconds. Despite the high accuracy and efficiency of automated segmentation algorithms, there is still a notable research gap which is that very few studies have been carried out in Nigeria or Africa to evaluate how well these automated segmentations perform on local CT

datasets. Currently, most literatures are based on data from Europe, Asia and North America. In these regions' patient characteristics, imaging protocols and pathological patterns differ from those in Africa. As a result, it remains a question as to how automated segmentation will perform compared to manual segmentation in a Nigerian Clinical setting and real-world practice.

### **1.3 RESEARCH QUESTIONS**

- How accurate is automated liver segmentation compared to manual segmentation?
- How variable is manual liver segmentation among observers?
- How much faster is automated liver segmentation than manual segmentation?

### **1.4 HYPOTHESIS**

H<sub>0</sub> (Null Hypothesis):

There is no significant difference between the accuracy of automated segmentation algorithms and manual segmentation in CT imaging.

H<sub>1</sub> (Alternative Hypothesis):

There is a significant difference between the accuracy of automated segmentation algorithms and manual segmentation in CT imaging.

H<sub>0</sub> (Null Hypothesis):

There is no significant difference in the time taken to perform liver segmentation using automated methods compared to manual segmentation.

H<sub>1</sub> (Alternative Hypothesis):

Automated liver segmentation significantly reduces the time required to perform liver segmentation compared to manual segmentation.

H<sub>0</sub> (Null Hypothesis):

There is no significant difference between liver segmentation results produced by two different manual observers.

H<sub>1</sub> (Alternative Hypothesis):

There is a significant difference in liver segmentation results between two different manual observers, indicating notable inter-observer variability.

## **1.5 AIMS AND OBJECTIVES OF STUDY**

Aim:

To compare the efficiency, accuracy, overlap and reproducibility between manual and automated segmentation algorithms in CT imaging.

Objectives:

- To evaluate the accuracy of automated segmentation relative to manual segmentation (ground truth).
- To assess the inter-observer variability in liver segmentation between ten different observers.
- To measure and compare the time of segmentation of the liver using automated and manual segmentation.

## **1.6 SIGNIFICANCE OF STUDY**

- Improved Clinical Efficiency.

This study will show how well automated liver segmentation can help reduce the time spent on manual segmentation, ultimately boosting the overall efficiency of clinical workflows. Faster segmentation leads to quicker diagnosis and treatment planning, benefiting patients in time-sensitive or emergency situations.

- Minimizing Inter-Observer Variability

This study is going to assess how different manual observers may produce varied liver segmentations, highlighting the extent of inter-observer variability. Understanding this is very critical in standardizing segmentation practices and could promote greater consistency in clinical interpretations across different healthcare professionals.

- Cost Effectiveness

Automated segmentation may reduce the need for labor in terms of manual segmentation of structures. Due to this development, the cost of segmentation reduced due to the reduced use of trained practitioners in the segmentation of anatomical structures. This study is going to highlight how effective segmentation is without the need of manual labour.

- Support For Artificial Intelligence Adoption In Radiography

This study will contribute massively to the growing body of evidence which supports the integration of AI-driven tools in medical imaging. The findings will encourage the adoption of automated systems, which can assist radiologists and radiographers in achieving better outcomes with enhanced precision and efficiency.

- Importance In Radio-Therapy

By producing accurate and reproducible liver segmentation, automated segmentation can help optimize the distribution of radiation dose to the tumor while sparing critical organs e.g. kidneys, stomach. This results in better treatment outcomes by improving tumor control while minimizing the risk of radiation-induced toxicity to surrounding tissues.

## 1.7 SCOPE OF THE STUDY

This study is a comparative study between manual and automated segmentation algorithms used in Computed Tomography imaging. The primary objective of this study is to assess the accuracy, time-efficiency and reproducibility of automated segmentation of the liver relative to its manual counterpart using the Dice Similarity Co-efficient, times of both segmentation and inter-observer variability. Suggestion: Manual segmentation is done manually by trained practitioners (radiographers, radiologists) using the 3D slicer software while automated segmentation is done with a deep-learning based model known as TotalSegmentator, using a cloud based CPU or GPU, Google colab. The research is done using Computed Tomography imaging dataset gotten from the University of Benin Teaching Hospital (UBTH) Radiology department. This study will not include any other imaging modality such as Magnetic Resonance Imaging and Positron Emission Tomography. The results of this study aim to shed light on how accurate and reliable automated segmentation is in comparison to manual segmentation.

## 1.8 OPERATIONAL DEFINITION OF TERMS

**Segmentation:** This is the process of dividing a medical image such as CT, MRI, PET into regions of interest, identifying anatomical structures within these regions, delineating these structures for individual analysis.

**Computed Tomography:** This is a medical imaging modality that uses ionizing radiation to view the internal structures of the body in form of slices. These slices are then reconstructed to give a 3-dimensional image of the structures. X-rays are taken from different angles to produce these slices.

**Mask:** This is a labeled image that highlights a specific region of interest within a scan. Each pixel or voxel indicates whether it belongs to a particular structure or not.

ROI (Region Of Interest): This is a specific area within an image focused on for analysis.

**Manual Segmentation:** This is a segmentation technique that involves human expertise in delineating structures for individual analysis. The expert delineates these structures slice by slice which is time consuming. It is considered the gold standard due to its accuracy.

**Automated Segmentation:** This segmentation technique uses deep learning to identify and delineate these structures without human input.

**Semi-Automated Segmentation:** This combines algorithms with user guidance to segment structures.

**DSC (Dice Similarity Co-efficient):** This is a statistical measure used to compare the similarity between 2 datasets (in our case segmentation masks).

**Python:** This is a high-level programming language widely used in medical imaging for data analysis, visualization and machine learning.

**Deep Learning:** The process by which an algorithm is trained to perform certain tasks. In our case trained with already manually segmented images to help in the automated segmentation of structures.

**Slices:** These are 2-dimensional cross-sectional images obtained from CT scans. They are stacked together to form 3-dimensional volumes which can be analyzed in detail.

## CHAPTER TWO

### LITERATURE REVIEW

#### 2.1 CONCEPTUAL REVIEW

##### **Anatomy of the liver**

For accurate segmentation of the liver, we need to know about its normal anatomy. Located in the upper abdomen the liver is the biggest and weightiest viscera in the body. It extends from the right hypochondrial to some part of the left hypochondrial region of the abdomen and is guarded by the lower ribs (Bazira, 2023). It is located majorly at the right upper quadrant of the abdomen and extends to the left upper quadrant. It is very essential such that it plays vital roles in the normal function of the body system. Some of these functions are: nutrient breakdown, energy storage, reduction of toxic drug effects (Mohajan, 2025). The liver has a weight of about (1,400 - 1,800)g which is about 2-3% of the total body weight and has a smooth surface. It is divided into 2 lobes by the falciform ligament. These are the right and left lobes of the liver. The right lobe of the liver is further divided into 2 parts: the caudate lobe and the lobe quadratus. These 2 lobes are divided by the hilar fissure which is transverse in orientation. This transverse hilar fissure is where the portal hepatis enters the liver (Sumadewi, 2023).

Although the liver performs a variety of functions, alcohol consumption, hepatitis, obesity and diabetes are huge risk factors of liver cancer (Araújo et al., 2021).

## **Radiographical representation of the liver in CT images**

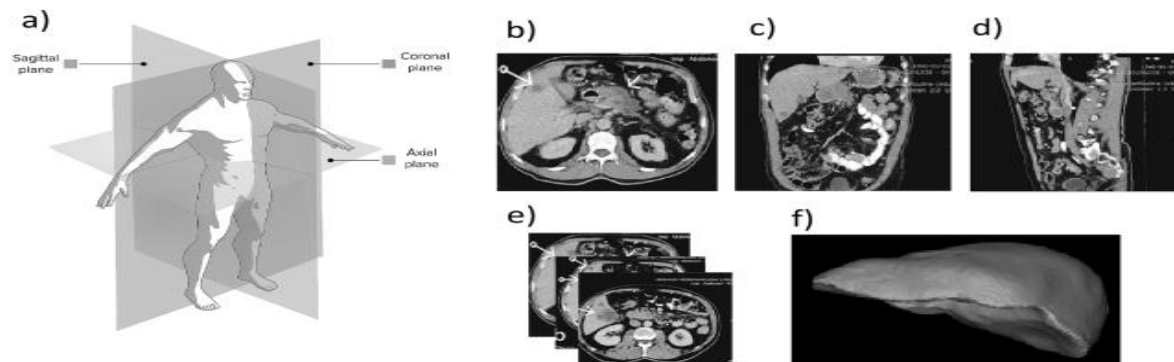
The liver is a very important organ in the human body and has several functions. Considering this, several radiographic modalities invented has been put into practice when assessing the structure and functions of the liver. Such modalities are: Computes Tomography (CT) , Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), PET-CT combination and Ultra sound. For tumor visualization, CT is the best while to soft tissue contrast visualization, MRI is superior. CT is the most preferred and traditionally used due to the fact that : a) It has a higher spatial resolution when compared to other modalities. b) Faster than MRI and PET-CT (Moghbel et al., 2017).

Traditionally CT involves acquiring 2d images in form of slices which represents tissues gotten consecutively in one of the 3 standard anatomical planes: coronal, sagittal or axial. Any of the structures in these slices can be marked (ROI) and then stacked on each other to form a 3-dimensional representation of the structure marked out (Delmoral & Tavares, 2024). The liver has intermediate soft tissue homogenous density with a Hounsfield unit of about 50 -70. It is denser than fat but less dense that bone and metals.

## **Segmentation and its types**

Segmentation is the partitioning or division of an image into different regions and then identifying anatomical structures within these regions and delineating them for analysis (Wang et al., 2021). These structures are then analyzed specifically, if possible reconstructed into a 3-dimensional

representation. Liver segmentation is a very important step in cancer diagnosis and treatment planning. Proper delineation and accurate identification of organs such as the liver aids in the medical assessment of the benefits and risks involved in interventional procedures (Araújo et al., 2021).



Source: Delmoral & Tavares, 2024

**Fig 2.1 Abdominal CT image, with a reconstructed liver. (a) All anatomical planes of the body. (b) An axial slice of an abdominal CT scan. (c) A coronal slice of an abdominal CT scan. (d) A sagittal slice of an abdominal CT scan. (e) Multiple coronal slices of an abdominal CT scan. (f) A 3-dimensional representation of a liver, reconstructed from the slices of a CT scan.**

There are majorly 3 types of segmentation: Manual, semi-automated and automated segmentation.

## **Manual Segmentation**

This is known as the “gold standard” of segmentation. It is done by trained individuals manually, without the aid of a computer program (Vinayahalingam et al., 2023). Although being the gold standard, manual segmentation has its drawbacks or limitations. These limitations can be: inter-observer variability, poor efficiency, cost, time inefficiency and stress to the individual segmenting.

## **Automated Segmentation**

This type of segmentation uses algorithms and deep learning techniques to segment anatomical structures automatically with little to no manual input. Unlike manual segmentation, it is more efficient and saves time but it needs high computing power and is a difficult task, especially when dealing with heterogeneous tumors (Dionísio et al., 2020).

## **Semi-automated Segmentation**

This type of segmentation combines both manual and automated segmentation by using both manual input and automatic algorithms in segmentation of anatomical structures. A good example is seeding and thresholding. Software algorithms are used to process data provided manually, these algorithms create a label which allow individuals edit the output from the algorithm. (Dionísio et al., 2020).

## **Semantic segmentation of liver structures in CT images**

In liver segmentation, the main objective is to assign semantic labels (names given to each region segmented e.g. the tumor, liver, vessels and background) to the pixels in the structures segmented or to the input image as a whole (Delmoral & Tavares, 2024).

Making computers fully able to automatically identify and separate parts of the liver in medical images is still very hard. The big challenge is making these methods work well on all kinds of images, not just the ones they were trained on (He et al., 2021). Big breakthroughs in segmentation algorithms have only happened recently, mostly in the last five years thanks to the rise of neural networks, especially deep learning.

## **Use of Totalsegmentator in automated tumor segmentation**

Totalsegmentator is an artificial intelligence-based tool used for the automated segmentation of structures in an MRI or CT image. This tool comes pre-trained on some specific segmentation softwares. Some of these segmentation softwares are: 3D slicer and MITK. Totalsegmentator is a pretrained artificial intelligence tool in the sense that it has already been trained by high quality segmented masks from different datasets. Due to this training and learning from these masks, it is able to automatically segment various anatomical structures in CT and MRI imaging. It is a deep learning model based on Convolutional neural network and U-Net architecture (Wasserthal et al., 2023).

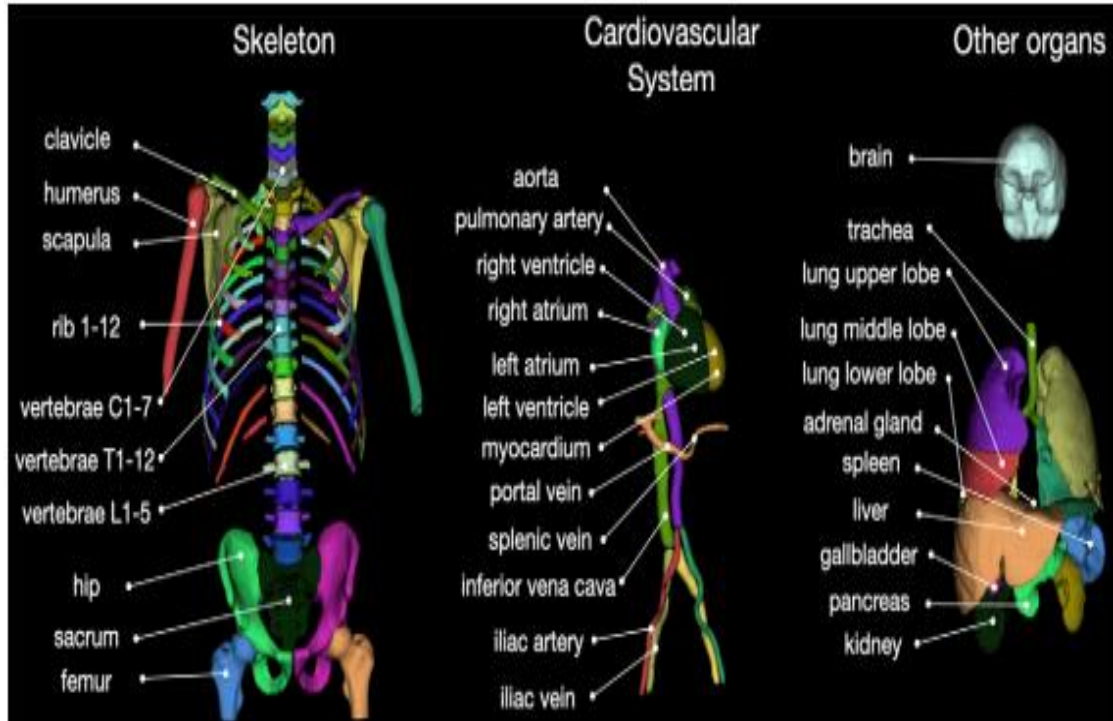
**Table 2.1 Differences between manual and automated segmentation**

<b>Manual Segmentation</b>	<b>Automate Segmentation</b>
<b>Done by trained professionals (radiologists, radiographers)</b>	Done by algorithms
<b>Time consuming</b>	Faster
<b>Less consistent due to inter-observer variability</b>	Reproducible
<b>Less efficient in segmentation of many structures</b>	Highly efficient, especially when segmenting alot of structures.
<b>Uses tools like brush, fill, grow from seed, ROI tracing, region-growing etc.</b>	Uses AI/ deep learning based tools like Totalsegmentator, MONAI label, etc.
<b>Requires prior knowledge or experience.</b>	Beginner friendly.
<b>Very easy to setup.</b>	Might be difficult to setup.

## **Clinical applications of segmentation**

### **Liver Tumor**

Hepatocellular carcinoma is cancer of the liver cells (hepatocytes), it is number one on the list of primary liver cancer and number five on the list of the most common tumors in the world (Okada et al., 2024). Hepatocellular carcinoma (HCC) develops in a complex way because it can be caused by many different things and influenced by several risk factors. Scientists still don't fully understand how it forms at the molecular level.



Source: Wasserthal et al., 2023

**Fig. 2.2 A 3-dimensional representation of segmented structures.**

HCC is a cancer that can result from various types of long-term liver damage, with viral hepatitis being the most common cause. No matter what starts the damage, the liver goes through a repeated cycle of injury and repair, which eventually leads to genetic changes in the cells. These changes are what lead to cancer (Dhanasekaran et al., 2016). CT scans are now commonly used to detect HCC (liver cancer) and help plan treatment. To do this, doctors use a special type of CT called dynamic CT, where a contrast dye (iodine) is injected into the bloodstream.

The scan is done in four phases:

1. Unenhanced phase (before contrast)
2. Arterial phase
3. Portal venous phase
4. Equilibrium phase
5. Using all four phases helps doctors accurately identify HCC.

The arterial phase is split into early and late arterial phases. This helps doctors clearly see the blood vessels in the liver before surgery (Okada et al., 2024).

Segmentation is very important in detecting and measuring tumor size, assessment of the tumor growth and also guide surgery or radiotherapy planning.

### **Liver hemangiomas**

Hepatic hemangiomas are the most frequently occurring benign tumors of the liver. These lesions consist of abnormally formed, slow-moving venous channels and are found in up to one-fifth of individuals, though some studies report much lower rates (Sadick et al., 2018; Kacała et al., 2024). These tumors mainly consist of cells lining blood vessels, and they usually come from branches of the hepatic artery (Li et al., 2015)

Thanks to improved imaging techniques, hepatic hemangiomas are now found more often. When they are small or medium-sized (less than 4 cm), they usually don't cause symptoms and are simply monitored without treatment (Kacała et al., 2024).

Segmentation can be used to differentiate between cystic and solid lesions (hemangiomas).

## **Liver Transplant**

Liver transplantation is a proven treatment for severe liver damage that can't be reversed. However, many problems can arise shortly after or long after the surgery, and these can lead to serious illness or even death (Rastogi, 2022). The severely damaged liver is replaced with another liver in good condition. Some common liver pathologies that call for liver transplant are: Chronic liver diseases, Liver cancer, Acute liver failure. The new liver is gotten from a deceased donor or a living person willing to donate a part of his or her full liver (Olthoff et al., 2015). This possible due to the fact that the liver tissue has the ability to regenerate/ regrow enabling both livers in the donor and recipient recover its full size.

Segmentation has a huge role in liver transplant, this importance can be seen in:

### **Donor Evaluation (Living Donor Transplant)**

- The liver is delineated into segments based on the Couinaud system, which identifies eight distinct segments.
- Surgeons utilize this segmentation to determine which segment(s) or lobe can be safely excised without jeopardizing the donor's health.
- Precise segmentation is crucial for calculating the graft volume, ensuring it is sufficiently large for the recipient while remaining safe for the donor's survival.

## **Graft Volume Calculation**

- Segmentation enables accurate evaluation of the future liver remnant (FLR) as well as the transplanted graft.
- It aids in preventing complications such as:
  - Small-for-size syndrome (when the graft is too small for the recipient)
  - Donor liver failure (occurs if excessive liver tissue is removed)

## **Surgical Planning**

- Surgeons utilize liver segmentation maps to:
  - Strategize resection lines (indicating where to make cuts)
  - Prevent injury to major vessels (such as hepatic veins or portal branches)
  - Ensure sufficient bile drainage and vascular supply following a transplant.

## **Postoperative Monitoring**

- Following a transplant, segmentation plays a crucial role in:
  - Tracking the growth of the transplanted segment, particularly in cases involving living donors
  - Detecting localized complications such as rejection, infarcts, or biliary leaks by identifying which segment is impacted.

(Oh et al., 2024).

## Evaluation metrics for liver segmentation

These metrics are used to compare between manual and automated segmentation algorithms. They are:

- **Dice Similarity Coefficient (DSC)**

This tells / measures how similar two segmentation masks are. Automated and manual segmentation masks are usually the focus point for this measurement. It is the most popular metric used in segmentation comparison and has a value between 0 and 1. The higher the value the more similar both segmentation masks are. Automated segmentation masks are compared with manual segmentation masks, this is due to the fact that manual segmentation is the “gold standard”. A value of 0 tell us that there is no overlap or similarity between the two masks and a value of 1 tell us that both masks are a perfect match (Dreizin et al., 2019). The DSC value can be gotten using a DSC formula or by using a python code.

$$DSC = (2|A \cap B|) / (|A| + |B|)$$

Where:

A = predicted segmentation

B = ground truth segmentation

Though this formula can be used, using a python code to calculate the DSC value is way faster and easier. This is an example of a code used to calculate the DSC value of two segmentation masks of the liver.

```
import nibabel as nib
import numpy as np

def dice_coefficient(mask1, mask2):
    mask1 = mask1.astype(bool)
    mask2 = mask2.astype(bool)

    intersection = np.logical_and(mask1, mask2).sum()
    return 2. * intersection / (mask1.sum() + mask2.sum())

# Load the segmentations
manual_nii = nib.load('manual_seg.nii')
auto_nii = nib.load('auto_seg.nii')

manual_mask = manual_nii.get_fdata()
auto_mask = auto_nii.get_fdata()

# Optional: binarize if values are not strictly 0 or 1
manual_mask = manual_mask > 0
auto_mask = auto_mask > 0

# Compute DSC
dsc = dice_coefficient(manual_mask, auto_mask)
print(f'Dice Similarity Coefficient: {dsc:.4f}')
```

**Fig 3. Python code used for calculating the DSC value of an automatically segmented liver.**

- Processing Time.

This is the total time it takes to create a segmentation mask or the total time it takes to segment a particular structure. This particular metric is crucial when dealing with trauma patients or patients who are on transplant planning.

Processing time is shorter in automated segmentation. This can be measured using a stop watch in manual segmentation or directly from the segmentation software used for automated segmentation.

On an average it takes an individual 15 -30 mins to segment manually and 30 seconds -3 minutes to segment automatically. (Çelik et al., 2023).

## **2.2 EMPERICAL REVIEW**

When evaluating segmentation techniques for liver CT imaging, it is necessary to look into other studies that provide measurable data on performance, efficiency and clinical relevance. This section presents review of selected peer-reviewed studies which compares manual and automated segmentation algorithms using standard metrics such as: The Dice similarity coefficient, segmentation time and reproducibility. These studies tell us how each method performs in practice, highlighting their strengths, limitations and applicability in clinical workflows such as liver transplant, planning and trauma care.

- Accuracy of automated segmentation.

Alot of studies report high segmentation accuracy using algorithms. In 2023, Xie et al. performed a retrospective study where a 3D U-Net automated segmentation model was used to segment 178 CT images of the liver. Manual segmentation was first carried out by radiologists on 170 CT images of the liver, then the automated segmentation algorithm was trained with these manual segments and then used to segment 178 CT liver images. The dice similarity coefficient was used

to determine the accuracy of the automatically segmented mask. The DSC values gotten for segments I to VIII of the liver were  $0.93\pm 0.01$ ,  $0.94\pm 0.01$ ,  $0.93\pm 0.01$ ,  $0.93\pm 0.01$ ,  $0.94\pm 0.00$ ,  $0.95\pm 0.00$ ,  $0.95\pm 0.00$ , and  $0.95\pm 0.00$ , respectively. These DSC values indicate that automated segmentation gives an accurate and almost similar mask (segments) when compared with manual segmentation.

Another study done by Yan et al.2023, where a deep learning network was trained to segment Magnetic Resonance Imaging images of tissues including: tumors, pituitary glands, sphenoid sinus, brain, superior saddle cistern and the lateral ventricle. A dataset of about 133 images of patients having craniopharyngioma was used. A tumor and mean tissue segmentation DSC values of 0.951 and 0.8668 were gotten. This result gotten also signifies how accurate automated segmentation is when compared to the “gold standard”.

Kumaraswamy and Patil in 2021 did a study on automated segmentation of an MRI prostate image using Res-Net. The trained a U-Net and Res-Net algorithm with a 116-patient sample dataset. After segmentation and comparison, a DSC value of 0.938. This also shows high segmentation accuracy of automated segmentation algorithms, it also showed that the accuracy of segmentation can be enhanced greatly by utilizing a two-stage neural network.

- Segmentation time and efficiency

In 2022 Cayot et al. performed a retrospective study comparing the manual and automated segmentation of polycystic liver. They trained a CNN using 190 CT series. 41 CT series where

then used to test the algorithm, a radiologist performed manual segmentation also which was used as reference for all comparisons. While the manual needed about 22.4 mins for full segmentation, took the algorithms (CPU and GPU) an average of 5 and 2 Seconds respectively to segment the structures. This indicated that automated segmentation is way ahead in speed compared to manual segmentation.

- Inter-observer Variability

A study done in 2018 by Kiefer et al. on the inter and intra-observer variability of an anatomical landmark-based manual segmentation. The inter-observer variability was assessed using the Inter-class Correlation Coefficient value which was 0.99 for PDFF and 0.98 for CSA. This showed that manual segmentation is highly reproducible, indicating minimal inter-observer variability.

The empirical evidence given above has highlighted the accuracy, time efficiency of automated segmentation over the manual segmentation methods and also the inter-observer variability of manual segmentation. While manual segmentation remains a gold standard in terms of accuracy, especially in complex cases, automated tools have demonstrated high Dice scores, reduced processing times, and strong reproducibility across diverse imaging datasets. These findings support the growing integration of deep learning algorithms into clinical workflows, particularly in time-sensitive settings like transplant planning and trauma evaluation. Future research should continue refining these models for improved generalizability and real-time applicability in routine practice.

## CHAPTER THREE

### RESEARCH METHODOLOGY

#### 3.1 Research setting

This study was carried out using CT images gotten from the University of Benin Teaching Hospital (UBTH) Radiology department, Benin city, Edo state.

All processes from data processing, model deployment and evaluation were carried out in a cloud-based environment using Google Colab. Totalsegmentator was used to automatically segment the liver from the CT images. Manual segmentation and visualization tasks were conducted using the 3D slicer v5.10.0 software. Further data handling and analysis was performed using the latest python programming language (version 3.9, 3.10) with supporting libraries such as SimpleITK and NumPy. This study was done between July and August 2025. Since the CT images were gotten from UBTH radiology department, an ethical approval was required.

#### 3.2 Study Design

This study was a retrospective, quantitative and comparative study. It was aimed at evaluating the performance of automated liver segmentation tool in comparison with the manually annotated reference mask. This investigation was carried out using the CT images acquired from the University of Benin Teaching Hospital Radiology department.

### **3.3 Target Population**

This study focused on individuals that did abdominal CT scans. It also focused on radiographers in UBTH.

### **3.4 Sampling techniques/Sample size**

A random sampling technique was employed to minimize selection bias and ensure that the data used in this study represented a broad and unbiased range of cases. Abdominal CT images were obtained from the Radiology Department of the University of Benin Teaching Hospital (UBTH).

The first 10 abdominal CT images which satisfied the criteria below were selected:

- The images had to be of diagnostic quality, free from major motion artifacts or reconstruction errors.
- All scans were acquired using similar imaging parameters (e.g., slice thickness, matrix size, and field of view) to maintain uniformity during segmentation analysis.

A total of 10 radiographers were also selected from the Radiology Department of UBTH to perform the manual segmentation. The first 10 radiographers who satisfied the criteria below were approached directly and voluntarily agreed to participate in the study.

Inclusion criteria for observers included:

- Must be a licensed radiographer registered with the Radiographers Registration Board of Nigeria (RRBN).
- Must have a minimum of one year of clinical experience in CT imaging or diagnostic radiography.

Each of the ten selected radiographers manually segmented the same ten CT images, resulting in 100 manually segmented datasets (10 observers × 10 images). These manual results were then compared with automated segmentation outputs to assess accuracy, inter-observer variability, and segmentation time.

### **3.5 Instruments for data collection**

The CT unit in the University of Benin Teaching Hospital Radiology department was our main source of data. It gave us the CT data needed for this study.

The following instruments was used for data collection:

- Computed Tomography Machine: The CT machine has already been used to acquire images of the abdominal organs to be segmented. These images acquired are in the computer system in located in the CT suite within the radiology department.
- Compact Disc: The images acquired by the CT machine was transferred into the compact disc for proper data transfer to the various computer systems for segmentation.

- Data Collection Sheet: The data sheet was used to acquire the time each manual segmentation is done. The data sheet used is attached in appendix I.
- Stop watch: This instrument was used to measure the time of manual segmentation.
- Laptop: A Hp zbook 15 was used for this study with an intel core i5 processor and 8 gigabytes of ram.
- 3D slicer v5.10.0: This software was used for manual segmentation.
- Google Colab: This cloud-based Graphics processing unit and python version of 3.11 was used for automated segmentation.

### **3.6 Validity of instrument**

- Compact Disc: The compact disc ensured safe export and transfer of CT images for segmentation in DICOM format. It was checked by a senior radiographer at UBTH ensuring the data within it was accurate.
- Data Collection Sheet: The data collection sheet used in this study was checked and corrected by a medical physicist and project supervisor ensuring its validity.
- Stop Watch: The stopwatch used was properly calibrated, ensuring it was valid for this study.

### **3.7 Reliability of Instruments**

The instruments used in this study demonstrated a high degree of reliability. They consistently produced stable and reproducible outcomes. Each tool was selected based on its prior validation in similar research contexts and its capability to ensure consistent results.

### **3.8 Methods of data collection**

This study used a secondary data collection method (an existing data collected initially by someone else for a particular purpose, but now used for a new research objective). Existing data from the University of Benin Teaching Hospital (UBTH) Radiology Department CT machine was used. A data collection sheet was also used to record segmentation time manually.

### **3.9 Methods of data analysis.**

This study focused on evaluating accuracy, reproducibility, and inter-observer variability. The following statistical measures were used:

Dice Similarity Coefficient (DSC): The DSC was used specifically because it quantified the overlap between manual and automated segmentations, providing a measure of segmentation accuracy. Higher DSC values indicate greater similarity between the segmentations. To obtain the DSC value python version 3.11 was used.

Intra-Class Correlation (ICC): ICC was used because it assesses agreement and consistency among different observers in manual segmentation, indicating inter-observer reliability. To achieve an accurate inter class correlation coefficient value SPSS v28 was used.

Fleiss' Kappa: Fleiss' Kappa was used because it measures categorical agreement among multiple observers, accounting for chance agreement and providing a robust measure of reproducibility. SPSSv28 was used to obtain the fleiss' kappa value.

Mean Segmentation Time: The mean time for manual segmentation across all observers was calculated, as well as the mean time for automated segmentation. This comparison quantified the efficiency gains of automated methods relative to manual segmentation. SPSS v28 was used to calculate the mean segmentation time for manual and automated segmentation.

### **3.10 Ethical considerations**

Ethical approval was acquired from the Health Research Ethics Committee of University of Benin Teaching hospital (UBTH). Anonymity of patient and confidentiality was assured. No personal identifiers were collected and the data acquired was used strictly for research purposes. The ethical approval request is attached in appendix II.

## CHAPTER FOUR

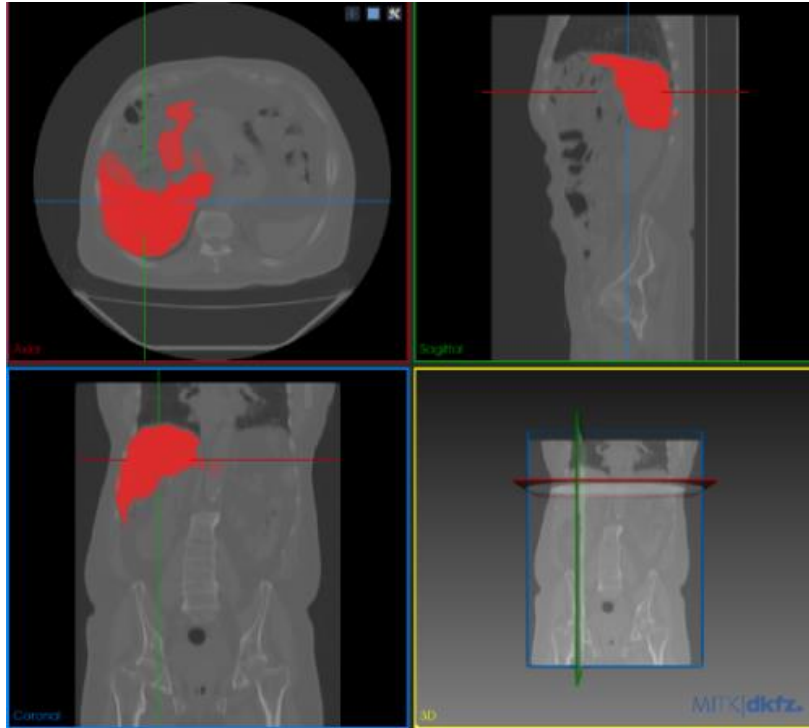
### RESULT AND DISCUSSION

In this chapter, we will discuss the results gotten by comparing manual and automated segmentation algorithms in computed tomography imaging of the liver. The analysis focuses deeply on the accuracy of automated segmentation, inter-observer variability and the time efficiency of automated segmentation compared to manual.

#### **4.1 Presentation (Data Representation)**

The following are the main results obtained during this study:

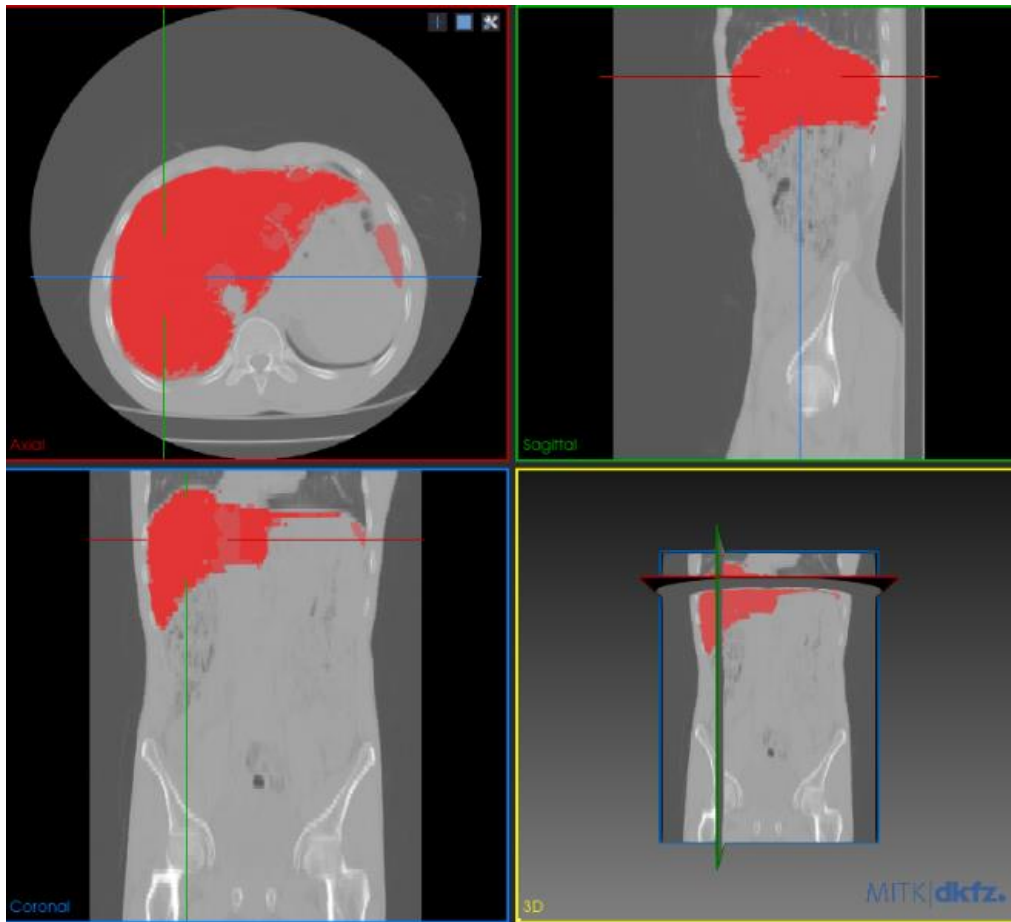
- Dice Similarity Coefficient (DSC):  
 $0.85 \pm 0.035$
- Average Segmentation Time:  
Manual: 25.9 minutes.  
Automated: 9 minutes.
- Inter-observer Agreement (Manual Segmentation):  
Intra-Class Correlation (ICC): 0.72  
Fleiss' Kappa: 0.68



**Figure 4.1. A patient chosen at random out of the 10 CT images used in this study. Different observers segmented the liver using 3D-slicer. All manually segmented liver is in red and overlap each other.**

Figure 4.1 presents manual segmentation overlap across different observers and figure 4.2 presents a manual and automated segmentation overlap between one observer and the automated algorithm used from the 10 patients used in this study. The 10 observers segmentation of the liver is seen in red in the first image while on the second image, the automated segmented structure is seen in light red and the manually segmented liver is seen in dark red. Both figure present manual and automated segmentation respectively in different planes (Axial, sagittal and coronal planes). The results obtained in this study will

help quantify the differences and similarities between observers and automated segmentation algorithms.



**Figure 4.2.** A patient chosen at random out of the 10 CT images used in this study. 1 observer's manual segmentation and an automated segmentation done using TotalSegmentator overlapped on each other. The manual segmentation mask is dark red and the automated mask is light red.

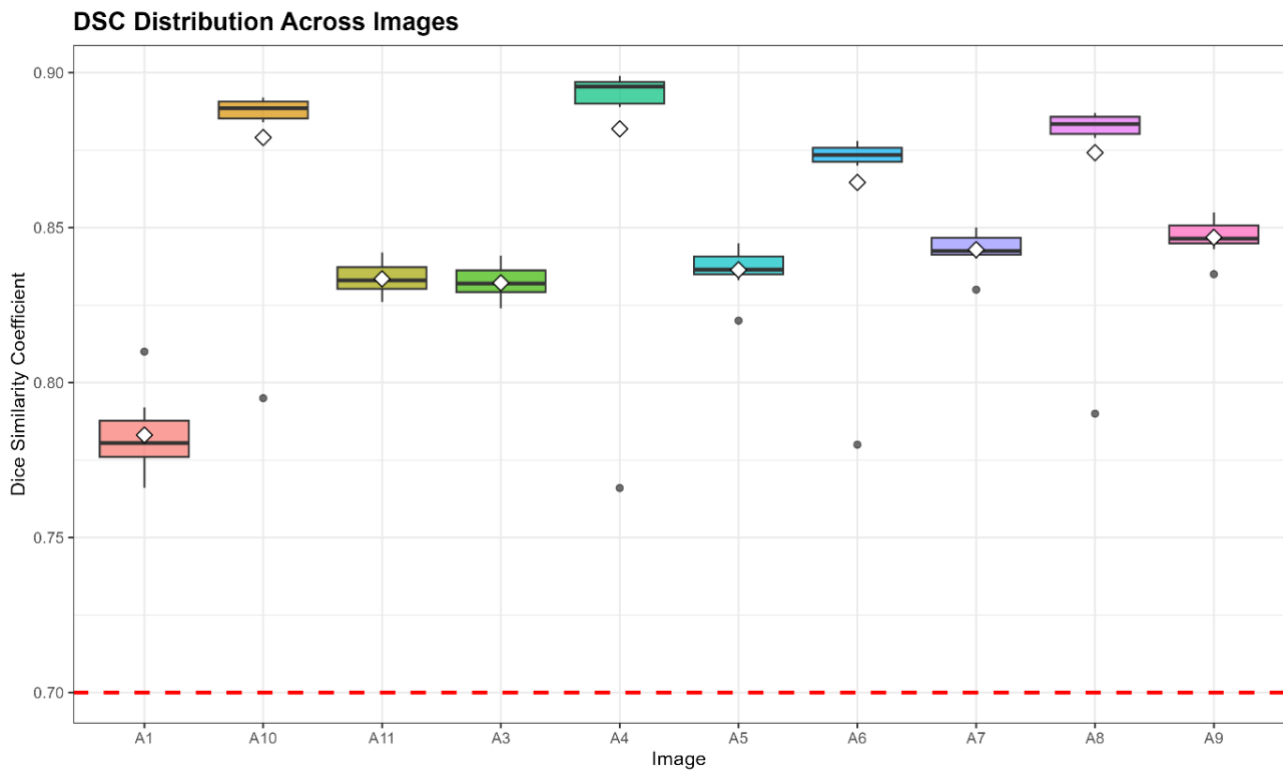
**Table 4.1: Descriptive Statistics of DSC (Dice Similarity Coefficient) values.**

**OVERALL PERFORMANCE METRICS (N=100 MEASUREMENTS)**

STATISTICAL PARAMETER	Value
MEAN $\pm$ SD	0.847 $\pm$ 0.035
MEDIAN (IQR)	0.843 (0.832 - 0.880)
RANGE	0.766 - 0.899
95% CI	0.841 - 0.854
COEFFICIENT OF VARIATION (%)	4.12%

From the table above, the mean DSC gotten was  $0.847 \pm 0.035$ . This indicates that there is a high level of overlap (85%) between manual and automated segmentation. A high level of overlap shows us that automated segmentation is highly accurate. All DSC values gotten ranged between 0.766 and 0.899 which are above the 0.7 threshold which is the accepted threshold for an accurate segmentation.

The results gotten are consistent across all samples, this is shown by the low coefficient of variation gotten (4.12%).

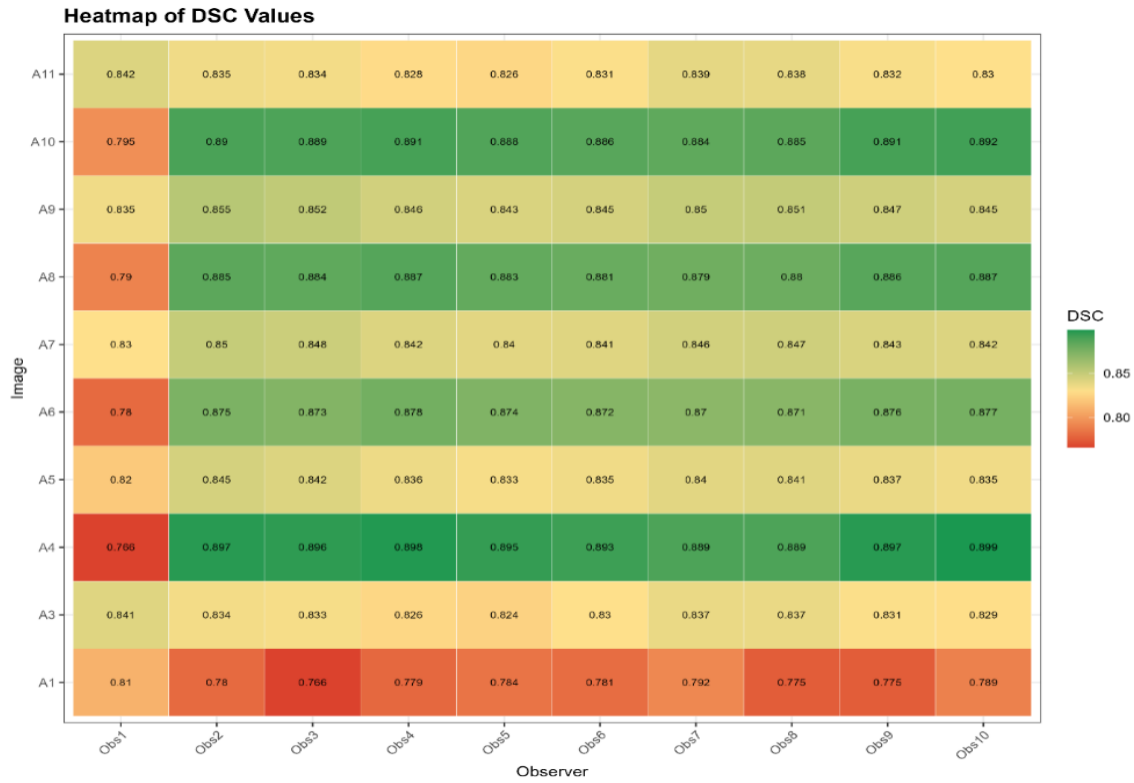


**Figure 4.3. Illustrates the overlap between manual and automated liver boundaries, confirming the accuracy of the DSC values presented.**

The boxplot shows a range of distribution of the DSC values with a line vertically passing through the box. The black dots sited above and below these vertical lines are representations of statistical outliers. The median value is shown by the horizontal black thick line in the box and the box is bounded above and below by the 75<sup>th</sup> and 25<sup>th</sup> percentile. The boxplot above shows us the distribution of the DSC with respect to each image segmented manually and automatically

**Table 4.12 DSC performance with respect to each image segmented.**

<i>Image ID</i>	<i>Mean DSC</i>	<i>Standard Deviation</i>	<i>Range</i>	<i>CV (%)</i>
<i>A1</i>	<b>0.783</b>	0.012	0.766 - 0.810	1.53%
<i>A3</i>	<b>0.832</b>	0.005	0.824 - 0.841	0.63%
<i>A4</i>	<b>0.882</b>	0.041	0.766 - 0.899	4.63%
<i>A5</i>	<b>0.836</b>	0.007	0.820 - 0.845	0.82%
<i>A6</i>	<b>0.865</b>	0.030	0.780 - 0.878	3.45%
<i>A7</i>	<b>0.843</b>	0.006	0.830 - 0.850	0.67%
<i>A8</i>	<b>0.874</b>	0.030	0.790 - 0.887	3.40%
<i>A9</i>	<b>0.847</b>	0.006	0.835 - 0.855	0.66%
<i>A10</i>	<b>0.879</b>	0.030	0.795 - 0.892	3.38%
<i>A11</i>	<b>0.833</b>	0.005	0.826 - 0.842	0.61%



**Figure 4.4. Heat map of DSC values, showing the most accurate and least accurate DSC values.**

For the 10 liver images segmented, the mean DSC for each image gotten ranged between 0.783 – 0.882. This shows that on each image segmented the automated segmentation algorithm used was very accurate putting into consideration the slight variation in segmentations done manually by different observers.

Figure 4.4 shows us a representation of the range of DSC values. The least DSC value is 0.766 and it is represented with a red color. The highest being 0.899 has a green representation. The colors in between ranges from orange to light green. The orange being closer to the minimum DSC value and light green closest to the highest DSC value.

**Table 4.13. Inter-Observer Variability/Reliability analysis**

<i>Reliability Measure</i>	<i>Value</i>	<i>Interpretation</i>
<i>ICC (Two-way, Consistency)</i>	0.720	Moderate
<i>Fleiss' Kappa</i>	0.683	Substantial
<i>ICC 95% CI Lower</i>	0.518	-
<i>ICC 95% CI Upper</i>	0.901	-

The Intraclass Correlation coefficient (ICC= 0.720) and Fleiss’ Kappa (0.683) shows that the agreement between all 10 observers were moderate to substantial. This confirms there was low variability in segmentations done by each observer. The ICC 95% CI lower value of 0.518 shows that at worst, there will still be a moderate agreement between all observers and an ICC 95% CI upper value of 0.901 shows us that at best, we will have an excellent agreement between all

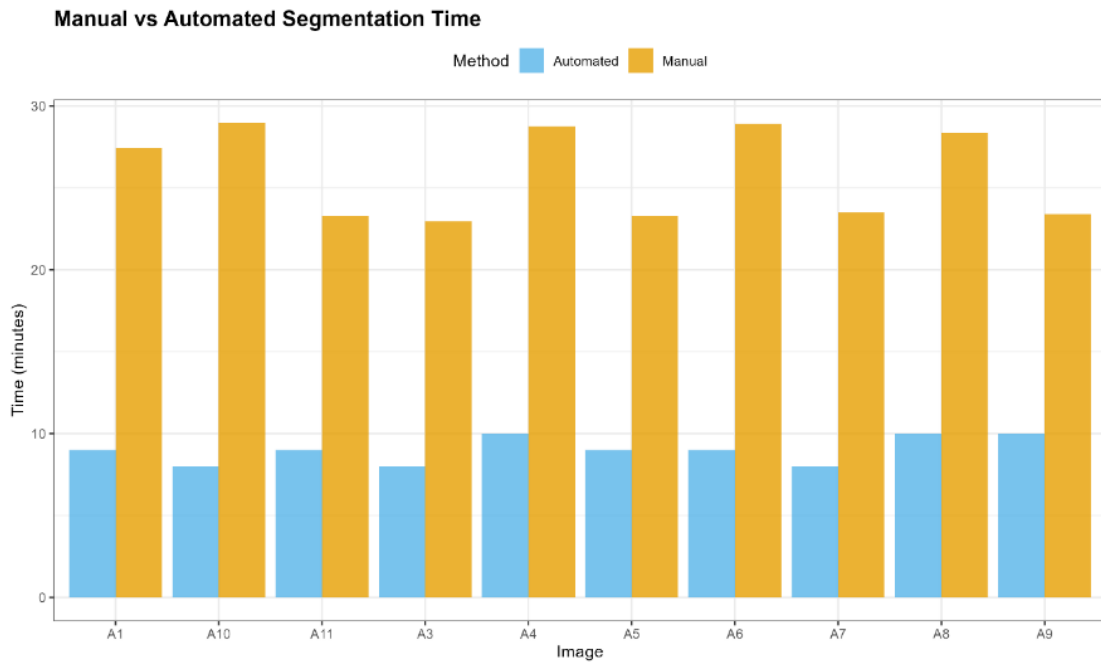
observers involved in the manual segmentation of the liver indicating that manual segmentation is reliable.

**Table 4.14. Segmentation Time Comparison.**

*Manual vs Automated segmentation efficiency (minutes)*

<b><i>Image ID</i></b>	<b>Manual (min)</b>	<b>Automated (min)</b>	<b>Time Saved (min)</b>	<b>% Time Saved</b>
<i>A1</i>	27.44	9	18.44	67.2%
<i>A3</i>	22.96	8	14.96	65.2%
<i>A4</i>	28.77	10	18.77	65.2%
<i>A5</i>	23.31	9	14.31	61.4%
<i>A6</i>	28.93	9	19.93	68.9%
<i>A7</i>	23.51	8	15.51	66.0%
<i>A8</i>	28.39	10	18.39	64.8%
<i>A9</i>	23.41	10	13.41	57.3%
<i>A10</i>	28.96	8	20.96	72.4%
<i>A11</i>	23.30	9	14.30	61.4%
<b><i>Overall Mean</i></b>	<b>25.90</b>	<b>9.00</b>	<b>16.90</b>	<b>65.2%</b>

An average of 25.9 minutes was used to manually segment each CT image, whereas it took automated segmentation an average of 9 minutes saving about 16.9 minutes which is a 65% reduction on the time spent on each image. This indicates that automated segmentation is way more efficient when compared to manual segmentation.



**Figure 4.5. A bar chart comparing manual and automated segmentation time.**

Figure 4.5 is a bar chart representing the comparison between the time efficiency of manual and automated segmentation. The yellow bars represent the time used to manually segment the liver while the bars in blue represent the time it took the algorithm to automatically segment the liver.

The value gotten from the one sample T-test ( $p < 0.001$ ), shows us that the DSC values gotten were significantly greater than 0.7. This confirms the high accuracy of automated segmentation.

The value gotten from the Paired T-test ( $p < 0.001$ ), shows that there is a significant difference between manual and automated segmentation time. This confirms the speed of automated segmentation.

**Table 4.15. Statistical test summary.**

*Hypothesis testing results*

<i>Statistical Test</i>	Test Statistic	p-value	95% CI	Result
<i>One-sample t-test (DSC vs 0.7 threshold)</i>	t(99) = 42.223	< 0.001	0.842 to Inf	Significantly > 0.7
<i>Paired t-test (Manual vs Automated time)</i>	t(9) = 19.866	< 0.001	14.97 to 18.82	Significant difference
<i>Pearson Correlation (Time vs DSC)</i>	r(8) = 0.455	0.186	-0.244 to 0.843	No significant correlation

The Pearson correlation value gotten ( $r=0.455$ ,  $p=0.186$ ), indicated that a longer manual segmentation time did not reflect on the accuracy of segmentation (no improvement in accuracy due to longer segmentation time).

**4.2 Discussion.**

In this study, we looked into evaluating the accuracy and time of automated segmentation relative to manual segmentation. Also, we assessed the inter-observer variability between 10 different observers who manually segmented the liver using 3D-slicer.

To determine the accuracy of automated segmentation, each segmentation mask generated by the automated algorithm was compared with the manual masks created by the observers. The mean DSC of  $0.85 \pm 0.03$  indicates an 85% overlap between the automated segmentation and the ground truth, demonstrating that the algorithm is highly accurate.

This result directly addresses the first research objective which was to evaluate the accuracy of automated segmentation relative to manual segmentation. It also answers the first research question regarding how closely automated liver segmentation matches manual (ground truth) segmentation. The high DSC value suggests that the automated method can reliably replicate expert manual segmentation, supporting its potential use in clinical workflows where speed and reproducibility are important.

The DSC gotten in this study indicates that there is a good agreement between both types of segmentations and when compared to the findings of a study carried out by Xie et al., 2023, which showed DSC values  $> 0.93 \pm 0.01$  for different structures segmented, there was a slight deviation. This deviation surpasses the measurement uncertainty, suggesting that between both studies there is a significant difference between automated segmentation accuracy. This difference may be as a result of dataset composition differences, quality of image and model training parameters. Although there is a slight deviation in DSC value, the results of this study reflect a satisfactory segmentation accuracy which is suitable for clinical evaluation.

In terms of inter-observer variability, the intra-class correlation (ICC) was 0.72 and Fleiss' Kappa was 0.68, indicating minimal variability between observers. These values demonstrate low interobserver variability in manual liver segmentation. This result directly addresses the second research objective which was to assess the inter-observer variability in liver segmentation among multiple observers. It also answers the corresponding research question regarding the extent of

variability between manual segmentations. The moderate-to-high agreement suggests that manual segmentation is relatively consistent among trained observers, providing a reliable baseline against which the accuracy of automated segmentation can be evaluated.

A study carried out by Kiefer et al., 2018 reported a low inter-observer variability with ICC values of 0.99 (PDF) and 0.98 (CSA) respectively. Compared to our study the variability between observers is slightly lower and this can be caused by anatomical complexity of the structure segmented and size of dataset used. Nonetheless, the result obtained from this study tells us that there is a moderate to substantial agreement between observers indicating a low inter-observer variability.

Automated segmentation was significantly faster than manual segmentation, achieving a 65% reduction in segmentation time. On average, it took 9 minutes per image for the automated algorithm, compared to 25.9 minutes per image for manual segmentation by each observer. This finding directly addresses the third research objective which was to measure and compare the time of liver segmentation using automated and manual methods. It also answers the corresponding research question regarding how much time automated segmentation saves relative to manual segmentation. The substantial decrease in time highlights the efficiency advantage of automated segmentation, suggesting its practical utility in clinical settings where time and resource optimization are critical.

With regards to segmentation time, the results gotten deviates from a study done by Cayot et al., 2021. The study showed a manual and automatic segmentation time of  $22.4 \pm 10.4$  min and  $5 \pm 2.1$

secs respectively. This deviation indicates that our system was relatively slower which can be caused by a difference in hardware configuration, experience difference of observers and algorithm differences. Despite this, the results of this study still show how fast and efficient automated segmentation is.

## CHAPTER FIVE

### CONCLUSION AND RECOMMENDATION

#### 5.1 Conclusion

The aim of this study was to compare manual and automated segmentation algorithms in CT imaging of the liver, with the goal of determining which approach provides greater efficiency, accuracy, and reliability. The results indicate that while manual segmentation remains a reliable reference standard, it is limited by inter-observer variability and long segmentation times. The use of Intra-Class Correlation (ICC) and Fleiss' Kappa highlighted the subjective nature of manual segmentation and confirmed differences between observers.

Automated segmentation, on the other hand, achieved a high Dice Similarity Coefficient (DSC), demonstrating strong overlap and similarity with the manual ground truth. Additionally, automated segmentation was highly efficient, reducing segmentation time by 65%.

These findings not only confirm that automated segmentation can achieve accuracy comparable to the gold standard, but also demonstrate its potential to streamline workflow and enhance clinical efficiency. In the context of radiology, this suggests that integrating automated segmentation algorithms into routine clinical practice could improve diagnostic speed, reduce observer workload, and support more consistent assessments across patients.

Furthermore, this study provides a foundation for future research, highlighting the need to evaluate automated segmentation across different algorithms, larger and more diverse datasets, and observers with varying levels of expertise. Such studies could further validate automated methods, optimize their clinical application, and potentially expand their use to other organs and imaging modalities.

## 5.2 Recommendations

- **Wide Clinical adoption:** From the results of this study, it is safe to recommend that automated segmentation algorithms such as TotalSegmentations be integrated into clinical practice to boost efficiency and clinical workflow.
- **Training of radiographers:** More radiographers should be trained in the manual segmentation of anatomical structures. They should also be trained in using and supervising automated segmentation algorithms.
- **Further Research:** Larger datasets and additional anatomical structures should be included in future studies to generalize findings on how accurate and efficient automated segmentation algorithms are.
- **Educational Inclusion:** Radiography students should be trained in using automated algorithms in segmentation, this will serve as a preparation for evolving digital practices.

## 5.3 Limitations of the study

- **Limited Observer pool:** Segmentations done manually was performed by 10 observers from almost the same training background. This might not really capture the variability

seen among individuals with different expertise levels. Future studies should include radiographers, radiologists, and radiotherapists with diverse levels of experience and training to enhance the generalizability of the results.

- **Single Segmentation Algorithm:** TotalSegmentator was the only automated segmentation algorithm used. Other algorithms might have performed differently. In future studies, multiple automated liver segmentation algorithms, such as nnU-Net, MONAI, or others, should be evaluated alongside manual segmentation to assess if performance is consistent across different platforms, addressing the limitation of using a single algorithm.

#### 5.4 Suggestions for further studies.

Based on the above limitations, here are some important suggestions to take into consideration when performing a deeper research:

- **Expand observer diversity:** Radiographers, radiologists, and radiotherapists with varying levels of experience and training backgrounds should be included in further studies to improve result generalizability.
- **Increase sample size:** In future studies, more observers should be involved to capture a wider range of inter-observer variability and improve the generalizability of results.
- **Evaluate multiple algorithms:** In future studies, the performance of several automated liver segmentation algorithms e.g., nnU-Net, MONAI, or other algorithms should be used

when comparing manual and automated segmentation to determine if results are consistent across platforms. This solves the single segmentation algorithm limitation.

## References

- Abel, L., Wasserthal, J., Meyer, M. T., Vosschenrich, J., Yang, S., Donners, R., Obmann, M., Boll, D., Merkle, E., Breit, H., & Segeroth, M. (2024). Intra-Individual reproducibility of Automated Abdominal Organ Segmentation—Performance of TotalSegmentator compared to human readers and an independent NNU-Net model. *Deleted Journal*. <https://doi.org/10.1007/s10278-024-01265-w>
- Araújo, J. D. L., Da Cruz, L. B., Diniz, J. O. B., Ferreira, J. L., Silva, A. C., De Paiva, A. C., & Gattass, M. (2021). Liver segmentation from computed tomography images using cascade deep learning. *Computers in Biology and Medicine*, *140*, 105095. <https://doi.org/10.1016/j.compbimed.2021.105095>
- Bazira, P. J. (2023). Anatomy of the liver. *Surgery (Oxford)*, *41*(6), 313–318. <https://doi.org/10.1016/j.mpsur.2023.02.024>
- Cayot, B., Milot, L., Nempont, O., Vlachomitrou, A. S., Langlois-Jacques, C., Dumortier, J., Boillot, O., Arnaud, K., Barten, T. R. M., Drenth, J. P. H., & Valette, P. (2022). Polycystic liver: automatic segmentation using deep learning on CT is faster and as accurate compared to manual segmentation. *European Radiology*, *32*(7), 4780–4790. <https://doi.org/10.1007/s00330-022-08549-1>
- Çelik, H., Odaman, H., Altay, C., Ünek, T., Özbilgin, M., Egeli, T., Ağalar, C., Astarçioğlu, İ. K., & Barlık, F. (2023). Manual and semi-automated computed tomography volumetry significantly overestimates the right liver lobe graft weight: a

singlecenter study with adult living liver donors. *Diagnostic and Interventional Radiology*, 30(1), 3–8. <https://doi.org/10.4274/dir.2023.221903>

Delmoral, J. C., & Tavares, J. M. R. (2024). Semantic Segmentation of CT liver Structures: A systematic review of recent trends and bibliometric analysis. *Journal of Medical Systems*, 48(1). <https://doi.org/10.1007/s10916-024-02115-6>

Dhanasekaran, R., Bandoh, S., & Roberts, L. R. (2016). Molecular pathogenesis of hepatocellular carcinoma and impact of therapeutic advances. *F1000Research*, 5, 879. <https://doi.org/10.12688/f1000research.6946.1>

Dionísio, F., Oliveira, L., Hernandes, M., Engel, E., Rangayyan, R., Azevedo-Marques, P., & Nogueira-Barbosa, M. (2020). Manual and semiautomatic segmentation of bone sarcomas on MRI have high similarity. *Brazilian Journal of Medical and Biological Research*, 53(2). <https://doi.org/10.1590/1414-431x20198962>

Dreizin, D., Zhou, Y., Zhang, Y., Tirada, N., & Yuille, A. L. (2019). Performance of a deep learning algorithm for automated segmentation and quantification of traumatic pelvic hematomas on CT. *Journal of Digital Imaging*, 33(1), 243–251. <https://doi.org/10.1007/s10278-019-00207-1>

Fel, J., Ellis, C., & Turk-Browne, N. (2023). Automated and manual segmentation of the hippocampus in human infants. *Developmental Cognitive Neuroscience*, 60, 101203. <https://doi.org/10.1016/j.dcn.2023.101203>

- Germain, T., Favelier, S., Cercueil, J., Denys, A., Krausé, D., & Guiu, B. (2014). Liver segmentation: Practical tips. *Diagnostic and Interventional Imaging*, 95(11), 1003–1016. <https://doi.org/10.1016/j.diii.2013.11.004>
- Han, X., Wu, X., Wang, S., Xu, L., Xu, H., Zheng, D., Yu, N., Hong, Y., Yu, Z., Yang, D., & Yang, Z. (2022). Automated segmentation of liver segment on portal venous phase MR images using a 3D convolutional neural network. *Insights Into Imaging*, 13(1), 26. <https://doi.org/10.1186/s13244-022-01163-1>
- He, B., Yin, D., Chen, X., Luo, H., Xiao, D., He, M., Wang, G., Fang, C., Liu, L., & Jia, F. (2021). A study of generalization and compatibility performance of 3D U-Net segmentation on multiple heterogeneous liver CT datasets. *BMC Medical Imaging*, 21(1). <https://doi.org/10.1186/s12880-021-00708-y>
- Kacała, A., Dorochowicz, M., Matus, I., Puła, M., Korbecki, A., Sobański, M., Jacków-Nowicka, J., Patrzalek, D., Janczak, D., & Guziński, M. (2024). Hepatic Hemangioma: Review of Imaging and Therapeutic Strategies. *Medicina*, 60(3), 449. <https://doi.org/10.3390/medicina60030449>
- Kavur, A. E., Selver, M. A., Dicle, O., Barış, M., & Gezer, N. S. (2019). CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge data [Dataset]. In *Zenodo (CERN European Organization for Nuclear Research)*. <https://doi.org/10.5281/zenodo.3431873>
- Kiefer, L. S., Fabian, J., Lorbeer, R., Machann, J., Storz, C., Kraus, M. S., Wintermeyer, E., Schlett, C., Roemer, F., Nikolaou, K., Peters, A., & Bamberg, F. (2018). Inter-

and intra-observer variability of an anatomical landmark-based, manual segmentation method by MRI for the assessment of skeletal muscle fat content and area in subjects from the general population. *British Journal of Radiology*, 20180019. <https://doi.org/10.1259/bjr.20180019>

Kumaraswamy, A. K., & Patil, C. M. (2021). Automatic prostate segmentation of magnetic resonance imaging using Res-Net. *Magnetic Resonance Materials in Physics Biology and Medicine*, 35(4), 621–630. <https://doi.org/10.1007/s10334-021-00979-0>

Lee, J. W., Andersen, S. a. W., Hittle, B., Powell, K. A., Al-Fartoussi, H., Banks, L., Brannen, Z., Lahchich, M., & Wiet, G. J. (2024a). Variability in manual segmentation of temporal bone structures in cone beam CT images. *Otology & Neurotology*, 45(3), e137–e141. <https://doi.org/10.1097/mao.0000000000004119>

Lee, J. W., Andersen, S. a. W., Hittle, B., Powell, K. A., Al-Fartoussi, H., Banks, L., Brannen, Z., Lahchich, M., & Wiet, G. J. (2024b). Variability in manual segmentation of temporal bone structures in cone beam CT images. *Otology & Neurotology*, 45(3), e137–e141. <https://doi.org/10.1097/mao.0000000000004119>

Lenchik, L., Heacock, L., Weaver, A. A., Boutin, R. D., Cook, T. S., Itri, J., Filippi, C. G., Gullapalli, R. P., Lee, J., Zagurovskaya, M., Retson, T., Godwin, K., Nicholson, J., & Narayana, P. A. (2019). Automated Segmentation of tissues using CT and MRI: A Systematic review. *Academic Radiology*, 26(12), 1695–1706. <https://doi.org/10.1016/j.acra.2019.07.006>

- Li, Y., Jia, Y., Li, S., Wang, W., Wang, Z., Wang, Y., Liu, B., Wang, W., Chang, H., & Li, Z. (2015). Transarterial Chemoembolization of Giant Liver Haemangioma: A Multi-center Study with 836 Cases. *Cell Biochemistry and Biophysics*, 73(2), 469–472. <https://doi.org/10.1007/s12013-015-0680-y>
- Li, S., Li, X. G., Zhou, F., Zhang, Y., Bie, Z., Cheng, L., Peng, J., & Li, B. (2024). Automated segmentation of liver and hepatic vessels on portal venous phase computed tomography images using a deep learning algorithm. *Journal of applied clinical medical physics*, 25(8), e14397. <https://doi.org/10.1002/acm2.14397>
- Mahadevan, V. (2014). Anatomy of the liver. *Surgery (Oxford)*, 38(8), 427–431. <https://doi.org/10.1016/j.mpsur.2014.10.004>
- Milara, E., Gómez-Grande, A., Sarandeses, P., Seiffert, A. P., Gómez, E. J., & Sánchez-González, P. (2024). Automatic skeleton segmentation in CT images based on U-Net. *Deleted Journal*, 37(5), 2390–2400. <https://doi.org/10.1007/s10278-024-01127-5>
- Moghbel, M., Mashohor, S., Mahmud, R., & Saripan, M. I. B. (2017). Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artificial Intelligence Review*, 50(4), 497–537. <https://doi.org/10.1007/s10462-017-9550-x>
- Mohajan, H. K. (2025, February 27). *Anatomy of Human liver: A theoretical study*. <https://www.paradigmpress.org/jimr/article/view/1528>

- Oh, N., Kim, J., Rhu, J., Jeong, W. K., Choi, G., Kim, J. M., & Joh, J. (2024). Comprehensive deep learning-based assessment of living liver donor CT angiography: from vascular segmentation to volumetric analysis. *International Journal of Surgery*. <https://doi.org/10.1097/js9.0000000000001829>
- Okada, M., Aoki, R., Nakazawa, Y., Tago, K., & Numata, K. (2024). CT and MR Imaging of Hepatocellular Carcinoma and Liver Cirrhosis. *Gastroenterology Insights*, *15*(4), 976–997. <https://doi.org/10.3390/gastroent15040068>
- Olthoff, K. M., Emond, J. C., Shearon, T. H., Everson, G., Baker, T. B., Fisher, R. A., Freise, C. E., Gillespie, B. W., & Everhart, J. E. (2015). Liver regeneration after living donor transplantation: Adult-to-adult living donor liver transplantation cohort study. *Liver Transplantation*, *21*(1), 79–88. <https://doi.org/10.1002/lt.23966>
- Rastogi, A. (2022). Liver transplant biopsy interpretation: Diagnostic considerations and conundrums. *Indian Journal of Pathology and Microbiology*, *65*(2), 245. [https://doi.org/10.4103/ijpm.ijpm\\_1090\\_21](https://doi.org/10.4103/ijpm.ijpm_1090_21)
- Raudaschl, P. F., Zaffino, P., Sharp, G. C., Spadea, M. F., Chen, A., Dawant, B. M., Albrecht, T., Gass, T., Langguth, C., Lüthi, M., Jung, F., Knapp, O., Wesarg, S., Mannion-Haworth, R., Bowes, M., Ashman, A., Guillard, G., Brett, A., Vincent, G., . . . Fritscher, K. D. (2017). Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. *Medical Physics*, *44*(5), 2020–2036. <https://doi.org/10.1002/mp.12197>

- Sadick, M., Müller-Wille, R., Wildgruber, M., & Wohlgemuth, W. (2018). Vascular Anomalies (Part I): Classification and Diagnostics of Vascular Anomalies. *RöFo - Fortschritte Auf Dem Gebiet Der Röntgenstrahlen Und Der Bildgebenden Verfahren*, *190*(09), 825–835. <https://doi.org/10.1055/a-0620-8925>
- Shen, L. (2022). Implementation of CT image segmentation based on an image segmentation algorithm. *Applied Bionics and Biomechanics*, *2022*, 1–11. <https://doi.org/10.1155/2022/2047537>
- Sumadewi, K. T. (2023). Embryology, anatomy and physiology of the liver: Review. *Indian Journal of Clinical Anatomy and Physiology*, *10*(3), 138–144. <https://doi.org/10.18231/j.ijcap.2023.031>
- Van Sluis, J., Noordzij, W., De Vries, E. G. E., Kok, I. C., De Groot, D. J. A., Jalving, M., Hooge, M. N. L., Brouwers, A. H., & Boellaard, R. (2022). Manual versus Artificial Intelligence-Based Segmentations as a pre-processing step in whole-body PET dosimetry calculations. *Molecular Imaging and Biology*, *25*(2), 435–441. <https://doi.org/10.1007/s11307-022-01775-5>
- Vinayahalingam, S., Berends, B., Baan, F., Moin, D. A., Van Luijn, R., Bergé, S., & Xi, T. (2023). Deep learning for automated segmentation of the temporomandibular joint. *Journal of Dentistry*, *132*, 104475. <https://doi.org/10.1016/j.jdent.2023.104475>

- Wang, H., Minnema, J., Batenburg, K., Forouzanfar, T., Hu, F., & Wu, G. (2021). Multiclass CBCT Image Segmentation for Orthodontics with Deep Learning. *Journal of Dental Research*, *100*(9), 943–949. <https://doi.org/10.1177/00220345211005338>
- Wasserthal, J., Breit, H., Meyer, M. T., Pradella, M., Hinck, D., Sauter, A. W., Heye, T., Boll, D. T., Cyriac, J., Yang, S., Bach, M., & Segeroth, M. (2023). TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. *Radiology Artificial Intelligence*, *5*(5). <https://doi.org/10.1148/ryai.230024>
- Xie, T., Li, Y., Lin, Z., Liu, X., Zhang, X., Zhang, Y., Zhang, D., Cheng, G., & Wang, X. (2023). Deep learning for fully automated segmentation and volumetry of Couinaud liver segments and future liver remnants shown with CT before major hepatectomy: a validation study of a predictive model. *Quantitative Imaging in Medicine and Surgery*, *13*(5), 3088–3103. <https://doi.org/10.21037/qims-22-1008>
- Yan, X., Lin, B., Fu, J., Li, S., Wang, H., Fan, W., Fan, Y., Feng, M., Wang, R., Fan, J., Qi, S., & Jiang, C. (2023). Deep-learning-based automatic segmentation and classification for craniopharyngiomas. *Frontiers in Oncology*, *13*. <https://doi.org/10.3389/fonc.2023.1048841>
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116–1128. <https://doi.org/10.1016/j.neuroimage.2006.01.015>

Zareei, A., & Karimi, A. (2016). Liver segmentation with new supervised method to create initial curve for active contour. *Computers in Biology and Medicine*, 75, 139–150.

<https://doi.org/10.1016/j.combiomed.2016.05.009>

**APPENDIX I: Data Collection Sheet.**

<b>OBSERVERS</b>	<b>SEGMENTATION TIME</b>									
<b>1</b>										
<b>2</b>										
<b>3</b>										
<b>4</b>										
<b>5</b>										
<b>6</b>										
<b>7</b>										
<b>8</b>										
<b>9</b>										
<b>10</b>										

This sheet is used to record the time taken for each manual segmentation task during the study.

Consistent timing tools e.g. a stopwatch is used for accurate measurement of the time taken to segment these structures.

**APPENDIX II: Ethical Approval.**

**HEALTH RESEARCH ETHICS COMMITTEE (HREC)**  
**UNIVERSITY OF BENIN TEACHING HOSPITAL**  
P.M.B. 1111 BENIN CITY NIGERIA Telephone: 052-600418 Website: ubth.org

**CHIEF MEDICAL DIRECTOR**  
Prof. Darlington E. Obaseki  
E-mail: darlobaseki@gmail.com

**DIRECTOR OF ADMINISTRATION**  
Jim Uwadie, Esq

**CHAIRMAN**  
Prof. (Mrs.) Antoinette N. Ofili

**HREC OFFICE:**  
Committee email: ubthresearchethics@gmail.com  
Registration Number:  
NHREC-UBTH-HREC/24/12/2022B

PROTOCOL NUMBER: ADM/E 22/A/VOL.VII/2025/138

PROPOSAL TITLE: "COMPARISON BETWEEN MANUAL AND AUTOMATED SEGMENTATION ALGORITHMS IN COMPUTED TOMOGRAPHY IMAGING OF THE LIVER"

PRINCIPAL INVESTIGATOR(S): IYEMIFOKHAE CALEB EWOGBAI

DEPARTMENT/INSTITUTION: DEPARTMENT OF RADIOGRAPHY, SCHOOL OF BASIC MEDICAL SCIENCES UNIVERSITY OF BENIN, BENIN CITY, EDO STATE

DATE CONSIDERED: AUGUST 6<sup>TH</sup>, 2025

DECISION OF THE COMMITTEE: APPROVED

*THIS APPROVAL DATES 6/8/2025 TO 5/8/2026. IF THERE IS DELAY IN STARTING THE RESEARCH, PLEASE INFORM THE HREC SO THAT THE DATES OF APPROVAL CAN BE ADJUSTED ACCORDINGLY*

REMARK:

CHAIRMAN: PROF. (MRS) A.N. OFILI SIGNATURE & DATE: *A.N. Ofili, 6/8/2025*

SUPERVISOR (S): AKPOBASAHAN E.A. (DR)

DECLARATION BY INVESTIGATOR(S):  
PROTOCOL NUMBER (please quote in all enquiries)  
Note that no participant accrual or activity related to this research may be conducted outside of these dates. All informed consent forms used in this study must carry the HREC assigned number and duration of HREC approval of the study. In multiyear research, endeavor to submit your annual re-port to the HREC early in order to obtain renewal of your approval and avoid disruption of your research. No changes are permitted in the research without prior approval by the HREC except in circumstances outlined in the Code. The HREC reserves the right to conduct compliance visit your research site without previous notification

Signature & Date: *[Signature]* 6/8/25

ubthresearchethics@gmail.com Registration Number: NHREC/24/01/202

### APPENDIX III: Python code used to calculate the DSC.

```
import nibabel as nib
import numpy as np

def dice_coefficient(mask1, mask2):
    mask1 = mask1.astype(bool)
    mask2 = mask2.astype(bool)

    intersection = np.logical_and(mask1, mask2).sum()
    return 2. * intersection / (mask1.sum() + mask2.sum())

# Load the segmentations
manual_nii = nib.load('manual_seg.nii')
auto_nii = nib.load('auto_seg.nii')

manual_mask = manual_nii.get_fdata()
auto_mask = auto_nii.get_fdata()

# Optional: binarize if values are not strictly 0 or 1
manual_mask = manual_mask > 0
auto_mask = auto_mask > 0
```

```
# Compute DSC
```

```
dsc = dice_coefficient(manual_mask, auto_mask)
```

```
print(f'Dice Similarity Coefficient: {dsc:.4f}')
```

This python code is used to calculate the DSC value which tells us how accurate the automated segmentation algorithm used is.