



Instituto Superior de Engenharia

Politécnico de Coimbra

DEPARTMENT OF INFORMATICS AND COMPUTER
ENGINEERING

2D Liver Segmentation in Abdominal Computed Tomography Scans

Dissertation to fulfill the Master of Science degree in Informatics
Engineering

Specialization in Software Engineering

Author

Stephanie Alexandra Batista Niño

Supervisor

Inês Campos Monteiro Sabino Domingues

Co-Supervisors

Jorge Miguel Tavares Couceiro de Sousa

Ricardo Ângelo Santos Filipe

Advisors at the Coimbra Hospital University Center

Paulo Jorge Botelho Moraes Rachinhas

Jorge Miguel Sousa Isidoro

Coimbra, September 2024



POLYTECHNIC INSTITUTE
OF COIMBRA

COIMBRA INSTITUTE OF
ENGINEERING

RESUMO

A detecção precoce de anomalias no fígado é crucial para um diagnóstico preciso e eficaz de doenças como o cancro do fígado, uma das principais causas de morte por cancro no mundo. Modalidades de imagiologia médica, especialmente a tomografia computadorizada, têm desempenhado um papel fundamental na identificação e caracterização de doenças. Nos Hospitais da Universidade de Coimbra, a segmentação precisa do fígado constitui uma técnica muito importante para o tratamento das doenças hepáticas. No entanto, as ferramentas comerciais de segmentação automática existentes apresentam um custo elevado pela segmentação em cada tomografia realizada e as soluções gratuitas carecem de uma interface para a visualização dos resultados. Assim, este projeto introduz uma abordagem designada por Incrementalização de Dados Privados, avaliando a capacidade de generalização dos modelos de segmentação de imagem U-Net, ResUNet++, FCN e BerDiff. Esta abordagem apresenta uma análise da exposição incremental dos modelos a vários dados clínicos do mundo real, aumentando a sua adaptabilidade e avaliando o seu desempenho em conjuntos de dados com dimensões e fontes diversificadas. Através de experiências sistemáticas, observou-se que o modelo mais eficaz para segmentar o fígado foi o modelo ResUNet++, apresentando o melhor desempenho, com uma exatidão de 0,9972, um DSC de 0,9449 e um ASSD de 0,0053 mm. Seguidamente, este modelo foi integrado numa interface gráfica desenvolvida no âmbito do projeto para a segmentação automática do fígado. Esta contribuição pretende otimizar o diagnóstico e monitorização de doenças, reduzindo os custos associados, melhorando o atendimento ao paciente e a alocação de recursos em ambientes clínicos.

Palavras-chave: Inteligência Artificial, Redes Neurais Artificiais, U-Net, ResUNet++, FCN, BerDiff, Tomografia Computorizada, Segmentação Automática do Fígado, Incrementalização de Dados Privados.

ABSTRACT

Early detection of liver anomalies is crucial for accurate and effective diagnosis of diseases such as liver cancer, one of the leading causes of cancer death worldwide. Medical imaging modalities, especially computed tomography, have played a fundamental role in the identification and characterization of diseases. At the Coimbra University Hospital, precise liver segmentation is an important technique for the treatment of liver diseases. However, existing commercial automatic segmentation tools have a high cost for each tomography and free tools lack a user-friendly interface for result visualization. Therefore, this project introduces an approach referred to as Private Data Incrementalization, evaluating the generalization capability of the U-Net, ResUNet++, FCN and BerDiff image segmentation models. This approach presents an analysis of the incremental exposure of the models to various real-world clinical data, increasing their adaptability and evaluating their performance on datasets with varying dimensions and sources. Through systematic experiments, it was observed that the most effective model for segmenting the liver was the ResUNet++ model, demonstrating the best performance, with an accuracy of 0.9972, a DSC of 0.9449 and an ASSD of 0.0053 mm. This model was subsequently integrated into a tool developed within the scope of the project for automatic liver segmentation. This contribution aims to optimize the diagnosis and monitoring of diseases, reducing associated costs, improving patient care and resources allocation in clinical environments.

Keywords: Artificial Intelligence, Artificial Neural Networks, U-Net, ResUNet++, FCN, BerDiff, Computed Tomography, Automatic Liver Segmentation, Private Data Incrementalization.

ACKNOWLEDGMENTS

I would like to thank all the people who have been essential to my continuous academic and professional growth throughout this journey.

I am grateful for the support I have received from the Radiotherapy and Nuclear Medicine departments of the Coimbra University Hospitals, especially from Eng. Paulo Rachinhas and Eng. Jorge Isidoro, as well as all the hospital workers who made it possible to collaborate on this project. I appreciate the availability, the trust, the exchange of knowledge and the opportunity.

The professors and coordinators of the master's program deserve a special acknowledgment for their commitment to sharing knowledge and for their encouragement and support.

My project supervisors, Professors Inês Domingues and Ricardo Filipe, have been crucial in the success of this project. Your expertise, patience, and insightful guidance challenged me throughout the process of developing the project to go further.

I also thank the employees and professors of the several departments at Coimbra Institute of Engineering (ISEC), whose friendliness and hospitality made my days brighter, with a special shoutout to Professor Miguel Couceiro, who more than a supervisor, after seven years of guiding me and supporting me unconditionally, I can say that he is a good friend and I am very grateful for having always accompanied me at every phase of my academic career, always betting on me and believing in me.

To my boyfriend, thank you for being my rock, encouraging me, and motivating me throughout this journey. His steady presence and love have been fundamental enabling me to face the challenges and overcome the obstacles throughout this path.

I am particularly grateful to my family, who stood by me unconditionally throughout my master's journey. I would not have made it this far without their support, comprehension, and love. They are my strength and the reason why I strive to achieve my goals.

Finally, my sincere thanks to everyone who played a part, directly or indirectly, in this work. Their involvement has been fundamental to this achievement, and I am thrilled to have shared this journey with them.

TABLE OF CONTENTS

Resumo	i
Abstract.....	iii
Acknowledgments	v
Table of Contents.....	vii
Index of Figures.....	xiii
Index of Tables	xviii
List of Acronyms	xxi
List of Symbols	xxv
1 Introduction.....	1
1.1 Motivation and Problem Statement.....	1
1.2 Main Objectives.....	2
1.3 Main Contributions.....	3
1.4 Project Planning	5
1.5 Document Structure	6
2 Background Knowledge.....	7
2.1 Medical Imaging.....	7
2.1.1 Radiology	7
2.1.2 Computed Tomography	7
2.1.3 Digital Imaging and Communications in Medicine	10
2.1.4 Picture Archiving and Communication Systems	11
2.1.5 Medical Imaging Segmentation.....	11
2.2 Liver.....	13
2.2.1 Definition	13
2.2.2 Importance	14
2.2.3 Importance of Liver Studies	14
2.2.4 Relevance of Liver Segmentation.....	14
2.3 Artificial Intelligence in Healthcare	15
2.3.1 Definition	15
2.3.2 Convolutional Neural Networks	16
2.3.3 Ground Truth	18

2.3.4	Application in Medical Image Segmentation	19
3	Review of Image Segmentation Techniques and Evaluation Methods	21
3.1	Image Segmentation Techniques	21
3.2	Performance Evaluation Metrics.....	21
3.2.1	Region-based Metrics.....	24
3.2.2	Contour-based Metrics	28
3.2.3	Findings	31
3.3	Literature Review	33
3.3.1	Research Methodology	33
3.3.2	Public Datasets	34
3.3.3	Historical Overview	35
3.3.4	Findings	35
3.4	Commercial Automated Segmentation Software Solutions.....	43
3.4.1	Annotate by Therapanacea.....	43
3.4.2	Contour+ by MVision AI	43
3.4.3	Limbus Contour by Limbus AI	44
3.4.4	Comparison.....	44
3.5	Public DICOM Viewers.....	45
3.6	Findings	47
3.6.1	Adoption of Artificial Neural Network-based Methods	47
3.6.2	Dimensionality Impact for Liver Segmentation	47
3.6.3	Main Limitations.....	48
3.6.4	Research Questions.....	49
4	Private Dataset Creation	51
4.1	Submission to the Hospital's Ethics Committee	51
4.2	Overview	51
4.3	ARIA Oncology Information System.....	52
4.4	Labeling Protocol.....	52
4.5	Structured Query Language Statement.....	52
4.6	Initial Scans Visualization.....	53
4.7	Outcome of the Initial Scans Visualization	54
4.8	Exporting Abdominal Scans and Ground Truth	54

2D Liver Segmentation in Abdominal Computed Tomography Scans

4.9 Scans Anonymization Process.....	55
5 Methodology.....	57
5.1 Overview	57
5.1.1 Clinical Workflow.....	58
5.1.2 Artificial Intelligence Workflow	58
5.2 Experimental Setup.....	59
5.3 Data Preparation	59
5.3.1 Datasets	59
5.3.2 Relevant Concepts.....	64
5.3.3 Sort Slices	65
5.3.4 Masks Creation	65
5.4 Pre-processing	66
5.4.1 Hounsfield Windowing.....	67
5.4.2 Histogram Equalization.....	67
5.4.3 Data Normalization.....	68
5.5 Segmentation Models	68
5.5.1 U-Net	69
5.5.2 ResUNet++	70
5.5.3 FCN	71
5.5.4 BerDiff.....	72
5.6 Post-processing.....	73
5.6.1 Median Filtering.....	73
5.6.2 Thresholding.....	73
5.7 Experimental Specifications.....	74
5.8 Models Configuration and Hyperparameters	75
5.8.1 U-Net	75
5.8.2 ResUNet++	76
5.8.3 FCN	76
5.8.4 BerDiff.....	76
5.9 Evaluation	77
6 Results and Discussion.....	79
6.1 Quantitative Results.....	79

6.1.1 Models.....	79
6.1.2 Datasets	81
6.2 Qualitative Results.....	83
6.2.1 Models.....	83
6.2.2 Datasets	88
6.3 Discussion	96
6.3.1 Models Training Time	96
6.3.2 Models Segmentation Performance	98
6.3.3 Datasets Analysis	100
7 L+ver: Segmentation Application.....	103
7.1 Requirements Specification.....	103
7.1.1 Functional Requirements.....	103
7.1.2 Non-functional Requirements	104
7.2 Development	104
7.2.1 Environment.....	105
7.2.2 Application Build Process	105
7.2.3 Automation and Requirements.....	105
7.3 Main Features	106
7.3.1 Menu Files	106
7.3.2 Menu Export.....	108
7.3.3 Menu Zoom	108
7.3.4 Menu Segmentation	109
7.3.5 Menu Transformations	109
7.3.6 Menu Light Theme.....	110
7.3.7 Additional Features	111
8 Conclusions.....	115
8.1 Strengths.....	115
8.2 Limitations	116
8.3 Future Work.....	116
References	119
Appendices	133
Appendix A – Scans Exportation Strategy	135

2D Liver Segmentation in Abdominal Computed Tomography Scans

Appendix B – SQL Query for Scans Retrieval	139
Appendix C – Script for Scans Anonymization.....	141
Appendix D – RECPAD 2023 Publication	149
Appendix E – MDPI Sensors Journal Publication	153
Appendix F – MDPI Encyclopedia Journal Entry.....	173
Appendix G – Biomedical Engineering Conference	181
Appendix H – International Conference on Mathematical Analysis and Applications in Science and Engineering.....	185
Appendix I – Springer Proceedings Journal	195
Appendix J – RECPAD 2024 Publication	215
Appendix K – Bachelor of Science in Biomedical Engineering Presentation.	217
Appendix L – Liver Segmentation Application Packages	225
Appendix M – Training and Validation Results	227
Appendix N – Data Augmentation Techniques.....	235

INDEX OF FIGURES

Figure 1. Abdominal CT scan. Source: [29].	8
Figure 2. CT operating principle. (a) CT scanner constitution. (b) Acquisition of slices at various angles for image reconstruction. Source: [32].	8
Figure 3. Example of a CT scan in DICOM format. (a) CT scan image/slice. (b) Anonimized CT scan metadata.	10
Figure 4. Anatomy of the liver. Source: [43].	13
Figure 5. Relationship between AI, ML, DL and CV.	15
Figure 6. Illustration of the general architecture of CNNs. Source: [56].	17
Figure 7. Illustration of image segmentation results. (a) Original image. (b) Ground truth segmentation. (c) Automatic segmentation. (d) Classification outcomes.	21
Figure 8. Distribution of the metrics used in the selected studies.	24
Figure 9. Region-based illustration.	24
Figure 10. Illustration of the calculation of accuracy, with TP in green, TN in blue, FN in red, and FP in orange.	25
Figure 11. Illustration of the calculation of precision, with TP in green, and FP in orange.	26
Figure 12. Illustration of the calculation of DSC, with TP in green, FN in red, and FP in orange.	27
Figure 13. Illustration of the calculation of VOE, with TP in green, TN in blue, FN in red, and FP in orange.	28
Figure 14. Contour-based illustration.	28
Figure 15. Illustration of the calculation of ASSD, with TP in green, TN in blue, FN in red, and FP in orange.	29
Figure 16. Illustration of the calculation of MSSD, with TP in green, TN in blue, FN in red, and FP in orange.	30
Figure 17. Common segmentation challenges. (a) Under-segmentation. (b) Over-segmentation.	31
Figure 18. Distribution of the retrieved studies over the years.	33
Figure 19. Characterization of the selected studies.	34
Figure 20. Selected studies distribution according to the categories of image segmentation techniques.	34
Figure 21. Summary of the quantitative results presented in the reviewed documents.	36

Figure 22. Percentage of studies according to their level of automation (semi-automatic, fully automatic or both).....	37
Figure 23. Percentage of studies according to their dimensionality of the image (2D versus 3D).....	42
Figure 24. Percentage of dataset types.....	42
Figure 25. Number of studies using each public dataset.	43
Figure 26. Illustration of the position of the gall bladder and liver. Source: [129]..	49
Figure 27. Flowchart of the operating sequence for the dataset creation.	51
Figure 28. Methodology for the proposed solution.	57
Figure 29. Age distribution of the created dataset.....	61
Figure 30. Gender distribution of the created dataset.	61
Figure 31. Scan year distribution of the created dataset.	61
Figure 32. Example of a CT scan from the 3D-IRCADb01 dataset. (a) CT scan image/slice. (b) Liver segmentation ground truth.....	62
Figure 33. Age distribution of the 3D-IRCADb01 public dataset.	62
Figure 34. Gender distribution of the 3D-IRCADb01 public dataset.....	63
Figure 35. Example of a CT scan from the CHAOS dataset. (a) CT scan image/slice. (b) Liver segmentation ground truth.....	63
Figure 36. Gender distribution of the CHAOS public dataset.	64
Figure 37. Masks creation process.....	66
Figure 38. Example of mask creation. (a) Original image before pre-processing techniques. (b) Ground truth mask created.	66
Figure 39. Pre-processing flowchart.....	67
Figure 40. Hounsfield window application. (a) Original image. (b) Image with Hounsfield windowing application.....	67
Figure 41. Hounsfield windowing and histogram equalization application. (a) Original image. (b) Outcome of applying the Hounsfield window followed by histogram equalization.	68
Figure 42. Segmentation flowchart.....	68
Figure 43. U-Net model architecture. Source: [119].	70
Figure 44. ResUNet++ model architecture. Source: [136].	71
Figure 45. FCN model architecture. Source: [137].....	72
Figure 46. BerDiff model architecture. Source: [138].....	72
Figure 47. Post-processing flowchart.....	73

2D Liver Segmentation in Abdominal Computed Tomography Scans

Figure 48. Post-processing comparison for two images. (a) Ground truth mask. (b) Predicted segmentation. (c) Post-processed segmentation.	73
Figure 49. RuntimeError: CUDA out of memory description.	76
Figure 50. Performance of the U-Net experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E1. (d) Segmentation outcome with E2. (e) Segmentation outcome with E3. (f) Segmentation outcome with E4. (g) Segmentation outcome with E5.	84
Figure 51. Performance of the ResUNet++ experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E6. (d) Segmentation outcome with E7. (e) Segmentation outcome with E8. (f) Segmentation outcome with E9. (g) Segmentation outcome with E10.	85
Figure 52. Performance of the FCN experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E11. (d) Segmentation outcome with E12. (e) Segmentation outcome with E13. (f) Segmentation outcome with E14. (g) Segmentation outcome with E15.	86
Figure 53. Performance of the BerDiff experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E16. (d) Segmentation outcome with E17. (e) Segmentation outcome with E18. (f) Segmentation outcome with E19. (g) Segmentation outcome with E20.	87
Figure 54. Loss values for each model across all datasets on unseen data.	88
Figure 55. Accuracy values for each model across all datasets on unseen data.	88
Figure 56. Precision values for each model across all datasets on unseen data.	88
Figure 57. DSC values for each model across all datasets on unseen data.	89
Figure 58. AOE values for each model across all datasets on unseen data.	89
Figure 59. ASSD values for each model across all datasets on unseen data.	89
Figure 60. MSSD values for each model across all datasets on unseen data.	90
Figure 61. Comparison of the performance from each model for the dataset 1 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.	91
Figure 62. Comparison of the performance from each model for the dataset 2 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.	92
Figure 63. Comparison of the performance from each model for the dataset 3 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.	93

Figure 64. Comparison of the performance from each model for the dataset 4 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.....	94
Figure 65. Comparison of the performance from each model for the dataset 5 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.....	95
Figure 66. Application architecture diagram.	105
Figure 67. Menu bar of the application main functionalities.	106
Figure 68. Single CT scan image (file) uploaded.....	106
Figure 69. CT scan folder uploaded.....	107
Figure 70. Multiple CT scan series visualization.....	107
Figure 71. DICOM View frame: (a) Original image visualization. (b) Image removed from DICOM View frame.	108
Figure 72. Export image options.....	108
Figure 73. Zoom feature: (a) 100% zoom. (b) 200% zoom.....	108
Figure 74. Segmentation feature: (a) Original image visualization. (b) Liver segmentation visualization.....	109
Figure 75. DICOM image transformation: (a) Negative operation. (b) Flip horizontal. (c) 90 CCW rotation. (d) Flip vertical. (e) 90 CW rotation.....	110
Figure 76. Examples of light screen theme.	111
Figure 77. Some of the application layout options.....	112
Figure 78. DICOM metadata.	112
Figure 79. HU modifications: (a) Original image. (b) Lower values of HU. (c) Higher values of HU. (d) Smaller range of values of HU.....	113
Figure 80. HU application: (a) ROI visualization based on HU values. (b) HU scale for different structures and tissues. Source: [137].	113
Figure 81. Rotation operation. (a) Original image. (b) Original mask. (c) Original image with mask overlay. (d) Rotated image. (e) Rotated mask. (f) Rotated image with mask overlay.	235
Figure 82. Reflection operation. (a) Original image. (b) Original mask. (c) Original image with mask overlay. (d) Reflected image. (e) Reflected mask. (f) Reflected image with mask overlay.	236
Figure 83. Combination of rotation and reflection operations. (a) Original image. (b) Original mask. (c) Original image with mask overlay. (d) Rotated and reflected image. (e) Rotated and reflected mask. (f) Rotated and reflected image with mask overlay.	236

INDEX OF TABLES

Table 1.1. Project schedule.....	6
Table 2.1. Advantages and disadvantages of using CNNs.	18
Table 3.1. Comparison of segmentation categories.	22
Table 3.2. Summary of segmentation methods and main performance metrics used in the selected studies.....	23
Table 3.3. Summary of the main evaluation metrics used by automatic liver segmentation studies.	32
Table 3.4. Essential information of each public dataset used by the studies analyzed.	35
Table 3.5. Summary of reviewed studies.	38
Table 3.6. Comparison between the different commercial software solutions considered by HUC.....	44
Table 3.7. Comparison between public DICOM viewers. Source for the comparison structure: [128].	46
Table 3.8. Usability assessment comparison provided by [128].	46
Table 3.9. Comparison between 2D and 3D dimensionality.	47
Table 4.1 Structure of the table created for the initial visualization of all the images.	53
Table 4.2. Result of the scans visualization.	54
Table 5.1. Comparison between the datasets used.....	59
Table 5.2. Datasets structure based on number of CT scans.....	59
Table 5.3. Datasets structure based on number of CT scans.....	60
Table 5.4. Demographic profile of the patients from the created dataset.....	60
Table 5.5. Configuration of the performed experiments.....	74
Table 6.1. Performance of the U-Net experiments on unseen data.	79
Table 6.2. Performance of the ResUNet++ experiments on unseen data.....	80
Table 6.3. Performance of the FCN experiments on unseen data.....	80
Table 6.4. Performance of the BerDiff experiments on unseen data.	81
Table 6.5. Comparison of the performance from each model on unseen data from the dataset 1.....	81
Table 6.6. Comparison of the performance from each model on unseen data from the dataset 2.....	82

2D Liver Segmentation in Abdominal Computed Tomography Scans

Table 6.7. Comparison of the performance from each model on unseen data from the dataset 3.....	82
Table 6.8. Comparison of the performance from each model on unseen data from the dataset 4.....	82
Table 6.9. Comparison of the performance from each model on unseen data from the dataset 5.....	83
Table 6.10. Training and validation time for each experiment.	97
Table 7.1. Functional requirements.	103
Table 7.2. Non-functional requirements.	104
Table 7.3. Components within the application.	105
Table 8.1. Performance of the U-Net experiments on training dataset.....	227
Table 8.2. Performance of the U-Net experiments on validation dataset.....	227
Table 8.3. Performance of the ResUNet++ experiments on training dataset.....	228
Table 8.4. Performance of the ResUNet++ experiments on validation dataset..	228
Table 8.5. Performance of the FCN experiments on training dataset.	229
Table 8.6. Performance of the FCN experiments on validation dataset.	229
Table 8.7. Performance of the BerDiff experiments on training dataset.	230
Table 8.8. Performance of the BerDiff experiments on validation dataset.....	230
Table 8.9. Comparison of the performance from each model on training dataset 1.	231
Table 8.10. Comparison of the performance from each model on validation dataset 1.	231
Table 8.11. Comparison of the performance from each model on training dataset 2.	231
Table 8.12. Comparison of the performance from each model on validation dataset 2.	232
Table 8.13. Comparison of the performance from each model on training dataset 3.	232
Table 8.14. Comparison of the performance from each model on validation dataset 3.	232
Table 8.15. Comparison of the performance from each model on training dataset 4.	233
Table 8.16. Comparison of the performance from each model on validation dataset 4.	233

Table 8.17. Comparison of the performance from each model on training dataset 5.
.....233

Table 8.18. Comparison of the performance from each model on validation dataset
5.234

LIST OF ACRONYMS

Acronym	Meaning
AAM	Axial Attention Module
ABC	Artificial Bee Colony
AC-Net	Attention Connect Network
ACM Digital Library	Association for Computing Machinery Digital Library
Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AOE	Area Overlap Error
ASPP	Atrous Spatial Pyramid Pooling
ASSD	Average Symmetric Surface Distance
CAD	Computer-Assisted Diagnosis
CFCN	Cascaded Fully CNN
CHAOS	Combined Healthy Abdominal Organ Segmentation
HUC	<i>Hospitais da Universidade de Coimbra</i> / Coimbra University Hospitals
CNNs	Convolutional Neural Networks
CRF	Conditional Random Fields
CT	Computed Tomography
CV	Computer Vision
DGS	<i>Direção-Geral da Saúde Portuguesa</i> / Portuguese Directorate General of Health
DICOM	Digital Imaging and Communications in Medicine

Acronym	Meaning
DL	Deep Learning
DRLSE	Distance Regularized Level Set Evolution
DSC	Dice Similarity Coefficient
DSN	Deeply Supervised Network
ECC	Enhanced Correlation Coefficient
FCM	Fuzzy C-Means
FCN	Fully Convolutional Network
FN	False Negatives
FP	False Positives
GA	Genetic Algorithm
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
GUI	Graphical User Interface
GVF	Gradient Vector Flow
HCC	Hepatocellular Carcinoma
HD	Hausdorff Distance
HIS	Hospital Information System
HU	Hounsfield Units
ICMASC	International Conference on Mathematical Analysis and Applications in Science and Engineering
IHE	Integrating the Healthcare Enterprise
II	Individually Identifiable Information
ILSVRC	ImageNet Large Scale Visual Recognition Challenge

2D Liver Segmentation in Abdominal Computed Tomography Scans

Acronym	Meaning
IMRT	Intensity-Modulated Radiation Therapy
IS	Instance Number
ISEC	<i>Instituto Superior de Engenharia de Coimbra</i> / Coimbra Institute of Engineering
JC	Jaccard Coefficient
JEB	<i>Jornadas de Engenharia Biomédica</i> / Biomedical Engineering Conference
MDPI	Multidisciplinary Digital Publishing Institute
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MSSD	Maximum Symmetric Surface Distance
MSSQL	Microsoft SQL Server
NEMA	National Electrical Manufacturers Association
PACS	Picture Archiving and Communication Systems
Q	Question
RAVD	Relative Absolute Volume Difference
RIS	Radiology Information System
RMS	Root Mean Square
RMSD	Root Mean Square Symmetric Surface Distance
RMSPop	Root Mean Square Propagation
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RSST	Radiotherapy Structure Set
RVD	Relative Volume Difference

Acronym	Meaning
SI	Similarity Index
SOP	Service-Object Pair
SQL	Structured Query Language
std	Standard Deviation
SVM	Support Vector Machines
TN	True Negatives
TP	True Positives
UID	Unique Identifier
ViT	Vision Transformers
VOE	Volumetric Overlap Error
VTM	Vision Transformer Module
WHO	World Health Organization

LIST OF SYMBOLS

Symbol	Meaning
2D	Two-dimensional
3D	Three-dimensional
$d(G_j, S)$	Shortest distance from the point G_j on the ground truth surface to any point on the segmented surface S
$d(G_j, S_i)$	Denotes the Euclidean distance between a point G_j on the ground truth surface to a point S_i on the segmented surface
$d(S_i, G)$	Shortest distance from the point S_i on the segmented surface to any point on the ground truth surface G
$d(S_i, G_j)$	Denotes the Euclidean distance between a point S_i on the segmented surface and a point G_j on the ground truth surface
G	Ground truth
G_j	Denotes a point on the ground truth liver surface
K	Constant that defines the contrast factor (1000)
M	Total number of points on the ground truth liver surface
N	Total number of points on the segmented liver surface
r	Correlation Coefficient
S	Segmented liver region
S_i	Denotes a point on the segmented liver surface
μ	Linear attenuation coefficient
μ_{tissue}	Represents the linear attenuation coefficient of the tissue
μ_{water}	Linear attenuation coefficient of water

1 INTRODUCTION

This chapter presents the context and motivation, as well as the main objectives and contributions of the project. Additionally, an overview will be provided to ease navigation through the dissertation.

1.1 Motivation and Problem Statement

Medical imaging has been playing a fundamental role in the diagnosis and treatment of diseases in recent years, through the use of advanced scanners and image reconstruction software. Physicians require tools that allow them to localize organs and tissues with higher precision and rapidity, while identifying and characterizing pathologies for accurate diagnosis. As one of the most important organs in the digestive system, the liver performs critical functions such as breaking down nutrients, producing bile and eliminating toxic substances. However, liver-related diseases, particularly oncological diseases, pose health risks, and liver cancer is a leading cause of cancer-related mortality worldwide [1]-[3]. With advances in technology, the invasive techniques used for diagnosis are being gradually replaced by non-invasive medical imaging techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI), with benefits for the patient. These techniques have become an integral part of the diagnosis, treatment planning and monitoring the progress of oncological diseases [4][5], providing detailed cross-sectional images for accurate visualization of internal structures, including liver tumors [6][7].

With the advancement of technology and artificial intelligence (AI) in medicine, there is a growing need to optimize the identification of oncological diseases [8]. Medical image segmentation is emerging as a fundamental step in the pipeline [9]-[12]. Liver segmentation in CT scans has emerged as a critical area, requiring accurate identification and delimitation of the liver region for treatment planning and progress monitoring, as well as for early detection of liver lesions and metastases to other organs [13]-[15]. However, accurate liver segmentation on CT scans is challenging due to factors such as artefacts, varying soft tissue densities, the complexity of adjacent organ proximity and dynamic anatomical modifications, such as tumor contraction during treatment, there is a compelling need for reliable and fast segmentation to facilitate adaptive radiotherapy [16]-[20]. Within this context, the development of a method for 2D automatic liver segmentation in CT scans is intended based on liver scans acquired at the Department of Radiotherapy of the Coimbra University Hospitals (HUC). The results will be compared with a reference segmentation performed by expert physicians, yielding comparative metrics with other state of the art methods.

1.2 Main Objectives

The main purpose of this project is to implement and compare algorithms to perform the automatic segmentation of the liver, which is of particular importance when carrying out treatments, either through external radiation in the Radiotherapy Department or through the application of HOLMIO-166 or Yttrium-90 seeds in the Nuclear Medicine Department.

An additional objective is the implementation of a Graphical User Interface (GUI) tool. This tool would allow healthcare professionals to avoid the costs of using a commercial software, with each segmented scan potentially costing about 40 euros, depending on the software. The eventual goal is to provide a more sustainable and cost-effective solution for the liver segmentation process in CT scans.

Through this project, it is intended to address the following questions (Q):

- Q1 — What are the challenges and limitations associated with accurate liver segmentation in CT scans?
- Q2 — How does the choice of the method impact the accuracy and efficiency of liver segmentation in CT scans?
- Q3 — What are the evaluation metrics commonly used to assess the performance of AI models and traditional methods for liver segmentation in CT scans?
- Q4 — What can be done to contribute to optimizing the process of treatment planning for oncological diseases, especially for the liver?
- Q5 — What are the best specifications and project conditions according to the HUC's temporal needs and priorities?
- Q6 — Which is the most efficient automatic segmentation model for the case study within the chosen algorithms?
- Q7 — How does the data source (CT scans) impact on the performance of the segmentation models?

To answer the first three questions (Q1, Q2 and Q3), a literature review was conducted [21], summarizing relevant work for this project. Additionally, informal interviews and brainstorming sessions with some of the physicians from the Department of Radiotherapy of HUC were held to understand the importance and challenges of liver segmentation in CT scans and their respective needs and requirements in order to answer the two following questions (Q4 and Q5). The last two questions (Q6 and Q7) were addressed at the end of the project, through a thorough comparative analysis of the performance of the different models given the *Private Data Incrementalization* from different sources into the same configuration (same data preparation, pre-processing, hyperparameters definition and post-processing), varying the model to be trained.

Private Data Incrementalization consists of an innovative approach focused on improving the performance and adaptability of Machine Learning (ML) models in the field of image segmentation (easily extended to other types of data, such as tabular). This method gradually exposes models to diverse real-world medical data, evaluating their ability to adapt and generalize to new information.

The approach emphasizes the evaluation of the performance of the models while being trained on an increasing number of medical scans (clinical scenarios). By systematically introducing varied datasets from different sources, the purpose is to build more robust and versatile segmentation models.

These questions are designed to provide a better understanding of the context and scope of the project, so it can be implemented complying with the real needs of the hospital and understand what currently exists and its limitations. This understanding is crucial to implementing an effective solution that responds to both the existing challenges and the expectations of the HUC.

1.3 Main Contributions

The main contributions of the project are associated to providing the development of an algorithm for 2D liver segmentation in abdominal CT scans for the Department of Nuclear Medicine and the Department of Radiotherapy of the HUC.

In particular, the main outcomes of the project are:

- Definition of a guideline to be followed for creating a Radiotherapy Structure Set for specific structures segmented in a CT scan (see Appendix A).
- Creation of a script (written in Bash) for automatic scans anonymization when exporting from the HUC Radiotherapy department's ARIA network (from Varian Medical Systems), where the images are stored. The ARIA network server contains all the information on the patients treated there, particularly dose records, CT scans conducted, structures designed by the physicians, among other relevant information (see Appendix C).
- Creation of a dataset with CT scans including ground truth (chapter 4).
- Optimization and comparison of four algorithms for automatic liver segmentation (U-Net, ResUNet++, FCN and BerDiff) (sub-chapter 5.8).
- Generalization assessment of the segmentation algorithms with *Private Data Incrementalization* (chapter 6).
- Integration of the best segmentation model trained with a GUI with easy user interaction, designed to minimize user effort and reduce time spent on the segmentation process, providing a free tool to be used by HUC (chapter 7).

- Scientific contributions (almost 10 publications in different sources).

Furthermore, the following scientific contributions were achieved within the scope of the master's project:

- Poster presentation at *Portuguese Conference on Pattern Recognition (RECPAD 2023)* with the initial project proposal (**Artificial Intelligence-driven Approaches for Radiation Dose Prediction in Computed Tomography Scans**) [22] (available on Appendix D), which was modified due to a new proposal defined in collaboration with the HUC, consisting in the implementation of the current project.
- Publication in Multidisciplinary Digital Publishing Institute (MDPI) journals:
 - Publication of a paper in advance of peer-review (**Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective**) in *Preprints* platform supported by MDPI [23].
 - Publication of a paper (**Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective**) in *MDPI Sensors* journal, more specifically in the *Biomedical Sensing and Bioinformatics Processing Special Issue* [21], available on Appendix E.
 - Publication of an entry (**Algorithms for Liver Segmentation in Computed Tomography Scans**) in *MDPI Encyclopedia* journal [24], available on Appendix F.
- Poster presentation at the *Biomedical Engineering Conference 2024 (JEB'24)* held at ISEC (**Application of Artificial Intelligence in Medical Imaging**) [25], available on Appendix G.
- Oral presentation at the *International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMASSC'24)* (**Performance Evaluation Metrics for Automatic Liver Segmentation on Medical Imaging Scans**) [26], available on Appendix H.
- Submission of a paper (**Comparative Analysis of Quantitative Assessment Metrics for Liver Segmentation**) in *Springer Proceedings* journal, available on Appendix I.
- Paper submitted to the *Portuguese Conference on Pattern Recognition (RECPAD 2024)* (**Adoption of Artificial Neural Network-based Methods for Medical Image Segmentation**), available on Appendix J.

- Paper in progress (**Private Data Incrementalization Approach to Evaluate the Generalization of Liver Segmentation Models**) to be submitted in *MDPI Sensors* journal, more specifically in the *Pattern Recognition Applications in Medical Image Analysis Special Issue*.
- Presentation to the second-year students of the Bachelor of Science in Biomedical Engineering at ISEC within the scope of the Introduction to Clinical Databases course (**Acquisition, Collection and Preparation of Medical Data for Biomedical Engineering Application in the Hospital Environment**), available on Appendix K.

1.4 Project Planning

The main tasks implemented during the development of the project are presented below:

T1 – State of the art – literature review in the fields of medical imaging and AI, more specifically related to its application in liver segmentation in CT scans.

T2 – Analysis of existing public datasets and summary of their characteristics – study of public datasets to better understand the structure and characteristics they tend to present, considering the most relevant data, to give guidance for the creation of our dataset with the HUC scans.

T3 – Creation of a dataset from scratch – elaboration of all the documentation related to the process of submitting the proposal to the HUC Ethics Committee to formally apply for the sharing of medical data, selection, and filtering of adequate abdominal CT scans within the scope of the project, process of anonymization of patients' personal data and scans exportation.

T4 – Analysis of evaluation metrics – study of the liver segmentation evaluation metrics for the selection of the most appropriate ones for the context of the project.

T5 – Implementation and evaluation of segmentation models – search for existing code. Adaptation to the dataset. Performing experiments and adjusting parameters, as well as the evaluation and comparison of the developed models with the manual segmentation by HUC Radiotherapy Department's professionals.

T6 – Application development – implementation of the liver segmentation solution.

T7 – Dissertation – dissertation writing for public exams.

T8 – Scientific community knowledge sharing – scientific publications writing.

Table 1.1 shows the tasks implemented during the development of the project over time.

Table 1.1. Project schedule.

	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Tasks												
T1												
T2												
T3												
T4												
T5												
T6												
T7												
T8												

1.5 Document Structure

Besides introduction, this document is structured in 6 chapters:

Chapter 2 discusses the theoretical background and fundamental concepts related to medical imaging, liver importance, and AI application in healthcare, particularly for semi-automatic and fully automatic medical imaging segmentation.

Chapter 3 details the image segmentation techniques and the most commonly used performance evaluation metrics, the research methodology for the literature review, as well as the main related works that were a reference for defining the methodology of the project, providing inspiration for the innovation approach of this work, and compares this review with other existing summaries of the state of the art. Finally, it is presented a comparison of the commercial automated segmentation software solutions that have been considered by the Department of Radiotherapy of the HUC.

Chapter 4 explains the steps for creating the dataset used for the project.

Chapter 5 describes the methodology adopted in the project, covering aspects such as the experimental setup, including the data preparation, data pre-processing and post-processing, datasets used and the steps followed for the implementation of the experiments with the segmentation models.

Chapter 6 presents the results achieved through the project and a thorough analysis of them.

Chapter 7 introduces the GUI to be provided to the final users (physicians).

Finally, Chapter 8 discusses the conclusions of the project, emphasizing its contributions for improving the quality of health services for oncological pathologies at the HUC hospital, as well as future work directions.

2 BACKGROUND KNOWLEDGE

This chapter describes the fundamental aspects of medical imaging, liver and AI, particularly the application of Deep Learning (DL) in healthcare, with the aim of understanding the scientific area in which the work is contextualized. The principles of these areas will be covered, providing a solid basis for understanding the concepts and techniques used in the research and development of the project.

2.1 Medical Imaging

A brief overview of the principal concepts associated with medical imaging can be found below, providing some insights into what medical imaging is, its relevance, the main modalities or types of medical scans, the main limitations, and challenges in medical image processing.

2.1.1 Radiology

Medical imaging is an essential field for clinical practice, allowing the visualization and quantification of anatomical structures.

Radiology, also known as radiodiagnosis, is a medical specialization whose main purpose is to generate morphological images of the inside of the body (medical imaging) using various physical agents, so that these images can later be used for the diagnosis, prognosis, and treatment of pathologies [27].

2.1.2 Computed Tomography

Since CT scans will be used in this work, a brief explanation on how a CT scan is obtained follows.

A CT scan is a radiological examination based on a computerized X-ray imaging process, in which an X-ray beam is projected onto the patient and rotates around the body, producing slices that are reconstructed and then processed in a computer to obtain images of the inside of the body, which are digitally combined to form a 3D image of the patient, facilitating the identification and location of anatomical structures, as well as possible tumors (Figure 1) [28]. Compared with other X-ray modalities, CT involves much higher radiation exposure due to the reconstruction of the CT scan from the combination of numerous individual X-ray projections from different angles, but has the advantage of providing highly detailed images of internal structures, which enhances the reliability of diagnosis and treatment planning. This level of detail allows for better visualization of soft tissues, blood vessels, and bones, making CT scans particularly useful for detecting and assessing complex conditions such as tumors, internal injuries, and vascular diseases. Additionally, the ability

to create 3D images aids in surgical planning and intervention, contributing to improved patient outcomes.



Figure 1. Abdominal CT scan. Source: [29].

The CT scanner works by taking images from different angles using the gantry (Figure 2), which rotates 360° around the patient, reducing the exposure time of the patient, which is the minimum time needed to obtain quality images [28][30][31].

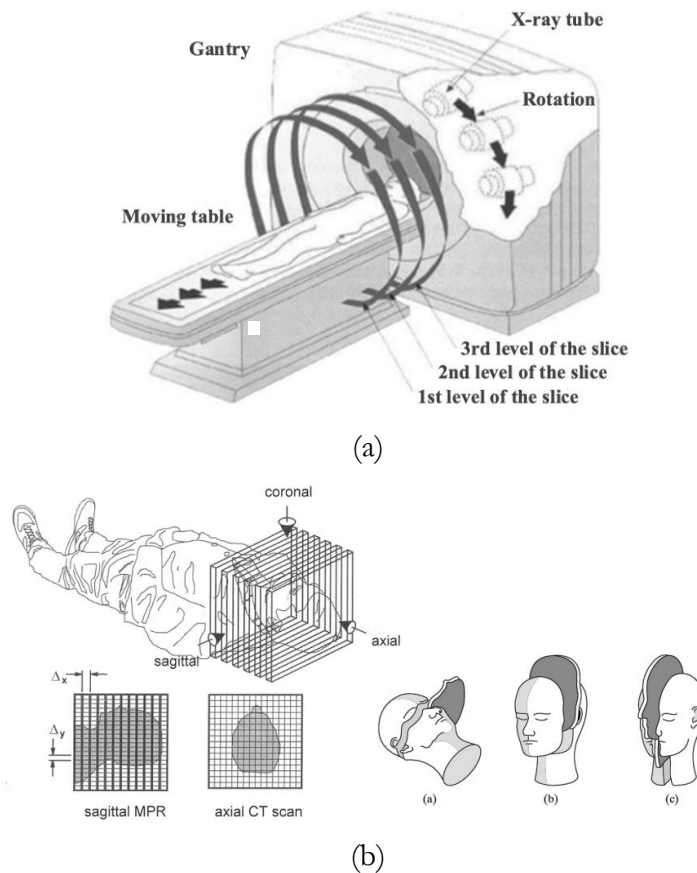


Figure 2. CT operating principle. (a) CT scanner constitution. (b) Acquisition of slices at various angles for image reconstruction. Source: [32].

2D Liver Segmentation in Abdominal Computed Tomography Scans

To produce X-rays, the filament of the ampoule's cathode is heated, causing the energy of the electrons in the filament material to increase, causing them to be released from the filament. In turn, through the potential difference between the cathode and anode, these electrons are accelerated and directed towards the anode target, acquiring kinetic energy in the process.

When these accelerated electrons collide with the anode target, they lose energy through braking by the nuclei of the anode material, producing braking X-rays, and by colliding with the orbital electrons of the anode material, causing them to be excited to higher energy levels, generating gaps in the lower energy levels, which are then filled by electrons of a higher energy level, producing characteristic X-rays.

The accelerating voltage of the X-ray tube plays a fundamental role in the production of X-rays, defining the energy spectrum of the photons produced.

For the absorption of photons by different tissues, it is important to consider the linear attenuation coefficient (μ) of the material passed through by the X-rays. This coefficient measures the material's ability to absorb photons and depends on the material's average atomic number and density, being inversely proportional to the photon's energy, specifically the lower the energy, the higher the linear attenuation coefficient. The importance of this parameter lies in its relationship with the image since it corresponds to a map proportional to the attenuation coefficients. In the case of materials with a low attenuation coefficient, scattered radiation can be generated, which degrades the contrast of the image, since the scattered photons have a different direction to the incident photons.

To produce X-rays, it is essential to consider beam hardening, which consists of an increase in the average energy of the spectra during their penetration into a given material, due to greater attenuation of low-energy photons. For this reason, a material may appear darker in the center of the image and lighter on the periphery.

To correct for this artifact caused by beam hardening, the Hounsfield Units (HU) is used, also known as the CT number, which consists of converting the value of the linear attenuation coefficient to a dimensionless scale, based on the density of water. The value for air is -1000 HU and the value for fat is -120 HU, compared to the value given for water (0 HU). The CT number is defined by Equation (1):

$$HU = \frac{\mu_{tissue} - \mu_{water}}{\mu_{water}} \times 1000 \quad (1)$$

where:

- **1000** is the constant that defines the contrast factor
- μ_{tissue} represents the linear attenuation coefficient of the tissue
- μ_{water} is the linear attenuation coefficient of water

The liver typically exhibits HU values between 40 and 60 HU under normal conditions. These values can vary depending on factors such as the presence of fat, fibrosis, or other pathologies [33].

2.1.3 Digital Imaging and Communications in Medicine

Digital Imaging and Communications in Medicine (DICOM) is a standard that facilitates the communication of biomedical information using digital images and associated data, as shown in Figure 3. Developed by the DICOM Standards Committee, which includes members from the National Electrical Manufacturers Association (NEMA), DICOM aims to establish and maintain national standards in this field [34]. This standard finds applications in various medical modalities such as radiology, cardiology, oncology, dentistry, surgery, neurology, breast imaging, radiotherapy, ophthalmology, pathology, pneumology, among others.

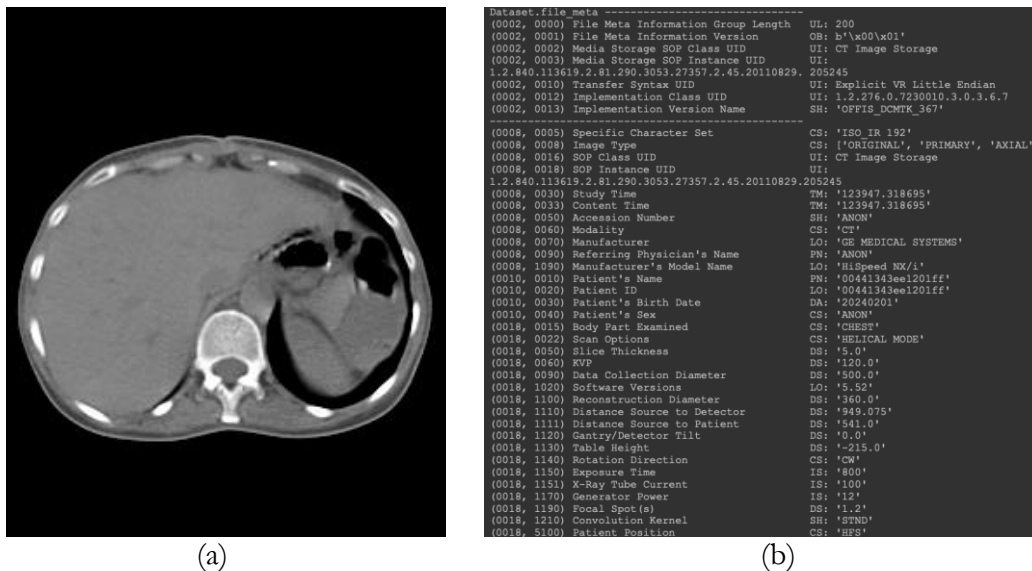


Figure 3. Example of a CT scan in DICOM format. (a) CT scan image/slice. (b) Anonimized CT scan metadata.

The DICOM Standard is built upon several fundamental principles [35]. Firstly, it emphasizes global applicability and localization, meaning that it is designed to be a worldwide standard that can be used in any locale. Secondly, the standard undergoes continuous maintenance in accordance with the procedures outlined by the DICOM Standards Committee. Additionally, each instance of information objects, such as images, is identified by a globally unique object identifier that remains consistent throughout all exchanges, ensuring traceability (e.g., SOP Instance UID, being one of the tags used in this project, the meaning of which is described in the sub-chapter 5.3.2.3 (SOP Instance UID)).

2.1.4 Picture Archiving and Communication Systems

Picture Archiving and Communication Systems (PACS) have revolutionized the storage and retrieval of medical imaging data. With the continuous advancement of imaging technology, PACS offers a cost-effective solution for storing vast amounts of medical images and enables quick retrieval from various modalities [36]. Moreover, PACS allows healthcare professionals to access these images from multiple locations, enhancing collaboration and improving patient care.

A PACS system comprises several components that work together seamlessly. These components include image acquisition devices, data management systems, image storage devices, transmission networks, display stations, and devices for producing hard-copy images. However, for optimal performance, PACS must also integrate with other management systems like the Hospital Information System (HIS) and the Radiology Information System (RIS) [37].

The HIS handles administrative aspects of hospitals, while the RIS manages medical imagery and associated data. To ensure data integrity and streamline information flow, initiatives like Integrating the Healthcare Enterprise (IHE) have been established. These initiatives aim to improve information sharing and transmission among computer systems in healthcare [38].

Standards such as DICOM play a crucial role in addressing specific clinical needs within PACS. This collaboration enhances the seamless flow of information between PACS, HIS, RIS, and other healthcare systems, ultimately improving patient care and outcomes.

2.1.5 Medical Imaging Segmentation

Medical imaging segmentation is the process of dividing digital images into multiple segments to simplify and modify the representation of the image into something that is easier to analyze. This technique is especially useful in medical diagnostics and treatment planning. It helps in identifying boundaries and structures within different imaging data. By segmenting specific areas or structures, such as organs or tumors, healthcare professionals can obtain precise measurements, assess the scope and severity of diseases, and customize treatments for individual patients. This segmentation process makes it easier to see complex tissues and helps in achieving a more accurate diagnosis and better patient outcomes.

The principal constraints and challenges associated with this process are detailed below [39]-[41]:

1. Acquisition artifacts

Some of the major constraints in segmentation arise from the artifacts inherent in medical imaging acquisition technologies.

These artifacts may involve:

- Partial volumes

The representation of an anatomical structure can be distorted due to overlapping adjacent tissues in the same pixel or voxel, which complicates the accurate identification of organ boundaries.

- Noise

Random fluctuations in the signal can cause a grainy appearance in the image, making it difficult to distinguish fine or low-contrast structures.

- Low sharpness

The limited resolution of some medical imaging equipments may result in blurred edges, where the transition between different tissues is not very clear.

2. Low contrast

Segmentation becomes challenged by the low contrast inherent in some medical scans. In certain instances, the intensity difference between adjacent tissues may be small, making it extremely difficult to discern the exact boundaries between distinct anatomical structures. This situation is especially problematic in organs areas where multiple structures overlap or are in close proximity.

3. Anatomical complexity and variability

Human anatomy presents complexity and variability, which implies that there are not two exactly equal organs among different individuals. This anatomical diversity introduces an additional level of complexity.

4. Time and resources involved

Manual segmentation is a time-consuming process that requires meticulous attention to detail. This process can require hours of work by highly skilled specialists, which limits the amount of analysis that can be performed in a given period. Such an investment of time and human resources could be more efficiently employed in diagnostic tasks and clinical decision making if automated or semi-automated segmentation tools were available.

5. Divergences in segmentation assessment

In the medical field, it is common for different professionals to have divergent opinions on the exact delineation of structures on images:

- Subjectivity in interpretation

Medical scans interpretation involves a degree of subjectivity, which can lead to discrepancies in the identification and delineation of anatomic structures.

- Experience and education

Differences in training, experience, and specialization can influence how clinicians approach the segmentation assignment, resulting in variations in the final results. These differences in interpretation can affect the consistency and reproducibility of segmentation results.

6. Inaccuracy and human error

Accuracy in manual segmentation of medical scans is inherently limited by susceptibility to human error. Despite the high skill and experience of healthcare professionals, factors such as fatigue, workload, and variability in approach can introduce inaccuracies:

- Absence of consistency

Fatigue and stress can affect consistency in manual segmentation, especially in prolonged sessions or with a high volume of images to be analyzed.

- Precision and accuracy

Human ability to discern and handle fine detail has limits, which can lead to the omission or misrepresentation of small structures or subtle features.

2.2 Liver

As this project focuses on its segmentation, this sub-chapter provides relevant information to better understand its main functions and relevance for study.

2.2.1 Definition

The liver is a vital organ located in the upper right quadrant of the abdomen, beneath the diaphragm (Figure 4). It is the largest internal organ and gland in the human body, weighing approximately 1.5 kilograms in an average adult. Structurally, the liver is divided into two main lobes: the right lobe, which is larger, and the left lobe. This organ is composed of specialized cells called hepatocytes, which perform a myriad of essential functions that are crucial for maintaining homeostasis and overall health [42].

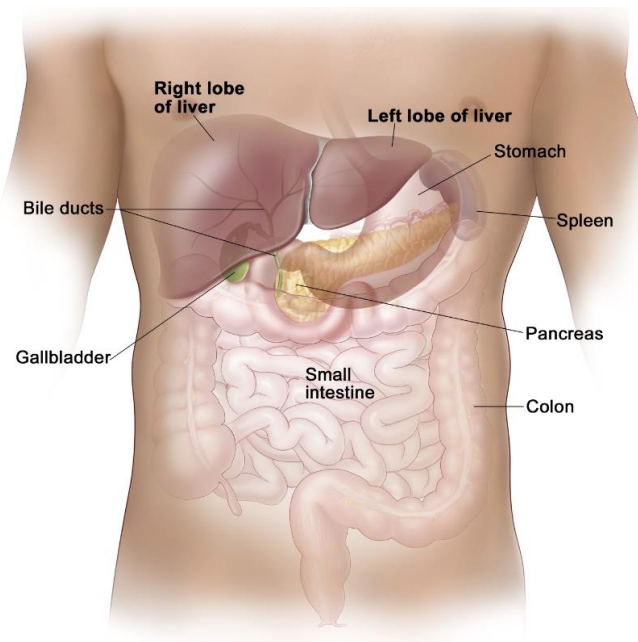


Figure 4. Anatomy of the liver. Source: [43].

2.2.2 Importance

The liver plays a critical role in several processes, making it essential for human survival. One of its primary functions is detoxification. The liver processes toxins and drugs, rendering them harmless or facilitating their excretion from the body. Additionally, it is responsible for the production of bile, which is crucial for the digestion and absorption of fats and fat-soluble vitamins in the small intestine [43][44].

The liver also performs an important function in metabolism. It regulates blood sugar levels by storing and releasing glucose as needed and is involved in the metabolism of proteins and lipids. Furthermore, the liver synthesizes essential proteins such as albumin and clotting factors, which are vital for maintaining blood volume and proper clotting function. The organ also stores vitamins and minerals, including iron and vitamin A, and converts ammonia, a toxic byproduct of protein metabolism, into urea for excretion.

2.2.3 Importance of Liver Studies

Studying the liver in detail is extremely important due to its critical functions and the crucial impact of liver diseases. Liver diseases, including cirrhosis, hepatitis, and liver cancer, are among the leading causes of death worldwide, being estimated that about 2500 people die each year due to liver diseases [45][46].

As Portugal is considered to be the 10th country in the world with the highest alcohol consumption and alcohol is one of the main causes of liver diseases, the Portuguese Society of Gastroenterology estimates a five-year mortality rate of 95% for people with liver diseases in Portugal, and liver cancer is the 5th leading cause of death in Portugal according to the Directorate General of Health (DGS) [47], mostly associated with men, approximately 75% [48].

These diseases not only affect millions of individuals, but also represent a substantial burden on health systems.

Understanding the complex physiology and pathology of the liver is essential for developing effective treatments and preventive measures. Advances in liver research have the potential to improve diagnostic techniques, therapeutic strategies, and patient outcomes. Additionally, studying the liver can provide insights into related conditions, such as metabolic disorders, cardiovascular diseases, and obesity, given the organ's integral role in metabolism and detoxification.

2.2.4 Relevance of Liver Segmentation

At HUC, liver segmentation in medical imaging is paramount due to its critical role in oncological treatment planning and progress monitoring [43][49][50].

The complex vascular structure and vital functions of the liver necessitate precise imaging to accurately localize and characterize tumors. This precision enables tailored radiotherapy and surgical planning, ensuring effective treatment while preserving healthy liver tissue. Furthermore, ongoing segmentation allows for meticulous monitoring of treatment response and early detection of disease progression, facilitating timely adjustments to therapeutic strategies. These capabilities are essential for improving patient outcomes and advancing liver cancer treatment. Despite its importance, liver segmentation in CT scans remains a challenging task [51]-[55].

2.3 Artificial Intelligence in Healthcare

This sub-chapter provides a concise overview of how AI is transforming healthcare by enhancing diagnostic and treatment methods. Through the integration of AI into medical imaging, AI can analyze large volumes of medical data to detect patterns and anomalies with greater accuracy than traditional methods.

2.3.1 Definition

AI consists of a domain of science intended to provide computers with the ability to perform functions such as logic, reasoning, organization, learning and conceptualization [1]. It comprises methodologies such as ML, DL, and computer vision (CV), which relationship is presented Figure 5.

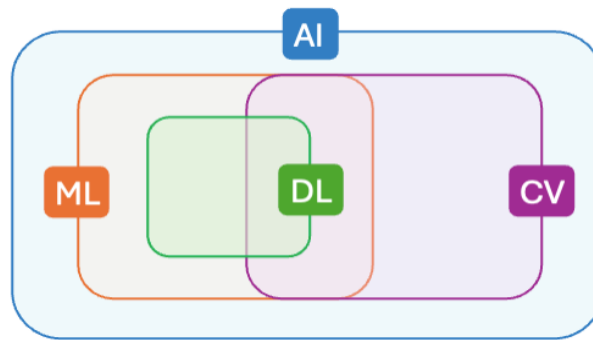


Figure 5. Relationship between AI, ML, DL and CV.

AI methodologies include:

- **ML:** branch of AI focused on creating algorithms that improve performance on tasks by learning from data patterns, without being explicitly programmed.
- **DL:** subset of ML using Artificial Neural Networks (ANNs) with multiple layers to automatically learn features from raw data.
- **CV:** branch of AI aimed at enabling computers to interpret and analyse visual data from images or videos.

These methodologies allow machines to perform tasks which usually require human intelligence, involving the ability to reason, discover meaning, generalize, or learn from previous experiences. In the context of healthcare, AI concerns the use of these advanced algorithms and software to emulate human cognition in the analysis, interpretation and understanding of complex medical data [2].

Applying AI in healthcare involves many fronts, redefining the paradigms of patient care through more accurate diagnosis, optimized treatment guidelines and better patient management in various domains. On a specific front, AI technologies are increasingly integrated into medical imaging, where they interpret large sets of health data, identifying patterns and predicting health outcomes with a level of precision and speed. This application not only aids in the detection and characterization of diseases, but also in monitoring progression and response to treatment, thus optimizing patient outcomes.

Other relevant concepts which might provide a broader understanding of the scope of the project are automation and computational costs:

- Automation

Refers to applying technology to perform tasks with minimal (semi-automation) or no (fully automation) human intervention. In the healthcare sector, automation can simplify processes, reduce manual errors and increase the efficiency of clinical workflows. In this regard, automated systems can manage patient records, schedule appointments and even help optimize the diagnostic procedure, allowing healthcare professionals focus on providing care to their patients.

- Computational costs

Involve the resources needed to process and analyze the data, including time, memory and power consumption. In the context of AI in healthcare, computational costs can be high given the complex algorithms and large datasets to be processed. The efficient management of these costs is essential for guaranteeing that AI applications are effective and sustainable in clinical contexts.

2.3.2 Convolutional Neural Networks

This sub-chapter introduces Convolutional Neural Networks (CNNs) and how they can be used in applications such as image classification, object detection, and segmentation.

2.3.2.1 Definition

CNNs are a subset of ANNs optimized for analyzing visual imagery. CNNs consist of multiple layers, including the input layer, convolutional layer(s),

pooling layer(s), and fully connected layer(s), each playing a crucial role in feature extraction and transformation [56].

2.3.2.2 Architecture

An illustrative representation of the architecture is shown in Figure 6 [57]-[59].

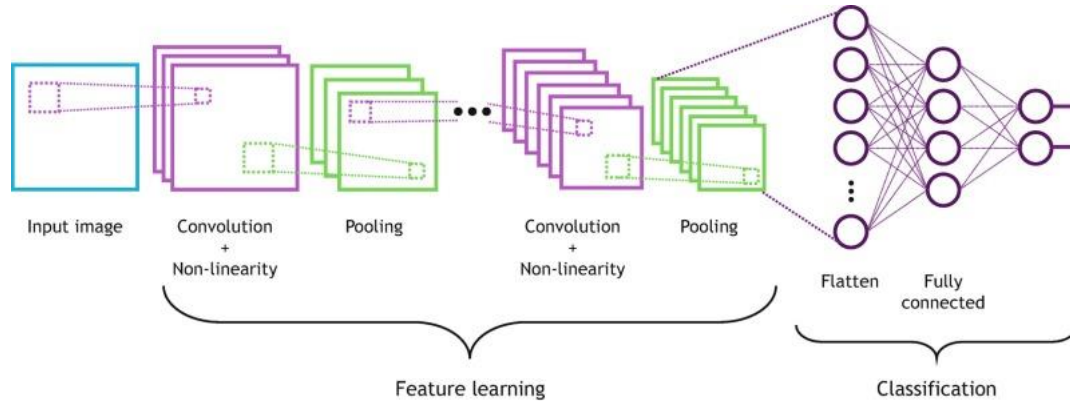


Figure 6. Illustration of the general architecture of CNNs. Source: [56].

The main layers used to build CNNs are:

1. Input layer

The input layer holds the raw input image data, with dimensions corresponding to the width, height, and depth (number of color channels) of the image.

2. Convolutional layer(s)

This layer applies filters to the input images, extracting features and creating feature maps. The filters are small matrices, and their sliding over the input image computes the dot product between the filter weights and the image patch.

3. Pooling layer(s)

Pooling layer(s) reduce the size of the feature maps, making computations more efficient and helping to prevent overfitting. Max pooling and average pooling are common techniques used to achieve this.

4. Flattening

After the convolution and pooling layer(s), the feature maps are flattened into a one-dimensional vector to be passed into fully connected layers for classification or regression tasks.

5. Fully connected layer(s)

Fully connected layer(s) take the flattened input and perform the final classification or regression tasks. Each neuron in this layer is connected to every neuron in the previous layer.

2.3.2.3 Training and Weight Optimization Processes

After the forward pass through the CNN's layers, the network employs backpropagation to optimize its weights, calculating the gradient of the loss function with respect to each weight using the chain rule, iterating backwards from the output layer to the input layer. The gradients are then used in gradient descent algorithms, such as Adam (Adaptive Moment Estimation) or RMSprop (Root Mean Square Propagation), to update the weights and minimize the loss function.

The entire dataset is passed through the network multiple times, with each complete pass called an epoch, allowing for iterative weight adjustments. The learning rate, a crucial hyperparameter, determines the size of these weight updates. Additionally, regularization techniques like L1/L2 regularization or dropout are often employed to prevent overfitting and improve the generalization capabilities of the model. Through this iterative process of forward passes, error calculation, backpropagation, and weight updates across multiple epochs, the CNN gradually refines its weights to minimize the loss function and enhance its performance on the given task.

2.3.2.4 Main Strengths and Weaknesses

Although CNNs have been increasingly used, they present both advantages and limitations [60]. Table 2.1 summarizes the main advantages and disadvantages of using CNNs, providing a balanced view of their capabilities and challenges.

Table 2.1. Advantages and disadvantages of using CNNs.

Advantages	Disadvantages
• Good capability of detecting patterns and features in images, videos, and audio signals. • Capable of end-to-end training, eliminating the need for manual feature extraction. • Efficiently handles large amounts of data and achieves high accuracy.	• Training CNNs can be computationally expensive and memory intensive. • Prone to overfitting if not enough data or proper regularization techniques are used. • Requires large amounts of labeled data for effective training. • Limited interpretability, making it challenging to understand what the network has learned.

2.3.3 Ground Truth

Ground truth in AI refers to the actual reality that a model aims to replicate or predict. It is essentially the target for training or validating models using labeled datasets. During the model inference phase, predictions are compared to the ground truth to evaluate the performance of the model. Creating ground truth datasets involves tasks like data labeling, model design, and training/testing. These datasets are usually annotated manually by experts, ensuring the data is accurate and reliable. This accurate data annotation is critical, particularly in

medical imaging, where precise labeling directly impacts the effectiveness of ML algorithms in diagnosing and understanding health conditions [61]-[63].

In medical imaging, ground truth datasets are invaluable. They consist of various types of medical scans, which are annotated to provide a reference for training models. High-quality ground truth datasets in medical imaging are essential because they enable ML and DL models to learn and recognize patterns within medical images, thereby aiding in the diagnosis and treatment of diseases. The process of creating these datasets is complex due to the need for diverse and high-resolution images that come with extensive metadata and clinical reports.

The application of ground truth in medical imaging involves using these well-annotated datasets to train AI models. These models are then employed to assist medical professionals in making more accurate diagnoses by recognizing patterns and anomalies in medical images that might be difficult to detect manually. High-quality datasets, which include images from various sources and devices, enhance the ability of the model to generalize and perform well in real-world scenarios. This reduces the risk of biases and errors, leading to better patient outcomes. Consequently, the effort put into creating and validating ground truth datasets is critical, as it forms the foundation for developing reliable and effective medical imaging models.

2.3.4 Application in Medical Image Segmentation

The number of publications using AI techniques to analyze static medical images has risen from hundreds to thousands in recent years. This growth is driven by the pattern-recognition demands of diagnostic tasks and the availability of organized images. Key applications in medical imaging include denoising, dose reduction, automatic segmentation, case triage, and image restoration [58][61]-[64].

The application of medical image segmentation spans various medical fields, impacting areas such as radiology, gastroenterology, histology, ultrasound imaging, and cancer detection. For instance, in radiology, AI-based segmentation helps in processing vast amounts of images from X-rays, CT scans, and MRIs, enabling more accurate and scalable diagnostic capabilities. Similarly, in gastroenterology, segmentation techniques can improve the detection of conditions like cancerous polyps and ulcers, often identifying issues that might be overlooked by human specialists. Histology benefits from segmentation by accurately applying staining protocols, thus aiding in the detailed analysis of tissue samples [65].

Additionally, the need for large, accurately labeled datasets necessitates extensive manual annotation, which is both time-consuming and expensive. Ensuring data quality and managing the vast amounts of data involved in training and validating these models further complicates the process.

Addressing these challenges is crucial for the successful integration of AI in medical image segmentation, ultimately enhancing diagnostic accuracy and improving patient outcomes.

3 REVIEW OF IMAGE SEGMENTATION TECHNIQUES AND EVALUATION METHODS

The present chapter contains the methodology adopted in order to carry out the analysis of studies and works related to liver segmentation in CT scans, the description of the medical imaging segmentation techniques considered for the literature review, the selected studies that were considered for the review of the state of the art, the commercialized software that was used by HUC during a trial period and, finally, a comparison of some public DICOM viewers.

3.1 Image Segmentation Techniques

The evolution of image processing technologies has given rise to a spectrum of segmentation categories. Inspired by existing segmentation techniques, authors in [66] noted that these methods can be systematically classified. While common types like instance and semantic segmentation stand alone, other techniques often involve ANNs, thresholding, clustering, or other region/edge-based approaches. Table 3.1 presents a brief description of each segmentation category, emphasizing their main advantages, and disadvantages, and presents some application details [23].

3.2 Performance Evaluation Metrics

There are two main types of evaluation metrics of segmentation performance, region-based and contour-based metrics. Examples of illustration of the original image and its respective ground truth and automatic segmentation is presented in Figure 7, with True Positives (TP) in green, True negatives (TN) in blue, False negatives (FN) in red, False positives (FP) in orange. Specific information related to the evaluation metrics can be found in Appendix H and Appendix I related to the presentation given at the ICMASC conference.



Figure 7. Illustration of image segmentation results. (a) Original image. (b) Ground truth segmentation. (c) Automatic segmentation. (d) Classification outcomes.

For 2D scans, these metrics are calculated based on the value of the pixels (picture elements). In the case of 3D scans, the metrics consider voxels (volume pixels). Voxels are analogous to pixels in 2D scans. Each voxel contains a value that represents some property (such as color, density, or intensity) in the 3D space.

Table 3.1. Comparison of segmentation categories.

Category	Description	Main advantages	Main limitations	Applicability
Threshold	Segments based on intensity thresholds	Simple, fast, easy to implement	Sensitivity to threshold selection, suffers from noise and artifacts	Commonly used in cases where clear intensity differences between the Region of Interest (ROI) and other regions exist
Cluster	Segments based on similar data patterns or clusters	Efficient grouping and identification of similar data patterns	Sensitivity to initialization and noise, limited to specific data distributions	Useful for identifying patterns and groups within the data, but may struggle with irregular shapes
Edge	Segments based on intensity transitions between regions	Accurate delineation of organ boundaries and structures	Sensitive to noise, difficulties with capturing complex structures	Suitable for images with clear organ boundaries and well-defined edges, but may struggle with low-contrast areas
Region	Segments based on homogeneous regions within an image	Simple implementation, intuitive methodology	Sensitive to initialization	Often used in cases where interpretability is a concern, but may struggle with fine details
Semantic	Segments based on semantic meaning of pixels	Pixel-level segmentation, fine-grained structural detail	Complexity in implementation, resource-intensive, and computationally demanding	Suitable for segmenting anatomical structures with distinct features
Neural Network	Learns models to segment images based on learned features	High accuracy, efficient learning from data	Requires large training datasets, computationally intensive	Suitable for various types of data due to its flexibility and adaptability

2D Liver Segmentation in Abdominal Computed Tomography Scans

Research on liver segmentation in medical imaging demonstrates promise, yet the lack of consistent evaluation metrics is often overlooked. The upcoming analysis will prioritize examining these metrics, focusing on those commonly used in automatic segmentation studies. The selected studies for the analysis of the evaluation metrics most commonly used, dated between 2014 and 2024, employ fully automatic liver segmentation techniques and metrics aligned with those in related works (Table 3.2). The research was sourced from **Google Scholar**, **Scopus**, and **Association for Computing Machinery Digital Library (ACM Digital Library)**, using keywords like “*Liver Segmentation*” and “*Segmentation Techniques*” to identify relevant studies.

Table 3.2. Summary of segmentation methods and main performance metrics used in the selected studies.

Author	Year	Main metrics used
Platero et al. [67]	2014	DSC
Mostafa et al. [68]	2015	Accuracy
Dou et al. [69]	2016	ASSD; VOE
Chartrand et al. [70]	2017	DSC
Christ et al. [15]	2017	DSC
Zhang et al. [71]	2017	DSC
Rafiei et al. [72]	2018	DSC
Hiraman [73]	2018	ASSD;MSSD;VOE
Ahmad et al. [74]	2019	DSC
Wang et al. [75]	2019	DSC
Lebre et al. [76]	2019	DSC
Almotairi et al. [77]	2020	Accuracy
Ayalew et al. [78]	2021	DSC
Pettit et al. [79]	2022	Accuracy
Sabir et al. [80]	2022	Accuracy; DSC; Precision
Scicluna [81]	2022	ASSD; DSC; MSSD
Olivares et al. [82]	2023	Accuracy; DSC
Ezzat et al. [83]	2023	Accuracy
Shao et al. [84]	2024	ASSD; DSC; Precision

ASSD: Average Symmetric Surface Distance DSC: Dice Similarity Coefficient

MSSD: Maxium Symmetric Surface Distance VOE: Volumetric Overlap Error

As a result, the most commonly used metrics in these studies were the ones shown in Figure 8, which have been selected for analysis in this sub-chapter.

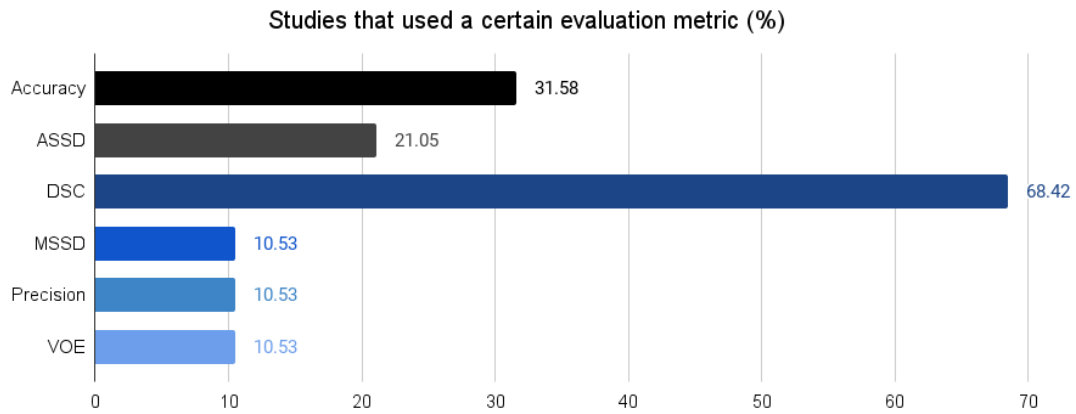


Figure 8. Distribution of the metrics used in the selected studies.

3.2.1 Region-based Metrics

As previously mentioned, region-based metrics assess the correctness of the area covered by segmentation, regardless of contour precision (Figure 9).

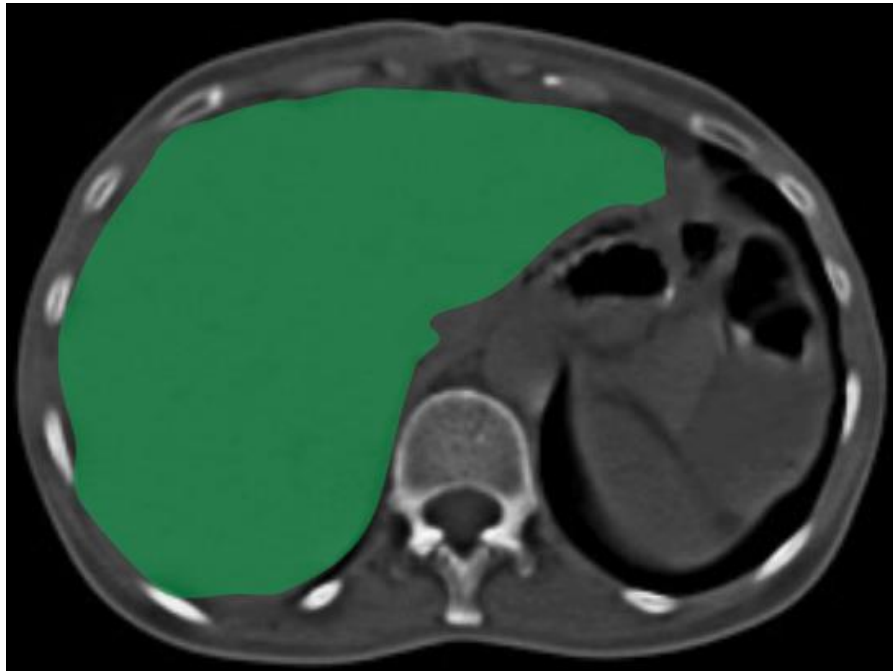


Figure 9. Region-based illustration.

These include Accuracy, Precision, Dice Similarity Coefficient (DSC), and Volume Overlap Error (VOE).

3.2.1.1 Accuracy

The accuracy is calculated as the ratio of the number of correctly identified pixels/voxels (both liver and non-liver) to the total number of pixels/voxels in the image or volume of interest. In mathematical terms, accuracy can be defined as presented in Equation (2) [85][86]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (2)$$

where:

- **TP (True Positives)** are the pixels/voxels correctly identified as liver tissue by both the algorithm and the ground truth.
- **TN (True Negatives)** are the pixels/voxels correctly identified as non-liver tissue by both the algorithm and the ground truth.
- **FP (False Positives)** are the pixels/voxels incorrectly identified as liver tissue by the algorithm but not by the ground truth.
- **FN (False Negatives)** are the pixels/voxels incorrectly identified as non-liver tissue by the algorithm but identified as liver tissue by the ground truth.

Figure 10 illustrates the calculation of the accuracy metric in order to translate the mathematical equation into a visual explanation.

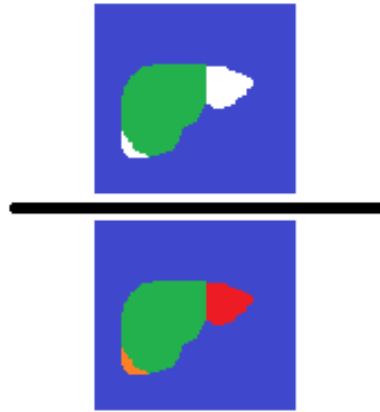


Figure 10. Illustration of the calculation of accuracy, with TP in green, TN in blue, FN in red, and FP in orange.

Accuracy provides a general overview of the segmentation performance, combining both the true positive rate (sensitivity) and the true negative rate (specificity) into a single measure. A high accuracy value, close to 1 (or 100% when expressed as a percentage), indicates a high level of agreement between the segmentation of the algorithm and the ground truth, implying an effective segmentation performance. Conversely, a lower accuracy value suggests

discrepancies between the output of the algorithm and the ground truth, highlighting potential areas for improvement in the segmentation process.

3.2.1.2 Precision

Precision is focused on the quality of the positive predictions made by the segmentation algorithm, meaning it assesses the proportion of pixels/voxels correctly identified as liver tissue out of all the pixels/voxels classified as liver by the algorithm [86]. Mathematically, the precision metric is defined as presented in Equation (3):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

Figure 11 illustrates the calculation of the precision metric in order to translate the mathematical equation into a visual explanation.

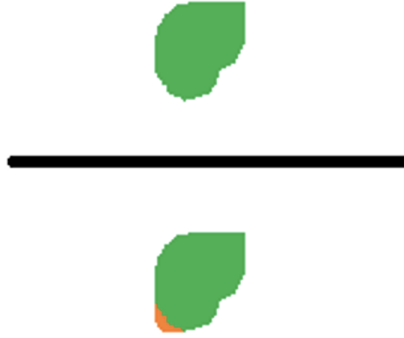


Figure 11. Illustration of the calculation of precision, with TP in green, and FP in orange.

A high precision indicates that a large majority of the pixels/voxels labeled by the segmentation algorithm as part of the liver are indeed liver tissue according to the ground truth. This implies that the algorithm is effective in avoiding over-segmentation, where non-liver tissues are mistakenly included within the segmented region of the liver.

3.2.1.3 Dice Similarity Coefficient

The DSC metric for assessing the performance of liver segmentation algorithms in CT scans, also known as the Sørensen-Dice coefficient/index or Dice coefficient, is a widely used evaluation metric in the field of medical imaging [86][87]. The DSC equation is presented in Equation (4):

$$\text{DSC} = \frac{2 * |S \cap G|}{|S| + |G|} \quad (4)$$

where:

2D Liver Segmentation in Abdominal Computed Tomography Scans

- $|S \cap G|$ represents the intersection, or the common pixels/voxels, between the segmented liver region S and the ground truth G . This intersection quantifies the number of pixels/voxels accurately identified as liver tissue by the segmentation algorithm.
- $|S|$ and $|G|$ denote the total number of pixels/voxels within the segmented liver region and the ground truth, respectively.

Figure 12 illustrates the calculation of the DSC metric in order to translate the mathematical equation into a visual explanation.

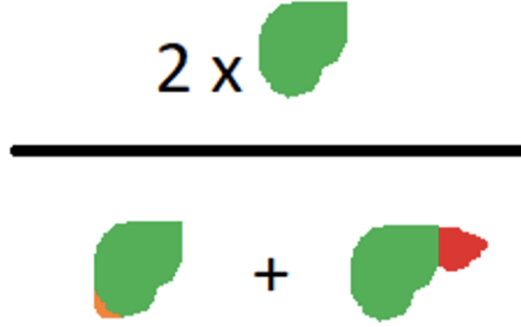


Figure 12. Illustration of the calculation of DSC, with TP in green, FN in red, and FP in orange.

The DSC ranges from 0 to 1, where 0 indicates no overlap between the segmented liver and the ground truth (indicating poor segmentation performance), and 1 represents perfect alignment or overlap (indicating ideal segmentation accuracy).

3.2.1.4 Volumetric Overlap Error

VOE is expressed as a percentage, representing the proportion of the total volume that does not overlap between the segmentation of the algorithm and the ground truth [88]. In 2D images, this is often referred to as the Area Overlap Error (AOE) or simply the overlap error, and it works on the same principle of comparing areas instead of volumes. It is calculated using the Equation (5):

$$\text{VOE} = \left(1 - \frac{|S \cap G|}{|S \cup G|}\right) * 100\% \quad (5)$$

where:

- $|S \cap G|$ denotes the intersection between the segmented liver region S and the ground truth G , as above.
- $|S \cup G|$ signifies the union of S and G , encompassing all the area/volume identified as liver tissue by either the segmentation algorithm or the ground truth.

Figure 13 illustrates the calculation of the VOE metric in order to translate the mathematical equation into a visual explanation.

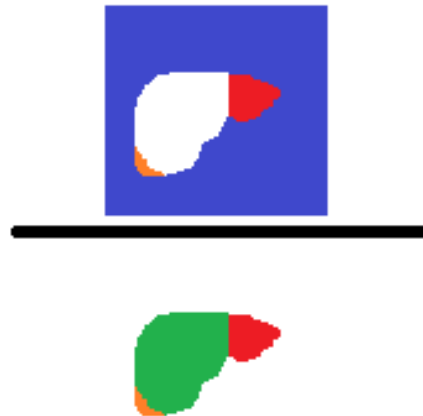


Figure 13. Illustration of the calculation of VOE, with TP in green, TN in blue, FN in red, and FP in orange.

A lower VOE indicates a higher degree of overlap between the segmented liver and the ground truth, reflecting a more accurate segmentation. Conversely, a higher VOE points to a greater discrepancy between the segmented region and the ground truth, signaling potential inaccuracies in the segmentation algorithm.

3.2.2 Contour-based Metrics

Also, as previously mentioned, contour-based metrics measure the correctness of boundary delineations (Figure 14).

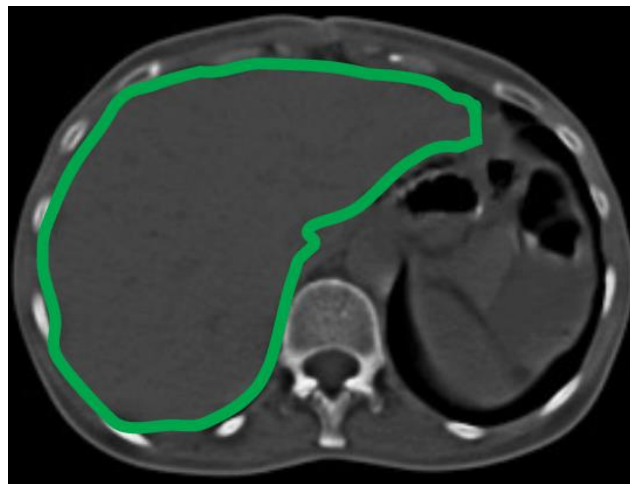


Figure 14. Contour-based illustration.

These include Average Symmetric Surface Distance (ASSD) and Mean Symmetric Surface Distance (MSSD).

3.2.2.1 Average Symmetric Surface Distance

ASSD is calculated by measuring the average distance from the surface points of the segmented liver to ground truth [87][88]. To ensure symmetry and a comprehensive evaluation, this process is often performed in both directions: from the segmented liver to the ground truth and vice versa. This bidirectional approach ensures that the metric accounts for all discrepancies, not just the most favorable ones. The ASSD equation is as presented in (6):

$$\text{ASSD} = \frac{1}{2} \left(\frac{\sum_{i=1}^N d(S_i, G)}{N} + \frac{\sum_{j=1}^M d(G_j, S)}{M} \right) \quad (6)$$

where:

- S_i denotes a point on the segmented liver surface.
- G_j denotes a point on the ground truth liver surface.
- $d(S_i, G)$ is the shortest distance from the point S_i on the segmented surface to any point on the ground truth surface G .
- $d(G_j, S)$ is the shortest distance from the point G_j on the ground truth surface to any point on the segmented surface S .
- N is the total number of points on the segmented liver surface.
- M is the total number of points on the ground truth liver surface.

Figure 15 illustrates the calculation of the ASSD metric in order to translate the mathematical equation into a visual explanation.

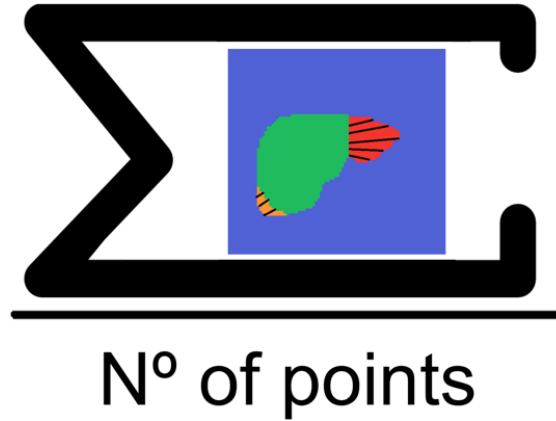


Figure 15. Illustration of the calculation of ASSD, with TP in green, TN in blue, FN in red, and FP in orange.

The ASSD value provides a numerical representation of the segmentation performance, with lower values indicating a better alignment between the segmented liver and the ground truth.

3.2.2.2 Maximum Symmetric Surface Distance

Maximum Symmetric Surface Distance (MSSD), also known as the Symmetric Hausdorff Distance, measures the maximum distance between the surfaces of the segmented liver region and the ground truth. MSSD focuses on the worst-case scenario by identifying the largest single discrepancy between the two surfaces [87][88].

The computation of MSSD involves calculating the maximum distance from any point on the segmented liver surface to the closest point on the ground truth surface and vice versa. This ensures that the metric captures the most impactful errors in segmentation, regardless of their direction. The MSSD is mathematically defined as follows in Equation (7):

$$\text{MSSD} = \max \left(\max_{i \in S} \min_{j \in G} d(S_i, G_j), \max_{j \in G} \min_{i \in S} d(G_j, S_i) \right) \quad (7)$$

where:

- S_i represents a point on the surface of the segmented liver region.
- G_j represents a point on the surface of the ground truth liver region.
- $d(S_i, G_j)$ denotes the Euclidean distance between a point S_i on the segmented surface and a point G_j on the ground truth surface.
- $d(G_j, S_i)$ denotes the Euclidean distance between a point G_j on the ground truth surface to a point S_i on the segmented surface.

Figure 16 illustrates the calculation of the MSSD metric in order to translate the mathematical equation into a visual explanation.

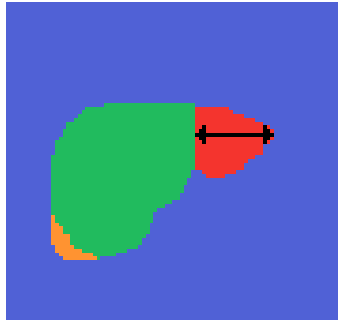


Figure 16. Illustration of the calculation of MSSD, with TP in green, TN in blue, FN in red, and FP in orange.

MSSD is expressed in physical units, such as millimeters, reflecting the spatial resolution of the CT scans. A lower MSSD value indicates a closer match between the segmented liver and the ground truth, signifying higher segmentation accuracy. Conversely, a higher MSSD value points to discrepancies between the segmented and ground truth surfaces, highlighting areas where the segmentation algorithm may need improvement.

3.2.3 Findings

Table 3.3 presents an overview of the metrics, along with their main advantages and shortcomings. Despite some studies validating the effectiveness of these metrics in tackling common challenges such as under-segmentation (Figure 17 (a)) and over-segmentation (Figure 17 (b)), which often appears when the HU of liver is close to the HU of other tissues or background, there remains an evident need for the adoption of standardized metrics in the field.

- Under-segmentation

Represents when the regions within an image have not been sufficiently segmented, leading to the merging of regions that should be separate.

- Over-segmentation

Occurs when the regions within an image are divided into too many segments, creating unnecessary and small regions that should be combined.

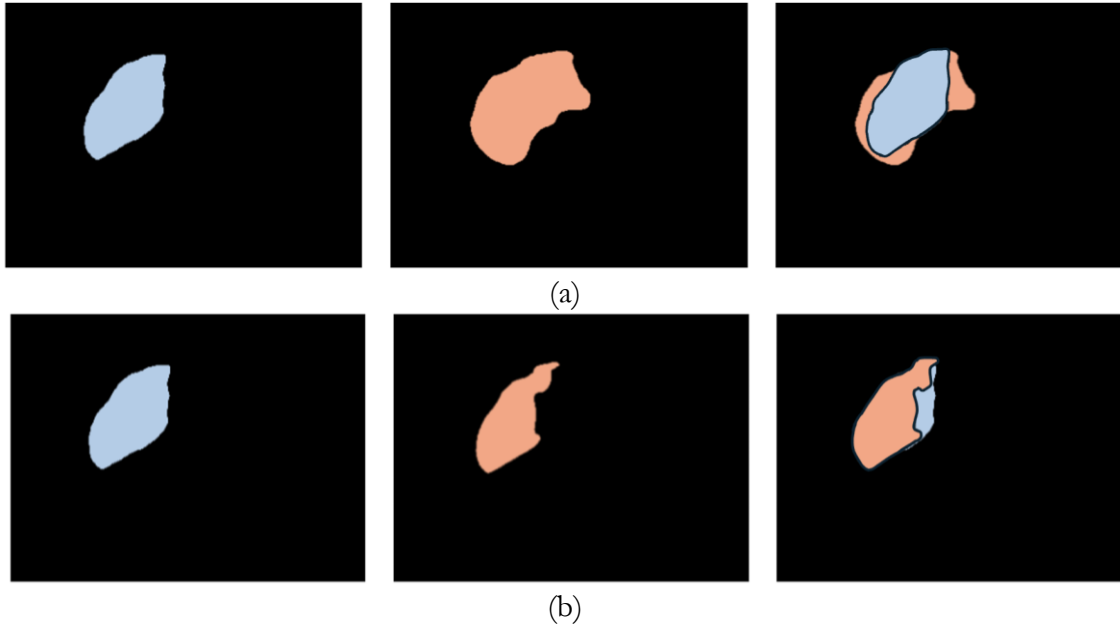


Figure 17. Common segmentation challenges. (a) Under-segmentation.
(b) Over-segmentation.

It is considered ideal to use at least one metric within each type of evaluation metric (region and contour). The introduction of such standards would greatly advance segmentation studies by highlighting key aspects, allowing for direct comparisons across various methods, and improving the reliability and effectiveness of automated liver segmentation in medical imaging.

Table 3.3. Summary of the main evaluation metrics used by automatic liver segmentation studies.

Category	Definition	Main advantages	Main disadvantages
Accuracy	Measures the proportion of correctly identified voxels (both liver and non-liver) to the total number of voxels, encompassing true positives and true negatives.	• Provides a comprehensive overview of the segmentation performance. • Easy to calculate and interpret.	• May be misleading in datasets with a large imbalance between liver and non-liver voxels. • Does not specifically focus on liver tissue accuracy.
ASSD	Calculates the average distance between corresponding points on the surfaces of the segmented liver region and the ground truth, in both directions.	• Offers a detailed assessment of boundary accuracy. • Highlights discrepancies in the surface alignment of the segmentation.	• More complex to compute and interpret compared to overlap-based metrics. • Can be influenced by outliers or extreme values.
DSC	Measures the overlap between the segmented liver region and the ground truth, calculated as the size of the intersection divided by the average size of the two regions.	• Provides a clear, straightforward measure of segmentation accuracy. • Easy to interpret and widely used in medical imaging.	• Might not fully capture small discrepancies in boundary delineation. • Can be overly optimistic in cases of large overlap but poor boundary alignment.
MSSD	Determines the maximum distance between any point on the segmented liver surface and the nearest point on the ground truth surface, and vice versa.	• Identifies the worst-case scenario in segmentation discrepancies. • Useful for assessing the reliability of segmentation in critical areas.	• Focused on the maximum error, potentially overlooking overall segmentation quality. • Can be overly punitive for generally accurate segmentations with minor outliers.
Precision	Assesses the proportion of voxels correctly identified as liver tissue out of all voxels classified as liver by the algorithm.	• Highlights the ability of the algorithm to avoid over-segmentation. • Crucial for applications where precision is more important than recall.	• Does not account for missed liver tissue (false negatives). • Can be misleading if the algorithm is overly conservative.
VOE	Quantifies the volumetric discrepancy between the segmented liver region and the ground truth, expressed as a percentage of the non-overlapping volume to the union volume.	• Directly related to clinical relevance, reflecting the actual volume error. • Intuitive and easy to understand.	• Can be sensitive to the size of the liver, with larger livers potentially showing higher absolute errors.

3.3 Literature Review

The literature review is organized into four main sections. It begins with a description of the research methodology in sub-chapter 3.3.1, followed by a brief overview and comparison of the public datasets used in the selected studies in sub-chapter 3.3.2. Sub-chapter 3.3.3 provides a historical perspective on the evolution of liver segmentation methods in medical imaging. Finally, sub-chapter 3.3.4 presents an analysis of the findings from the literature review in, analysing the main insights and gaps identified throughout the research process.

3.3.1 Research Methodology

This sub-chapter outlines the approach taken in conducting a systematic review of the literature. Details such as data sources, search queries, and inclusion and exclusion criteria can be found in Appendix E.

Once inclusion and exclusion criteria were applied to the data source tool, it was ensured that the studies selected were relevant to the research questions and met the needs of the project.

The oldest paper dates back to 1990 and the most recent to 2024. A representative paper was chosen for each year, with a potential total of 35 studies selected (the most relevant per year between 1990 and 2024, based on their content and results). A distribution of the total number of studies retrieved per year is given in Figure 18. As can be seen, some years do not have any studies, and thus a total of 30 documents were finally included in the historical review. It is also clear that interest in the topic has increased over the years.

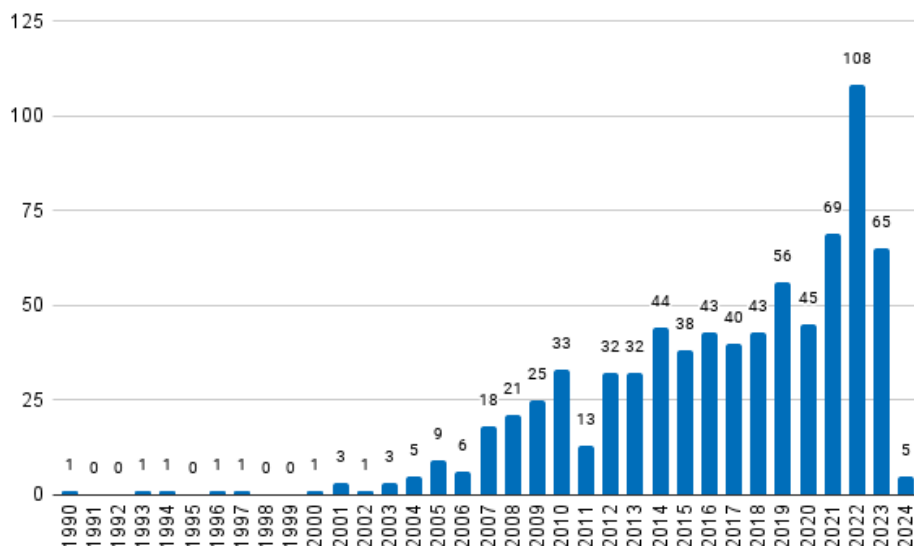


Figure 18. Distribution of the retrieved studies over the years.

A Word Cloud constructed from the titles of the selected studies is shown on the left side of Figure 19. On the right side is a pie chart showing the number of journals, conference proceedings and reports.

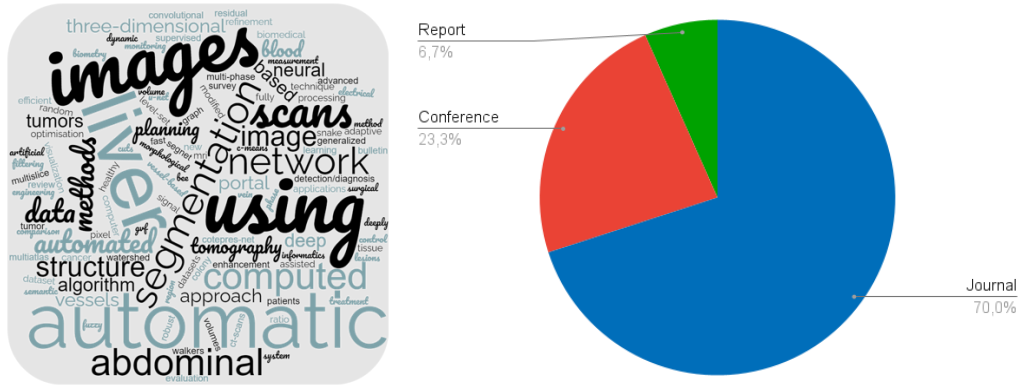


Figure 19. Characterization of the selected studies.

Figure 20 shows the selected studies categorized by image segmentation techniques according to the taxonomy of Sakshi and Kukreja (2023) [66].

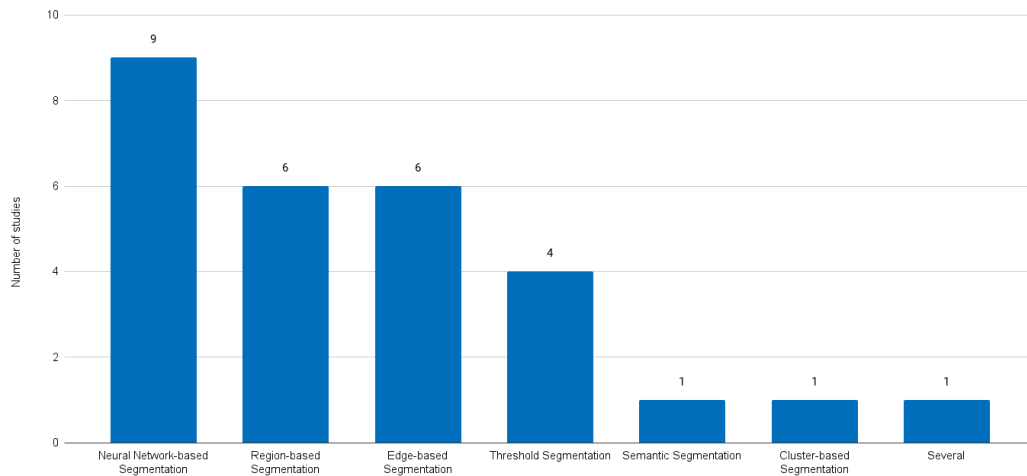


Figure 20. Selected studies distribution according to the categories of image segmentation techniques.

3.3.2 Public Datasets

The most common public datasets used for liver segmentation on CT scans include 3D-IRCADb01, LiTS17, MICCAI-SLiver07 and Combined Healthy Abdominal Organ Segmentation (CHAOS). Based on the comparison of the most used public datasets provided by Al-Saeed et al. [90], adding the CHAOS dataset (shown in Table 3.4), it is possible to identify several key differences that may have implications for data processing and analysis, such as different formats and differences in resolution between datasets which may require different approaches to processing and interpretation.

2D Liver Segmentation in Abdominal Computed Tomography Scans

Table 3.4. Essential information of each public dataset used by the studies analyzed.

Dataset	Format	Number of scans	Number of slices per scan	Resolution
3D-IRCADb01 [91]	DICOM	20	74 to 260 slices	512x512
CHAOS [92]	DICOM	40	77 to 105 slices	512x512
MICCAI-Sliver07 [93]	RAW	30	74 to 260 slices	512x512
LiTS17 [94]	RAW	200	42 to 1024 slices	Vary through subjects

3.3.3 Historical Overview

As mentioned in sub-chapter 3.3.1, one study per year between the years 1990 and 2024 was selected, summing to a total of 35 documents. For five years the search retrieved no results (see Figure 18) resulting in 30 documents. Unfortunately, the full text of two works was not available. The remaining 28 documents can be found in Appendix E, as well as the analysis of other studies review.

3.3.4 Findings

The most recent of the review studies found dates from 2022. One of the contributions of this work is to present a more up-to-date view of the works published since then. In addition, none of the other works presents a historical perspective on the subject, starting from 1990, as is the case with the present review.

A summary of the studies reviewed in sub-chapter 3.3.3 is given in Table 3.5. The columns of the table contain the following information: identification of each study (column *Authors*); publication year (column *Year*); general category or approach used in the segmentation method as defined in [66] (column *Segmentation Category*); the specific segmentation technique or algorithm (column *Method*); whether the method is fully automatic or semi-automatic (column *Automation Level*); whether the segmentation is performed in 2D or 3D (column *Dimensionality*); the dataset used for evaluation (column *Dataset*); and the key results of the best segmentation method in each paper (column *Results*). Graphical depictions of the metrics used in one or more studies are shown in Figure 21.

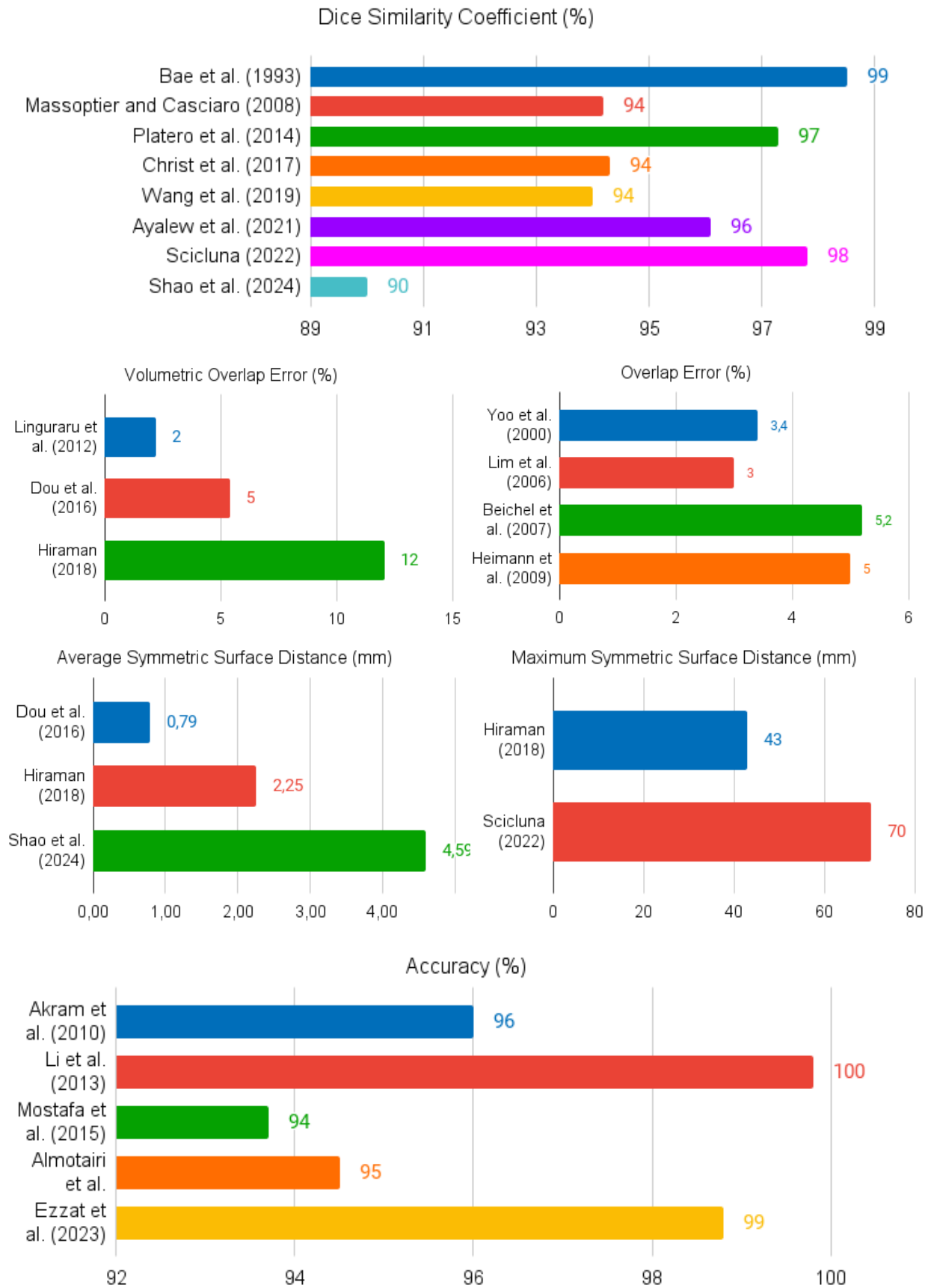


Figure 21. Summary of the quantitative results presented in the reviewed documents.

2D Liver Segmentation in Abdominal Computed Tomography Scans

It is clear from Table 3.5 that prior to 2016 there was no predominant category. Region-based Segmentation, Edge-based Segmentation, Threshold Segmentation, Semantic Segmentation, Cluster-based Segmentation and even combinations of several methods were tried. However, since 2016, Neuronal network-based techniques have dominated the field. Remembering the huge impact that AlexNet [117] had in winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [118], it is clear that liver segmentation in CT scans took a few years to catch up with the state of the-art research. This is probably due to the arrival of U-Net [119], which was proposed in 2015, and is specifically designed for biomedical image segmentation.

In terms of automation, both fully automated and semi-automated techniques have been explored (Figure 22). While it is nice to have fully automatic, accurate and fast techniques, the final decision should always belong to the specialist. Thus, the development of fully automatic methods for contour initialization is advocated, together with the development of intuitive tools that allow specialists to modify the fully automatically generated contour, if they feel the need to do so. This point of view is in line with current clinical practice, where the specialist follows contouring guidelines from respected entities such as the European Society for Radiotherapy and Oncology, the American Society for Radiation Oncology, or the Global Harmonization Group [120], while being allowed to use built-in auto-delineation and interpolation tools [121]. Therefore, it is agreed with Sarria et al. [121] in that, while AI can improve the accuracy and consistency of contouring, it cannot replace the knowledge and clinical judgement of radiation oncologists, physicists, or radiation therapists. AI should be used as a tool to support and optimize clinical decision-making and not as a substitute for human expertise.

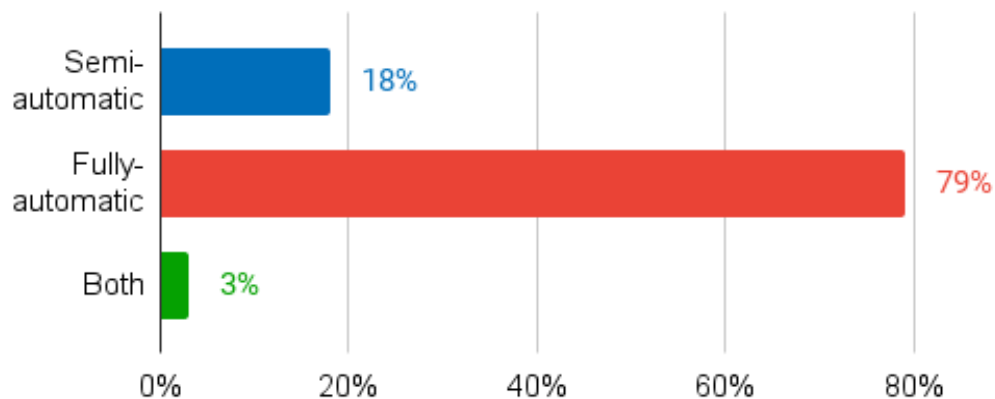


Figure 22. Percentage of studies according to their level of automation (semi-automatic, fully automatic or both).

Table 3.5. Summary of reviewed studies.

Author	Year	Segmentation category	Method	Level of automation	Dimensionality	Dataset	Main results
Bae et al. [95]	1993	Threshold	Gray-level Thresholding	Semi-automatic	2D	Private	0.985 DSC with mean percent error within 10%.
Gao et al. [96]	1996	Edge	Parametrically deformable contour model	Fully automatic	3D	Private	13.2% of the results required user modifications
Soler et al. [97]	1997	Region	Deformable models	Fully automatic	3D	Private	Claimed to be comparable to manual segmentations
Yoo et al. [98]	2000	Threshold	Threshold	Fully automatic	2D	Private	3.41% error
Pan and Dawant [99]	2001	Edge	Level sets	Fully automatic	Both	Private	[0.874,0.963] average similarities
Saitoh et al. [100]	2002	Threshold	Threshold	Fully automatic	3D	Private	~20 minutes computation time
Masumoto et al. [101]	2003	Region	Differences between time-phase images	Fully automatic	3D	Private	Volume ratio of 67% on average and 32% in the worst cases
Lim et al. [102]	2004	Region	Watershed	Fully automatic	2D	Private	Only qualitative

2D Liver Segmentation in Abdominal Computed Tomography Scans

Author	Year	Segmentation category	Method	Level of automation	Dimensionality	Dataset	Main results
Liu et al. [103]	2005	Edge	GVF snake	Semi-automatic	2D	Private	5.3% median value of the difference ratios
Lim et al. [104]	2006	Semantic	Labeling-based search	Fully automatic	2D	Private	96% average correctness; 3% average error rate
Beichel et al. [105]	2007	Region	Graph cuts	Semi-automatic	3D	Private	5.2% AOE
Massoptier and Casciaro [106]	2008	Edge	Active contour	Fully automatic	3D	Private	94.2% mean DSC
Heimann et al. [53]	2009	Several	Majority Voting	Both	Both	Private	5% Overlap error; -0.7 Volume difference; 0.8 ASSD; 1.7 RMS distance; 19.1 MSSD
Akram et al. [107]	2010	Threshold	Global Threshold	Fully automatic	3D	Private	0.96 average accuracy; 0.0017 std; 96% accurately segmented; 4% poorly segmented
Oliveira et al. [108]	2011	Edge	Level sets	Semi-automatic	2D	SLiver07	82.05 overall score
Linguraru et al. [109]	2012	Region	Graph cuts	Fully automatic	3D	Private; MICCAI-SLiver07	2.2 VOE

Author	Year	Segmentation category	Method	Level of automation	Dimensionality	Dataset	Main results
Li et al. [110]	2013	Edge	Fuzzy clustering and level set	Fully automatic	2D	Private	0.9986 average accuracy; 0.9989 average specificity
Platero et al. [67]	2014	Region	Graph cuts	Semi-automatic	3D	MICCAI-SLiver07	76.3 maximum score; 0.973 DSC
Mostafa et al. [68]	2015	Cluster	ABC optimization	Fully automatic	2D	Private	93.73% accuracy; 84.82% average SI
Dou et al. [69]	2016	Neural Network	3D DSN	Fully automatic	3D	MICCAI-SLiver07	5.42% VOE; 0.79mm ASSD
Christ et al. [15]	2017	Neural Network	CFCN	Fully automatic	2D	3D-IRCADb-01	94.3% mean DSC
Hiraman [73]	2018	Neural Network	CNN	Fully automatic	2D	MICCAI-SLiver07	12.07% average VOE; -1.96% RVD; 2.25mm ASSD; 2.60mm RMSD; 43.01mm MSSD
Wang et al. [75]	2019	Neural Network	CNN	Fully automatic	3D	Private	0.94 ± 0.06 DSC
Almotairi et al. [77]	2020	Neural Network	SegNet	Fully automatic	3D	3D-IRCADb-01	94.57% overall accuracy
Ayalew et al. [78]	2021	Neural Network	U-Net	Fully automatic	2D	3D-IRCADb-01; LiTS	0.9612 DSC

2D Liver Segmentation in Abdominal Computed Tomography Scans

Author	Year	Segmentation category	Method	Level of automation	Dimensionality	Dataset	Main results
Scicluna [81]	2022	Neural Network	U-Net and VGG16UNetC	Fully automatic	2D	CHAOS	85.84 mean score; 97.85 DSC; 80.33 RAVD; 94.80 ASSD; 70.38 MSSD
Ezzat et al. [83]	2023	Neural Network	CNN	Fully automatic	2D	Private	98.80% accuracy
Shao et al. [84]	2024	Neural Network	AC-Net	Fully automatic	3D	Private; LiTS	0.90 DSC; 0.82 JC; 0.92 recall; 0.89 precision; 11.96 HD; 4.59 ASSD

ABC: Artificial Bee Colony

AC-Net: Attention Connect Network

ASSD: Average Symmetric Surface Distance

CFCN: Cascaded fully CNN

CNN: Convolutional Neuronal Network

DSC: Dice Similarity Coefficient

DSN: Deeply Supervised Network

GVF: Gradient Vector Flow

HD: Hausdorff Distance

JC: Jaccard Coefficient

MSSD: Maximum Symmetric Surface Distance

NN: Neuronal Network

r: Correlation Coefficient

RAVD: Relative Absolute Volume Difference

RMS: Root Mean Square

RMSD: Root Mean Square Symmetric Surface Distance

RVD: Relative Volume Difference

SI: Similarity Index

std: standard deviation

VOE: Volumetric Overlap Error

Both 2D and 3D methods have been developed – see Figure 23. The 3D liver segmentation methods present two main advantages. On the one hand, fully 3D methods would make use of more contextual information and thus potentially provide better segmentations. On the other hand, the development of fully 3D techniques would avoid the need of the slice-based methods to aggregate all of the segmentations into a coherent volume. This aggregation could not only lead to errors and anatomically incorrect structures but would also increase computational time. Although the advantages of 3D methods are recognized, their disadvantages are also recognized, such as notably increased complexity of implementation, the need for more computational resources and possibly lower efficiency, increased data requirements, and difficulties in visualization. The choice must be an informed one and made on a case-by-case basis.

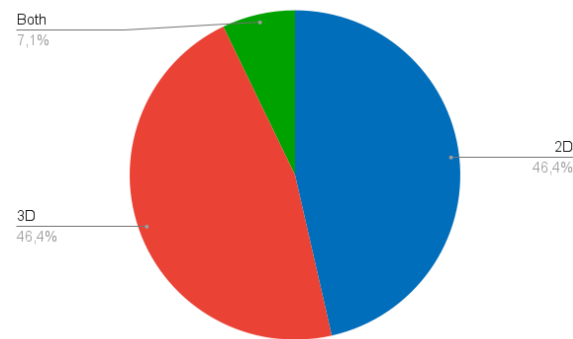


Figure 23. Percentage of studies according to their dimensionality of the image (2D versus 3D).

Data availability is a major concern, as most methods use private datasets (Figure 24). This inhibits reproducibility of the results. In addition, researchers who do not have access to hospitals or other facilities with CT scanners, cannot develop new techniques for this particular problem. It is advocated to make the data available, while respecting all the ethical considerations that are important when dealing with medical data and properly anonymizing any sensitive information. Some notable exceptions to publicly available datasets were previously listed in the sub-chapter 3.3.2.

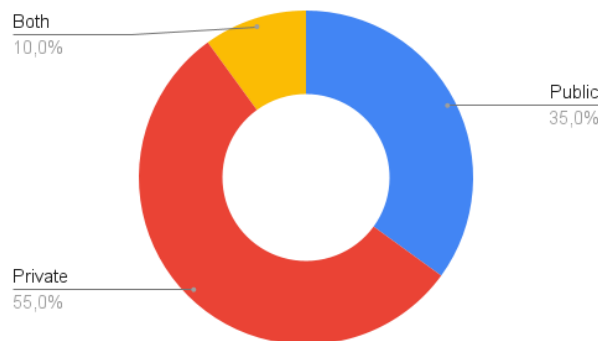


Figure 24. Percentage of dataset types.

2D Liver Segmentation in Abdominal Computed Tomography Scans

The most common public datasets used in studies of liver segmentation on CT scans include 3D-IRCADb01, LiTS17, MICCAI-SLiver07 and CHAOS (Figure 25).

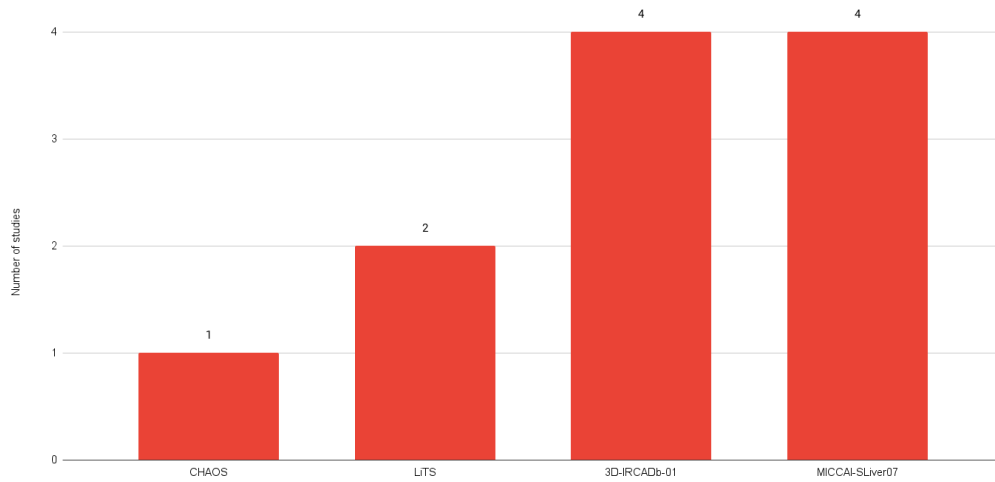


Figure 25. Number of studies using each public dataset.

3.4 Commercial Automated Segmentation Software Solutions

This sub-chapter seeks to evaluate and compare the three commercial automatic segmentation software solutions considered by the Department of Radiotherapy of the HUC: Annotate, Contour+ and Limbus Contour. Each of these platforms provides functionalities designed to improve processing times and the reliability of segmentation in abdominal CT scans, a common procedure in radiotherapy planning. The information presented in this sub-chapter is based on research carried out during the course of this project by me via the companies official websites, public pre-market notification information from the FDA.

3.4.1 Annotate by Therapanacea

Provides a labeling tool designed to enhance the precision of radiation therapy planning. The software is engineered to automatically segment anatomical structures and pathological areas from CT scans, allowing the user to edit manually and semi-automatically the contours for the ROIs. It is a web platform that can be launched from Google Chrome or Mozilla Firefox [122][123].

3.4.2 Contour+ by MVision AI

Contour+ is a software system for image analysis algorithms to be used in radiation therapy treatment planning workflows. The system includes processing tools for automatic contouring of CT scans. The produced segmentation templates for ROIs must be transferred to appropriate image visualization systems as an initial template for a medical professional to

visualize, review, modify and approve prior to further use in clinical workflows. This software is not intended for use with real-time adaptive planning workflows, and it is operated either by the manufacturer in the cloud of the manufacturer or by a healthcare provider in the healthcare IT network of the provider [124][125].

3.4.3 Limbus Contour by Limbus AI

Limbus Contour is an application for automatic contouring (segmentation) of CT/MRI DICOM images intended to be used by trained medical professionals to derive contours for input to radiation treatment planning. Limbus Contour does not display or store DICOM images and relies on existing radiotherapy treatment planning systems and DICOM image viewers for display and modification of generated segmentations. It is a stand-alone package which operates on Microsoft Windows operating system [126][127].

3.4.4 Comparison

Based on the detailed comparison table for the automated segmentation software present in Table 3.6, it is possible to observe that each of the software considered by the HUC has its particular advantages, they collectively underscore the role of AI-driven tools in improving the efficiency and accuracy of medical imaging segmentation.

Table 3.6. Comparison between the different commercial software solutions considered by HUC.

Aspect to Compare	Annotate	Contour+	Limbus Contour
Segmentation precision	High precision with robust contour adaptation	High precision	High precision
Processing Speed	≈ 3 minutes per scan in real-world scenarios		
Pricing	≈ 40 euros per scan		
Scalability	Well-suited for large hospital networks	Ideal for medium to large-scale settings	Adaptable to various hospital sizes
User needs	Designed for complex cases	Best for routine and repetitive tasks; needs minimal interaction	Suited for customizable workflows in research-oriented settings

2D Liver Segmentation in Abdominal Computed Tomography Scans

Aspect to Compare	Annotate	Contour+	Limbus Contour
Clinical outcomes impact	Enhances accuracy treatment planning	Focus on reducing planning time	Promotes consistent treatment planning
Advantages	Advanced algorithms for high adaptability; excellent for complex anatomical variations	Streamlined workflow integration; low operational training required	High customization; supports a wide range of imaging modalities
Disadvantages	Higher cost might limit accessibility for smaller institutions	May not handle rare anatomical cases effectively	Requires initial setup customization which can be time-intensive

3.5 Public DICOM Viewers

For the scope of this project, an assessment of the best approach for presenting the segmentation results has also been conducted, whether through a public DICOM viewer or through a self-developed application. A comparison initially made by [128] in 2016 but updated during the year 2024 within the implementation of this project between the different viewers is presented in Table 3.7, making it easier to choose which of them could be the best viewer to be used. The choice depended on the needs raised, as presented in sub-chapter 7.1 under requirements. Another factor is the usability assessment, considering the aspects previously assessed by [128] (Table 3.8).

OHIF, MITK, and 3DimViewer were selected for their essential features, such as cross-platform compatibility (Windows, macOS, and Linux), and their high usability scores. However, these publicly available tools had installation and troubleshooting challenges, especially for MITK and 3DimViewer due to limited documentation. Despite OHIF's larger community and resources, troubleshooting was still insufficient. Additionally, compatibility issues were identified with certain imaging equipment scans, including an inability to import and view images due to unrecognized DICOM Transfer Syntax UID tags.

Table 3.7. Comparison between public DICOM viewers. Source for the comparison structure: [128].

Aspect to Compare	3D Slicer	3DimViewer	MITK	Agnosco	Amide	MedINRIA	OHIF	Onis
Operating System	Windows, macOS, Linux	Windows, macOS, Linux	Windows, macOS, Linux	Windows	Windows, macOS, Linux	Windows, macOS, Linux	Windows, macOS, Linux	Windows
Image Transformation	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Export to DICOM files	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Export to non-DICOM files	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Anonymization	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
HU scale	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
More than a single series	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Plugin integration	Yes	Yes	Yes	No	No	Yes	Yes	Yes

Table 3.8. Usability assessment comparison provided by [128].

Aspect to Compare	3D Slicer	3DimViewer	MITK	Agnosco	Amide	MedINRIA	OHIF	Onis
Instalation	4	4	4	4	4	4	4	4
Import DICOM data	1	4	3	2	2	3	4	4
Working speed	1	1	1	2	4	4	4	4
Interface	2	2	3	3	2	3	4	4
Easy for use	1	3	1	3	2	2	4	4
Final assessment	9	14	12	14	14	12	20	20

3.6 Findings

In the field of liver segmentation, various techniques have been explored, each with its own strengths and limitations. While some methods have shown promise as previously presented, there are still several challenges that need to be addressed, along with opportunities for future research.

3.6.1 Adoption of Artificial Neural Network-based Methods

The growth of ANNs-based approaches has led to remarkable progress in liver segmentation, particularly with respect to CT scans. These models have led to a new era of accuracy and efficiency, outperforming traditional methods [117][119]. This improved accuracy has become critical in the field of medical imaging, where the correct interpretation of CT scans can directly affect the diagnosis and treatment plans of patients. Furthermore, the efficiency of these ANN models translates into faster processing times, allowing for more agile decision making in clinical settings.

Another positive aspect of ANNs in liver segmentation is their ability to cope with the complexity of liver anatomy. ANNs, with their systematic and complex pattern recognition capabilities, are able to navigate these anatomical variations. As a result, they are better able to deal with the variety of appearances that liver tissue can have on CT scans. This ability to handle complex datasets ensures that ANNs can provide consistent and accurate segmentation in a wide range of cases.

3.6.2 Dimensionality Impact for Liver Segmentation

In medical imaging, and in particular liver segmentation, the choice between slice-based 2D and volume-based 3D segmentation methods is crucial. Table 3.9 presents a comparison between both approaches, always considering that the choice between these two methods should be guided by the specific clinical requirements, and the available resources for implementing the technology [15][69][73][75][97].

Table 3.9. Comparison between 2D and 3D dimensionality.

Dimensionality	Advantages	Disadvantages
2D segmentation	Computationally easier and faster due to the less complex data structure associated to deal with individual 2D images rather than volumetric data, easier to integrate into clinical workflows.	Individual slices might not align perfectly, and it has an inability to account for the spatial relationships between different image slices, which can lead to less accurate

Dimensionality	Advantages	Disadvantages
		reconstructions of 3D structures.
3D segmentation	Processes the entire volume allowing for a more comprehensive analysis, maintains spatial relationships, preserves the spatial relationships within the data, and provides a continuous and more accurate representation of structures.	Requires more computational resources and time, especially as the volume of data increases and demands more advanced hardware and software, potentially increasing the cost and time of processing and development.

3.6.3 Main Limitations

The reliance on semi-automatic techniques that necessitate manual segmentation has been a persistent challenge in the field of liver segmentation. This approach not only consumes time but also introduces the potential for human error, impacting the overall efficiency and accuracy of the segmentation process. Addressing this limitation is crucial for streamlining the segmentation workflow and minimizing the risk of inaccuracies in medical imaging analysis.

The limitations imposed by smaller datasets have hindered the generalization of algorithms for broader applications. This constraint poses a barrier to the prevalent adoption of liver segmentation techniques, emphasizing the need for strategies to overcome dataset limitations and ensure the applicability of these methods across diverse clinical settings.

Additionally, the prevalence of issues such as over and under segmentation, and low accuracy, underscores the necessity for the development of more robust techniques.

Even commercially available software often has flaws when it comes to segmentation, and a very common complaint is the inclusion of the gall bladder in the segmentation of the liver, due to the proximity of both structures and the size of the vesicle, as can be seen in Figure 26, provided by [129].

The inconsistent evaluation metrics across studies hinders the comparison of the performance of the different methods. Additionally, the scarcity of user-friendly interfaces for physicians to edit segmentations limits the practical application in hospitals.

Overcoming these challenges is essential for enhancing the reliability and precision of liver segmentation, ultimately improving the quality of medical diagnoses and treatment planning.

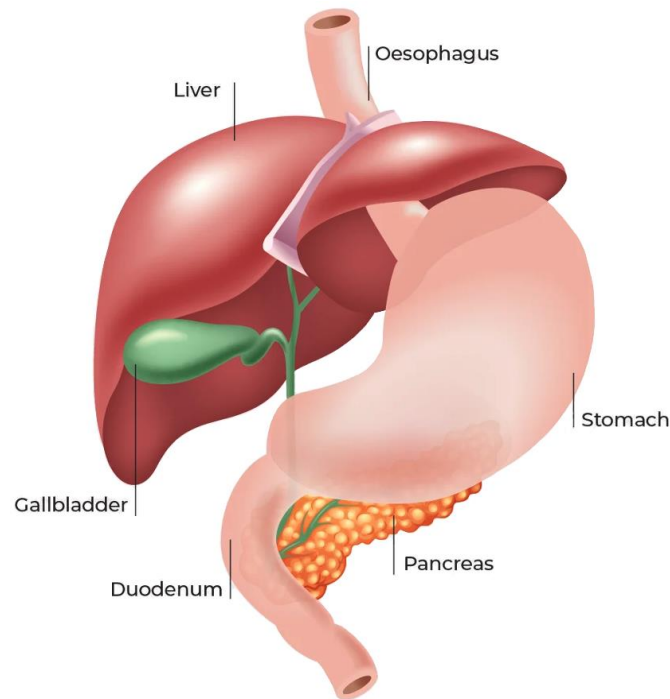


Figure 26. Illustration of the position of the gall bladder and liver. Source: [129].

3.6.4 Research Questions

It may be relevant to reflect on the answers to some of the research questions presented in chapter 1 (Introduction).

- **Q1 – What are the challenges and limitations associated with accurate liver segmentation in CT scans?**

Challenges and limitations associated with accurate liver segmentation in CT scans include under-segmentation, over-segmentation, low contrast, poor boundary detection and background segmentation due to noise. In addition, liver segmentation in CT scans is further challenged by the presence of artefacts, such as partial volumes, noise, and low sharpness and contrast between organs, making it difficult to identify the boundaries between different tissues.

- **Q2 – How does the choice of the method impact the accuracy and efficiency of liver segmentation in CT scans?**

The choice of the method has an important impact on the accuracy and efficiency of liver segmentation in CT scans. Traditional techniques such as image processing and region-growing approaches have shown varying degrees of sensitivity and specificity, with some challenges in dealing with large injuries. In contrast, newer methods such as FCN (Fully Convolutional Network) and techniques like ResU-Net and SegNet showed higher accuracy, with some reaching up the highest accuracy levels. Notably, the use of Graphics Processing Unit (GPU) has reduced processing times thus contributing towards more efficient and accurate liver segmentation methods.

- **Q3 – What are the evaluation metrics commonly used to assess the performance of AI models and traditional methods for liver segmentation in CT scans?**

Some of the key metrics used to measure the outcome of segmentation techniques include accuracy, precision, sensitivity, DSC, AOE/VOE, ASSD and MSSD. There is not much consistency in the metrics presented by the various studies except for DSC.

4 PRIVATE DATASET CREATION

This chapter describes the process of creating a dataset from scratch, step by step, based on the abdominal CT scans and their respective manual segmentation provided by the HUC Radiotherapy Department. All scans are anonymized, and the slices/images have been reviewed to preclude the presence of personal information.

4.1 Submission to the Hospital's Ethics Committee

In order to be granted access to the HUC's data/images, the project was submitted to a formal approval process at the hospital's Ethics Committee. This process is composed by a set of documents which allow an assessment of compliance with several criteria, notably:

- Objectives and relevance of the scope
- Methodology
- Curriculum vitae of the researchers involved
- Compliance with the European Union's General Data Protection Regulation (GDPR)
- Contribution to the HUC
- Scientific commitment and impact

4.2 Overview

Figure 27 shows a flowchart that broadly represents the process of selecting the scans, collecting them, and creating the final dataset.

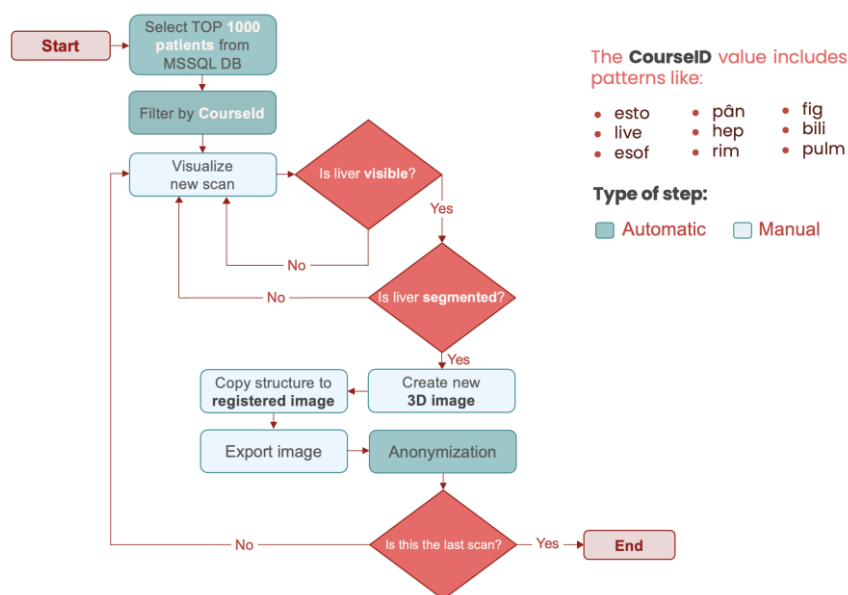


Figure 27. Flowchart of the operating sequence for the dataset creation.

4.3 ARIA Oncology Information System

The HUC Radiotherapy Department uses ARIA, an oncology information system developed by Varian Medical Systems, which supports the clinical needs of patients and the workflow requirements of treatment teams. ARIA integrates all information related to patients who have been and are being treated, including images acquired on the CT scanner for treatment purposes. The system enables the visualization of images associated with each CT scan for each patient through one of the applications on the ARIA network. Additionally, ARIA allows for the storage, viewing, and management of radiological images and data. An important feature of ARIA is its integration with an application for manual segmentation of structures, which aids in precise treatment planning [130].

4.4 Labeling Protocol

All the scans included in the dataset were manually labeled by a radioncologist, who meticulously labeled each scan slice by slice using a specific module/application from the ARIA network of HUC's Radiotherapy department. Any region of the scan without an associated label was designated as part of the "*Body*".

4.5 Structured Query Language Statement

The first step for creating the dataset involved defining the Structured Query Language (SQL) query for ARIA's MSSQL database presented in Appendix B.

This SQL query is executed on the ARIA database to fetch the first 1000 records from the 'Course' and 'Patient' tables within the 'VARIAN' schema, specifically selecting patient IDs, course IDs, and course start times. It performs an inner join on these tables using the 'PatientSer' column to associate each course with its respective patient. The query filters courses based on the 'CourseID' containing certain patterns, such as variations of 'esto', 'pân', 'fig', 'live', 'hep', 'bili', 'esof', 'rim', and 'pulm', irrespective of case sensitivity. The results are then ordered in descending order by the patient serial number, ensuring the most recent or highest patient serials are listed first.

This operation aims to identify and list specific types of treatment courses along with associated patient information, adhering to certain naming conventions within the course IDs, considering key words or characters related to the liver or nearby organs (stomach, pancreas, liver, hepatic, bile, esophagus, kidney, lung), in English and Portuguese.

4.6 Initial Scans Visualization

Based on the results of the previous SQL query, the scans obtained were structured in a table that allowed an introduction to all the most recent scans in which the liver could possibly be found, as shown in Table 4.1.

Table 4.1 Structure of the table created for the initial visualization of all the images.

Is liver completely visible?	Segmented liver?	Anonimization ID	Patient ID	Course ID	StartDate Time	Gender	Year of birth
Yes/No	Yes/No		**		yyyy-mm-dd hh:mm:ss	Female /Male	yyyy

where:

- Confirmation of whether or not the **scan shows the liver** in its totality and whether there has been **segmentation of the liver** previously carried out by a radioncologist.
- The **Anonymization ID** is auto generated when the CT scan is exported from ARIA to the folder used to collect all the CT scans, which is described in sub-chapter 4.9.
- The **Patient ID** is only available to the HUC, with no external access to this column when sharing the complete table for reasons of in-depth analysis of the database, but it is preserved on the HUC's side in case of need for specific analysis of a particular scan.
- The **Course ID** consists of a concise statement of the location of the treatment, e.g. "*liver*".
- The **StartDate Time** is the date in the format shown in the table, which is the date when the patient started receiving treatment for that course (localization) and therefore near to when the scan was acquired, giving an idea of the date from the most recent to the oldest, not the precise date when the scan was acquired.
- The patient's **Gender** and **Year of birth** are the only personal information of the patients to which it was decided to have access, in order to justify, if necessary, any aspect associated with the morphological constitution of the patient that may be relevant to the study, given that these are attributes that are also available for analysis and comparison with public datasets.

4.7 Outcome of the Initial Scans Visualization

As a partial result of this visualization of the images, the following broad overview was achieved, shown in Table 4.2.

Table 4.2. Result of the scans visualization.

Starting number of CT scans	Total number of scans with viewable liver	Total number of segmented scans	Total number of segmented slices/images
1000	821	246	11175

From the 1000 scans originally analyzed, only 24.6% could be used, which corresponds to the number of scans containing segmentation of the liver previously carried out by a radioncologist, where each scan has a minimum of 20 and a maximum of 132 slices with liver segmentation, achieving a total of 11175 slices (images).

4.8 Exporting Abdominal Scans and Ground Truth

Based on the results of the previous SQL query (subchapter 4.5), the scans obtained were structured in a table that allowed an introduction to all the most recent scans in which the liver could possibly be found, as shown in Table 4.2. Subsequently, various techniques for exporting scans that had manual liver segmentation were considered.

As these images had segmentation of more organs and structures than just the liver, the possibility of developing an automatic script that would filter each scan and export only the images and the segmented liver structure, instead of all the present structures, was considered. The problem was that the scans segmentation at HUC has been done by different professionals and not even the same professional defined the same ID for a given structure. For example, it was possible to find the word “*liver*” with different abbreviations, with numbers, in conjunction with other words, and in English or Portuguese, making it difficult to opt for developing a script, because basically each scan presented a different ID for the segmentation of the same structure.

Once it was decided that creating a script would be counterproductive to the time that had been defined for the process (considering all the time already invested in the database creation process, mainly for documentation for the hospital’s Ethics Committee), the possible alternatives for manual export were analyzed. The chosen strategy for the export process is presented in Appendix A, resulting in the least time-consuming alternative, being a process of approximately 8-10 minutes per scan/patient.

4.9 Scans Anonymization Process

DICOM standards stipulate stringent guidelines for the anonymization of input DICOM data to safeguard against the inadvertent disclosure of Individually Identifiable Information (III) concerning patients or affiliated individuals or entities, complying with General Data Protection Regulation (GDPR). These guidelines, outlined in Chapter E of the DICOM standard documentation [35], meticulously detail procedures for the removal and substitution of Attributes within a DICOM Dataset that possess the potential to disclose such sensitive information. By adhering to these protocols, healthcare providers and organizations can ensure the protection of patient privacy and confidentiality while facilitating the secure exchange and utilization of medical imaging data for therapeutic purposes. Based on this documentation, the anonymizing script can be found in Appendix C.

This script modifies and deletes all the pertinent DICOM tags for all the images of the same scan (each in its own DICOM file) and for the DICOM file containing the Radiotherapy Structure Set (RSST), saving each CT scan and its respective structure in the same folder named after the Anonymization ID of the corresponding patient, which is a random id generated to replace the patient's ID provided by the HUC radiotherapy department.

The “-m” command in the script represents the modification of the respective DICOM tag, while “-ma” modifies the same tag for all those found in the DICOM header. On the other hand, the “-ea” command removes all DICOM tags that exist in the DICOM image header.

In sub-chapter 5.3.1.1 (Private Dataset Analysis), a final characterization of the dataset is presented.

5 METHODOLOGY

This chapter provides a description of the methodology adopted to carry out the project, including stages such as preparing the data, creating the segmentation model for its respective training and evaluation.

5.1 Overview

Based on the Q4 initially formulated in sub-chapter 1.2:

- **Q4 — What can be done to contribute to optimizing the process of treatment planning for oncological diseases, especially for the liver?**

Interviews and brainstorming sessions were carried out with physicians from the HUC Radiotherapy Department with the aim of better understanding the needs and challenges these physicians face when it comes to planning treatments, with the liver being the organ requested by the HUC to be automatically segmented. This led to the creation of the workflow diagrams shown in Figure 28, illustrating the process from a clinical perspective (clinical workflow) and from a technical perspective (AI workflow), which were presented to the HUC to validate whether this solution could help to optimize their processes.

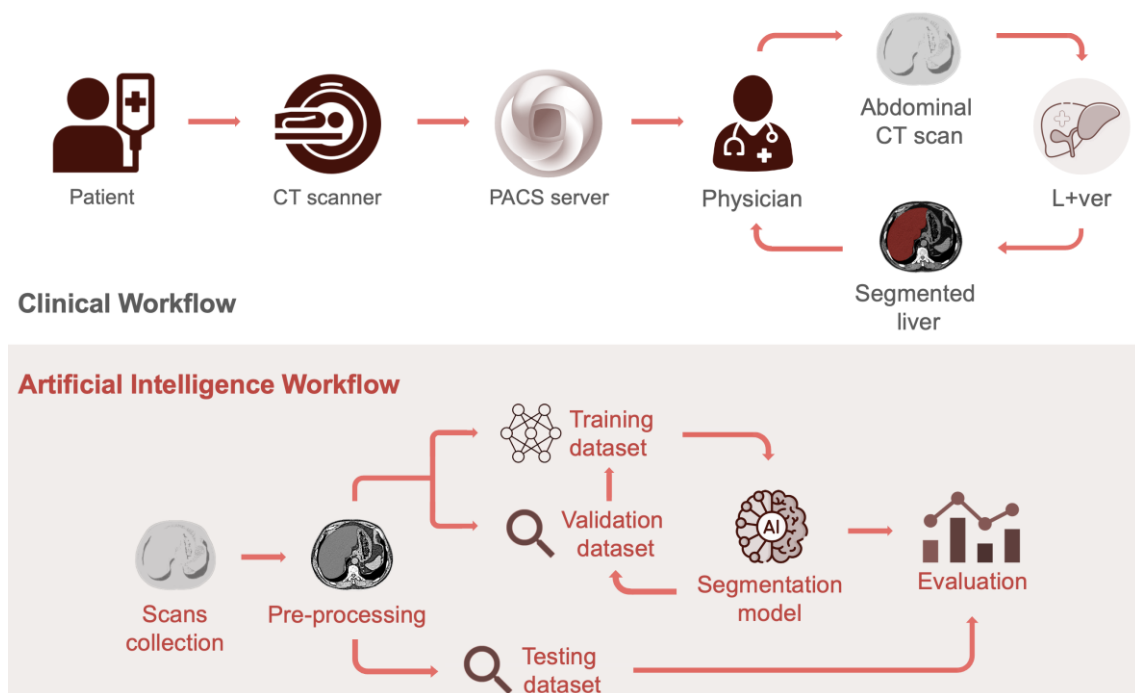


Figure 28. Methodology for the proposed solution.

5.1.1 Clinical Workflow

The clinical workflow is composed of the following elements:

- **Patient:** the process starts with a patient who needs an abdominal CT scan.
- **CT scanner:** the patient undergoes a CT scan to obtain detailed images of the abdominal area.
- **ARIA server:** the CT scan scan is stored on the ARIA server.
- **Physician:** a physician accesses the CT scan images from the ARIA server.
- **Upload CT scan:** the physician uploads the abdominal CT scan for liver identification and segmentation.
- **Liver segmentation:** the L+ver solution automatically segments the liver from the CT images for further analysis or treatment planning.
- **Evaluation and decision:** the segmented liver images are evaluated by the physician to make a treatment decision.

5.1.2 Artificial Intelligence Workflow

The AI workflow is composed of the following steps:

- **Scans collection:** the workflow begins with the collection of abdominal CT scans.
- **Pre-processing:** the collected scans are pre-processed to ensure they are suitable for training an AI model.
- **Training and validation dataset:** the pre-processed scans are divided into a training and a validation dataset. The training dataset is used for the model to learn patterns, relationships, and features. The validation dataset is used to fine-tune the model, helping to prevent overfitting and ensuring it generalizes well to unseen data.
- **Segmentation model:** the AI segmentation model is trained using the augmented training dataset.
- **Testing dataset:** a separate testing dataset is used to validate the effectiveness of the model.
- **Final evaluation:** the performance of the AI model is thoroughly evaluated to ensure it meets the required standards for clinical use.

An additional step implemented was post-processing, which aimed to optimize the outcome (see sub-chapter 5.6).

5.2 Experimental Setup

All experimental models were implemented using a NVIDIA® T4 Tensor Core GPU and a NVIDIA® A100 GPU, with Kaggle and Google Colab Pro+ serving as the programming environments, respectively.

5.3 Data Preparation

In this sub-chapter an analysis of the data that constitutes the CT scan datasets for the experiments will be presented, as well as some relevant concepts and important processes for data preparation, which creation process was previously described in chapter 4.

5.3.1 Datasets

To carry out the project, three datasets were used, whose comparison can be found in Table 5.1, highlighting the extensive volume of the datasets and their main characteristics.

- Private dataset created with HUC CT scans.
- 3D-IRCADb01 [91].
- CHAOS [92].

Table 5.1. Comparison between the datasets used.

Dataset	Format	Number of scans	Number of slices per scan	Total number of slices	Resolution
Private dataset	DICOM	246	20 to 132	11175	512x512
3D-IRCADb01 [91]	DICOM	20	74 to 260	2843	512x512
CHAOS [92]	DICOM	40	77 to 105	2915	512x512

Several combinations of these datasets were used in the experimental part of this project. Table 5.2 present the distribution and structure of each dataset used to train, validate and test the models.

Table 5.2. Datasets structure based on number of CT scans.

ID	Training and validation datasets (CT scans)			Testing dataset (CT scans)
D1	Private (200)			Private (46)
D2	3D-IRCADb01 (20)	CHAOS (20)		
D3	3D-IRCADb01 (20)	CHAOS (20)	Private (50)	
D4			Private (100)	
D5			Private (200)	

Given that the number of slices in each image differs according to the CT scan, Table 5.3 shows the distribution and structure of each dataset based on the total number of slices used to train, validate and test the models.

From the initial 5758 slices (images) that were included in the public datasets, only 4414 of these could be used, since the rest represented images in which the liver was not present, so the mask was blank. Therefore, the distribution of slices for the public datasets is also presented in Table 5.3.

Table 5.3. Datasets structure based on number of CT scans.

ID	Training and validation datasets (slices)			Testing dataset (slices)
D1	Private (9261)			Private (1914)
D2	3D-IRCADb01 (2073)	CHAOS (2341)		
D3	3D-IRCADb01 (2073)	CHAOS (2341)	Private (2445)	
D4			Private (4623)	
D5			Private (9261)	

5.3.1.1 Private Dataset Analysis

To analyze the private dataset, the tables below offer a detailed overview of the main characteristics of the dataset, emphasizing the scope of segmented scans, as well as the demographic profile of the patients. Table 5.4 presents descriptive statistics for the ages of the participants. These tables collectively provide a comprehensive snapshot of the breadth of the dataset and the age distribution of the patients.

Table 5.4. Demographic profile of the patients from the created dataset.

Mean age (years)	Standard deviation (years)	Minimum age (years)	Median age (years)	Maximum age (years)
71.67	11.91	11	73	99

The histogram in Figure 29 shows a similar concentration of ages around the late 60s to early 80s, with a few outliers, which aligns with the typical demographic for abdominal CT scans, as these procedures are more commonly performed on older individuals due to higher prevalence of related health conditions in this age group.

2D Liver Segmentation in Abdominal Computed Tomography Scans

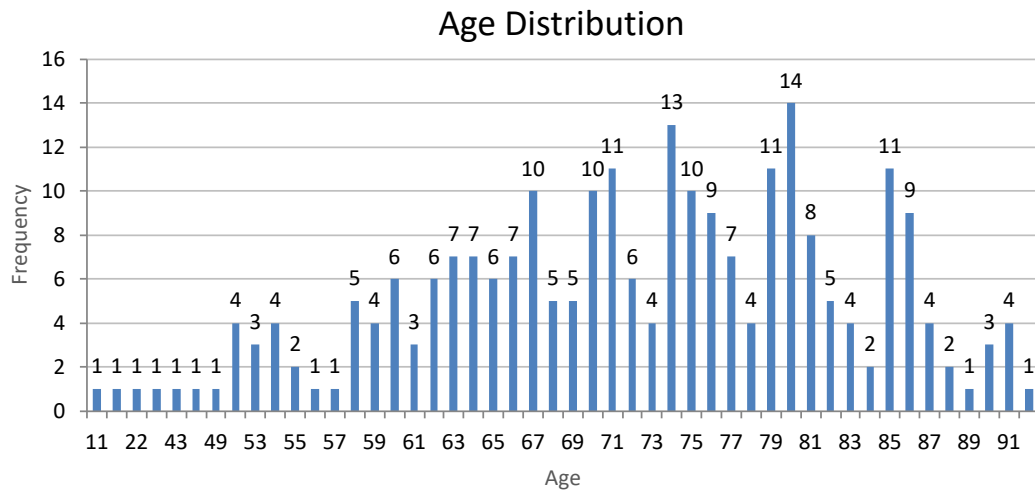


Figure 29. Age distribution of the created dataset.

The gender distribution presented in Figure 30 indicates a higher count of male patients compared to female patients.

Gender Distribution

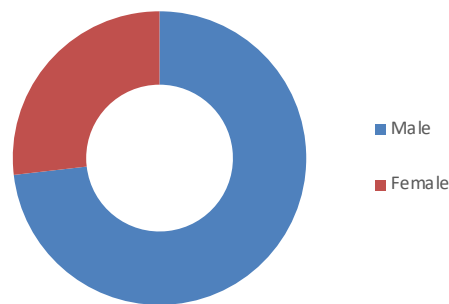


Figure 30. Gender distribution of the created dataset.

The bar chart for scan year in Figure 31 shows the frequency of scans conducted on different years.

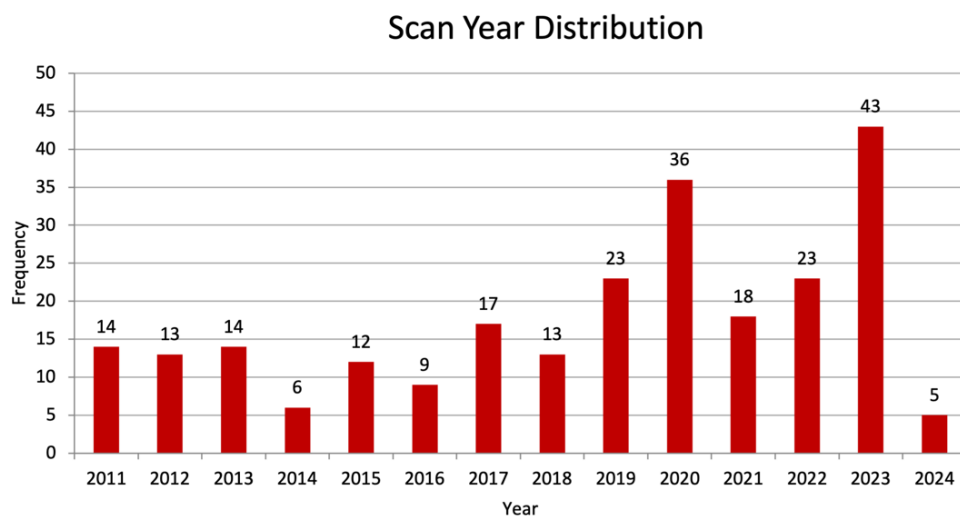


Figure 31. Scan year distribution of the created dataset.

5.3.1.2 3D-IRCADb01 Dataset

To analyze the 3D-IRCADb01 public dataset, an overview of the main characteristics of the dataset are presented below [91]. This dataset is composed of the 3D CT-scans of 20 patients with hepatic tumors in 75% of cases, as can be seen in the sample of the dataset shown in Figure 32.

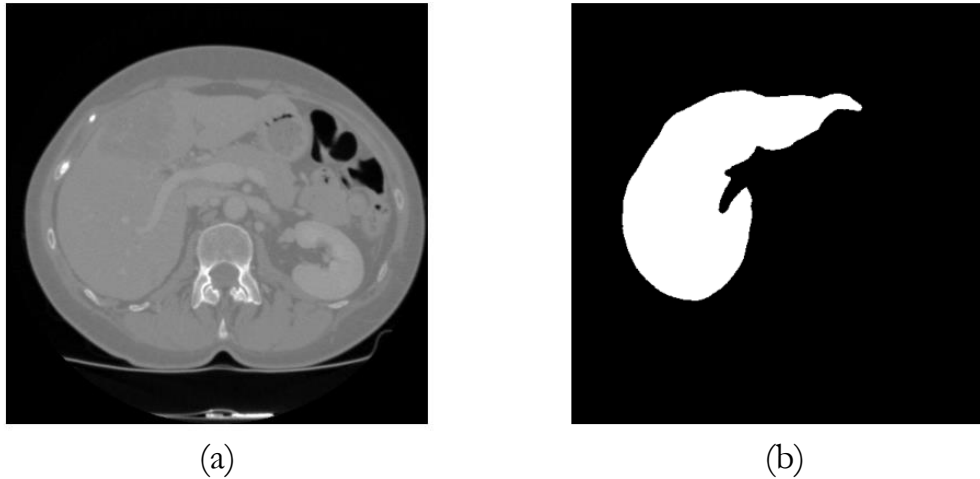


Figure 32. Example of a CT scan from the 3D-IRCADb01 dataset. (a) CT scan image/slice. (b) Liver segmentation ground truth.

The histogram in Figure 33 shows a balance between the proportion of patients aged between 37 and 95 years old, but due to the reduced dimension of the dataset, there were no notable variations, with the most frequent age being 54 years old.

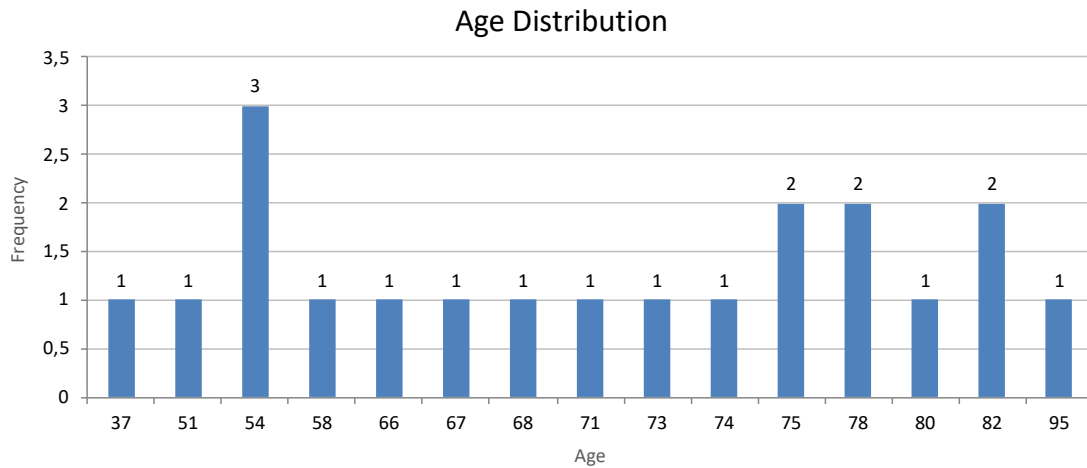


Figure 33. Age distribution of the 3D-IRCADb01 public dataset.

The gender distribution presented in Figure 34 indicates a balance between the number of male patients and female patients.

Gender Distribution

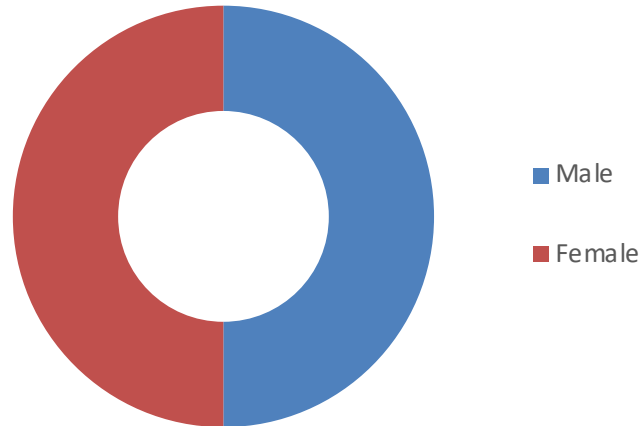


Figure 34. Gender distribution of the 3D-IRCADb01 public dataset.

5.3.1.3 CHAOS Dataset

CHAOS challenge aimed the segmentation of abdominal organs (liver, kidneys and spleen) from CT and MRI scans. CHAOS was held in the IEEE International Symposium on Biomedical Imaging (ISBI) on April 11, 2019, Venice, Italy. For this challenge, a public dataset was presented [92]. This dataset contains 80 medical imaging scans, but only 40 of them are CT scans from healthy abdomen organs without any pathological abnormalities, the 40 remainder consisted of MRI scans. From the 40 CT scans, only 20 were intended by the challenge to be used for training and the remaining 20 for testing, with no ground truth for the latter. Given the methodology defined for implementing this project, it was decided to discard the 20 test CT scans. A sample of a CT scan from the dataset is shown in Figure 35.

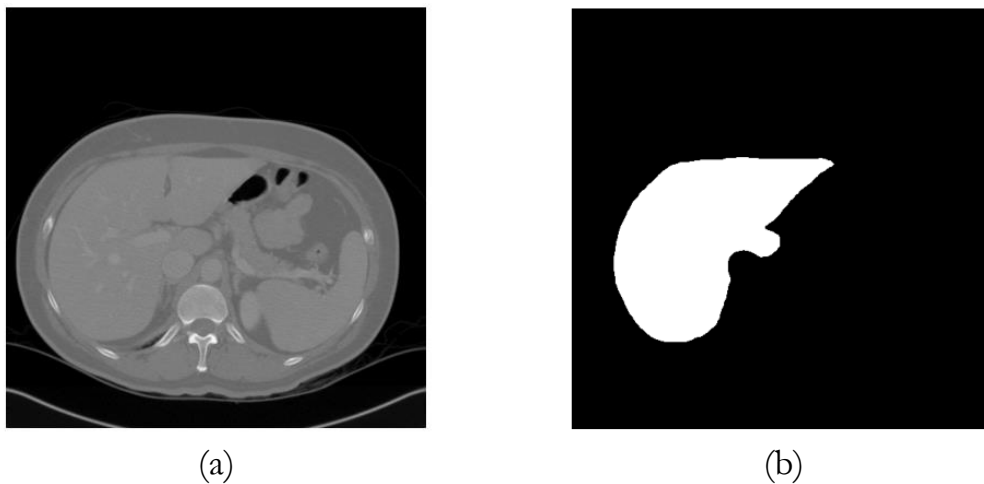


Figure 35. Example of a CT scan from the CHAOS dataset. (a) CT scan image/slice. (b) Liver segmentation ground truth.

Masks in this dataset are in PNG format, unlike the created dataset with the CT scans from HUC, which are in DICOM format, as is also the case with the scans contained in the 3D dataset. As such, these scans have been converted to provide consistency in the format of all the scans.

Regarding age distribution, no detailed information was found, except an indication of the participation of patients with ages between 18 and 63 with an average of around 45 years old.

The gender distribution presented in Figure 36 indicates a balance between the number of male patients and female patients, there was a slight predominance of males, comprising 22 men and 18 women.

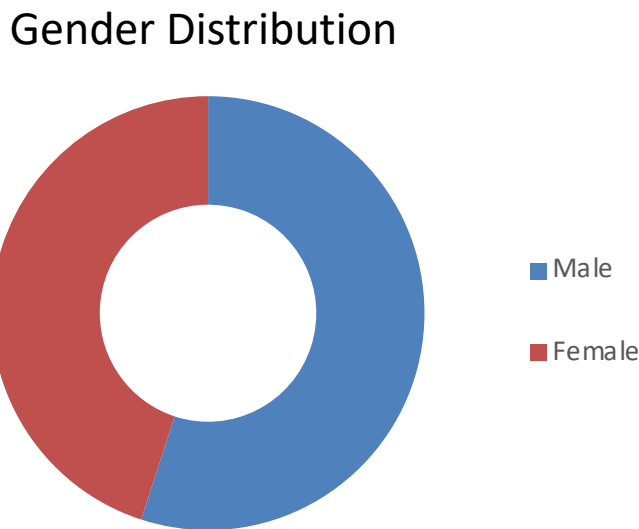


Figure 36. Gender distribution of the CHAOS public dataset.

5.3.2 Relevant Concepts

In order to perform the analysis and preparation of the private dataset, it was important to understand essential terms such as Radiotherapy Structure Set (RTSS), Instance Number and SOP Instance UID.

5.3.2.1 Radiotherapy Structure Set

The RTSS is a specialized type of DICOM object that contains information about the anatomical and clinical structures of a patient, typically derived from CT scans. These structures are essential for planning and delivering precise radiotherapy treatments by providing a comprehensive and detailed map of the anatomy of the patient, which is used to plan the delivery of radiation doses to target tumors while minimizing exposure to surrounding healthy tissues. RTSS includes the definition of various anatomical structures, such as organs at risk, target volumes, and other critical structures.

5.3.2.2 Instance Number

The Instance Number (IS) (0020, 0013) tag in DICOM metadata represents a unique identifier for a specific slice/image within a series. Each slice in a CT scan series is assigned an instance number to maintain its order and identification within the series.

The IS is assigned during the image acquisition stage. It is meant to indicate the sequential order of each slice relative to the others in the same scan. Ideally, the IS starts at 1 for the first slice and increments by one for each subsequent slice.

5.3.2.3 SOP Instance UID

The SOP Instance UID (0008, 0018) tag is a unique identifier for a specific instance of a DICOM object. SOP (Service-Object Pair) Instance UID (Unique Identifier) ensures that each DICOM object (such as a slice, a report, or an RTSS) can be uniquely identified and retrieved across different systems and databases.

5.3.3 Sort Slices

During a CT scan, multiple cross-sectional slices of the body of the patient are captured. Each slice is stored as a separate DICOM file, which contains metadata along with the image data. The metadata includes various attributes that describe the image. One such attribute is the IS, which is intended to uniquely identify each image within a series, as previously mentioned.

At the beginning, the slices of the CT scans from the private dataset were sorted by the name of the slices, however, after visualization, it was noted that the order of the slices was not appropriate. To optimize the presentation of the slices, the sorting process begins by reading the IS from the DICOM metadata of each slice.

The slices are then arranged in ascending order based on their IS value, guaranteeing that the slices are in the correct anatomical sequence from one end of the scanned region to the other.

5.3.4 Masks Creation

The process of creating precise masks based on RTSS contours for each CT scan slice is essential for the segmentation of structures for the private dataset. This is possible by using its corresponding RTSS contour begins with the collection and pairing of DICOM files, including both the CT scan slices and the RTSS file. The CT slices capture the anatomical details, while the RTSS contains contours that delineate the ROI, as previously explained. Each CT slice is identified by a unique SOP Instance UID and each RTSS contour is linked to a specific slice via the Referenced SOP Instance UID, ensuring each contour accurately corresponds to the correct slice. This matching is critical for

maintaining the anatomical accuracy of the overlays. A concise representation of this process is shown in Figure 37.

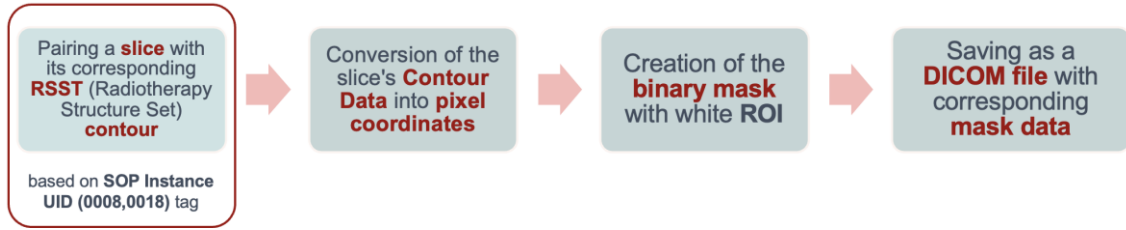


Figure 37. Masks creation process.

Once the contours are accurately associated with their corresponding CT slices, the next step involves converting the contour data of the slice into pixel coordinates suitable for the creation of binary masks. This involves translating the contour coordinates, which are typically specified in the coordinate system of the image, into the corresponding pixel indices on the CT slice.

This conversion is relevant for precise overlay as it maps the abstract contour data directly onto the tangible pixel grid of the CT scan. After conversion, these coordinates are used to create binary masks, where pixels inside the contour are set to 1 (indicating the ROI), and pixels outside are set to 0, with an example of the outcome in Figure 38.

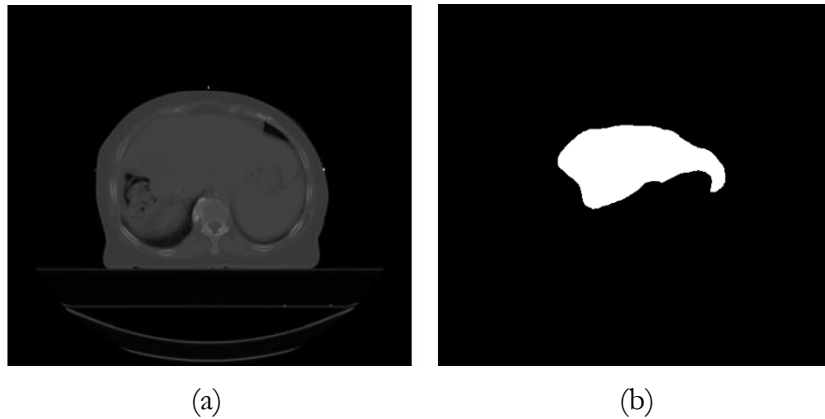


Figure 38. Example of mask creation. (a) Original image before pre-processing techniques. (b) Ground truth mask created.

5.4 Pre-processing

Pre-processing in medical imaging is a crucial step to prepare images for further analysis. This phase involved various techniques to enhance the quality and usability of the images. In this chapter, the pre-processing techniques employed will be explored: application of a Hounsfield window, histogram equalization and data normalization. Figure 39 represents the steps for the pre-processing stage.

2D Liver Segmentation in Abdominal Computed Tomography Scans



Figure 39. Pre-processing flowchart.

5.4.1 Hounsfield Windowing

Hounsfield windowing, also known as window-leveling or contrast stretching, is a technique used in CT imaging to enhance the visibility of specific tissues or structures. By focusing on a particular range of HU values, clinicians can better visualize ROIs [131][132].

To apply Hounsfield windowing, it was first defined the window parameters, the window center, which was the midpoint of the range, and the window width, which determined the span of HU values displayed. Pixels within this range were shown with varying intensities, while those outside the range were rendered fully black or white. Specifically, using a window range of -100 to 400 HU was effective for visualizing soft tissues and some bony structures. This interval was proposed by Sabir, et. al. [80] for their liver segmentation proposal, with the aim of optimizing the correct visualization not only of the liver but also of other structures. This range included most organs, muscles, fat, and blood, enhanced lung details at the lower end (-100 HU), and captured less dense bones at the upper end (400 HU) (Figure 40).

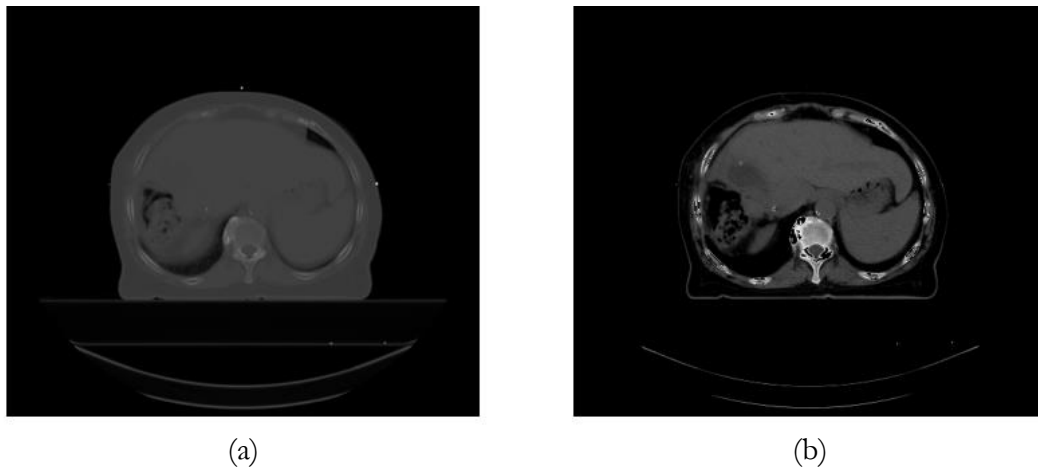


Figure 40. Hounsfield window application. (a) Original image. (b) Image with Hounsfield windowing application.

5.4.2 Histogram Equalization

Histogram equalization is a technique to improve image contrast by spreading out the most frequent intensity values. This process makes hidden features more visible, enhancing the quality of images [133]-[134].

To apply histogram equalization, it was necessary to redistribute pixel values so the histogram of the output image became approximately flat, thereby enhancing contrast by improving areas with lower contrast without notably altering high-contrast areas. This technique proved particularly useful for images with poor contrast, often dominated by bright or dark regions, and in medical imaging, it helped to make important structures within tissues or organs more visible, as shown in Figure 41.

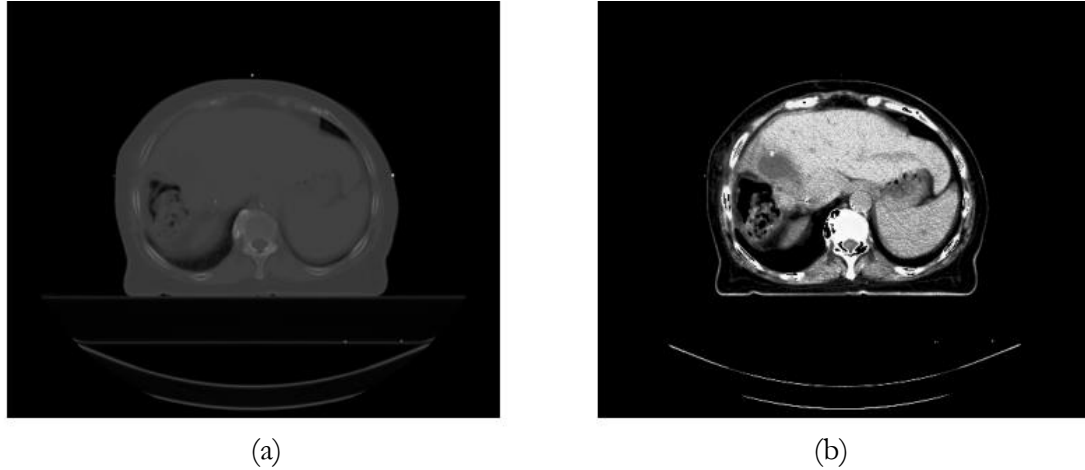


Figure 41. Hounsfield windowing and histogram equalization application. (a) Original image. (b) Outcome of applying the Hounsfield window followed by histogram equalization.

5.4.3 Data Normalization

Data normalization involves scaling the numerical data features to a uniform range. This process ensures that all features have similar scales, facilitating the ability of the segmentation algorithms to process the data effectively and produce accurate segmentation results. By transforming each feature so they fall within a common range, such as 0 to 1, normalization prevents any single feature from dominating the learning process due to its larger scale.

This uniformity is crucial for the ANNs to treat all features equally, leading to improved performance and more precise liver segmentation in the CT scans.

5.5 Segmentation Models

This sub-chapter describes the algorithms used for the project development, as shown in Figure 42.



Figure 42. Segmentation flowchart.

The chosen models for the implementation of the present project are U-Net, ResUNet++, FCN and BerDiff. The selection of these models has been supported based on an analysis of the latest and most impactful advances in medical image segmentation.

The FCN model was one of the first models to demonstrate the viability of using DL for dense prediction tasks like segmentation, being introduced in 2015. Despite being an older architecture, FCN remains relevant due to its simplicity and effectiveness, particularly as a benchmark in segmentation tasks.

The U-Net model, also introduced in 2015, marked a breakthrough in the field of medical imaging. This approach has been shown to be effective for segmentation tasks, providing better outcomes compared to previous methods.

The ResUNet++ model, introduced in 2019, presented the ability to capture and process detailed spatial and contextual information, which design ensures that it can handle the complex and varying structures found in medical imaging.

Most recently, in 2023, the BerDiff model emerged as an innovative adaptation of U-Net, presenting promising results. BerDiff incorporates recent advances in DL architectures, optimizing segmentation effectiveness and performance.

5.5.1 U-Net

U-Net is structured in a U-shape to efficiently use computing resources while balancing the depth of the network. The architecture consists of a contracting path, also known as the encoder, to capture the context of the image, and a symmetric expanding path, known as the decoder, for precise localization. The encoder path includes convolutional layers with ReLU activation and max-pooling layers to downsample the input image, capturing the context in lower resolutions. The decoder path, on the other hand, consists of up-convolution layers to upsample the features, restoring the spatial resolution and enabling the delineation of the liver boundaries [119].

Figure 43 presents a diagram of the U-Net architecture designed for medical image segmentation. This visual representation illustrates the U-shaped structure of the network, showcasing the contracting and expanding paths with respective components such as convolutions, max pooling, ReLU activations, up-convolutions, and skip connections, which can be used for a clearer understanding of the layout and function of the model.

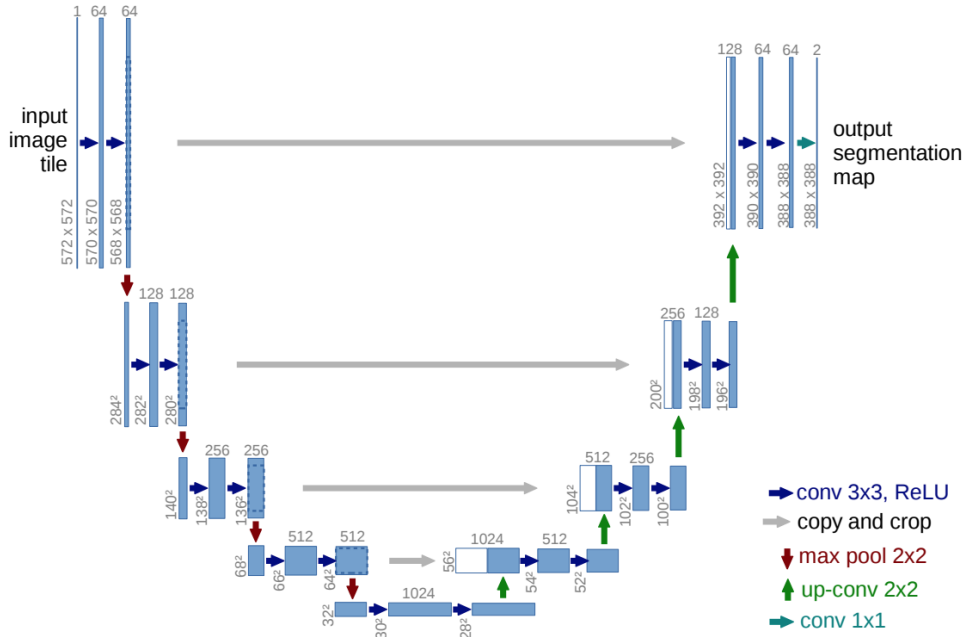


Figure 43. U-Net model architecture. Source: [119].

5.5.2 ResUNet++

ResUNet++ is a CNN which architecture builds on the U-Net and ResUNet models, incorporating residual blocks, squeeze and excitation blocks, an Atrous Spatial Pyramid Pooling (ASPP) bridge, which collectively enhance the ability of the model to capture and process detailed spatial and contextual information, and attention mechanisms to enhance the accuracy of pixel-wise segmentation [136].

The architecture of ResUNet++ consists of three main parts (Figure 44): Encoding, Bridge, and Decoding. In the Encoding phase, the model processes the input through a series of Conv2D layers (3x3 filters), batch normalization, and ReLU activations. Residual connections and squeeze-and-excite blocks enhance feature representation and channel-wise feature recalibration. The Bridge section employs an ASPP layer to capture multi-scale contextual information. The Decoding phase involves a sequence of Conv2D layers, batch normalization, ReLU activations, and attention mechanisms. Upsampling layers are used to restore the spatial resolution, and concatenation merges features from corresponding encoding stages. The final output is obtained through a sigmoid activation, indicating the segmentation map.

2D Liver Segmentation in Abdominal Computed Tomography Scans

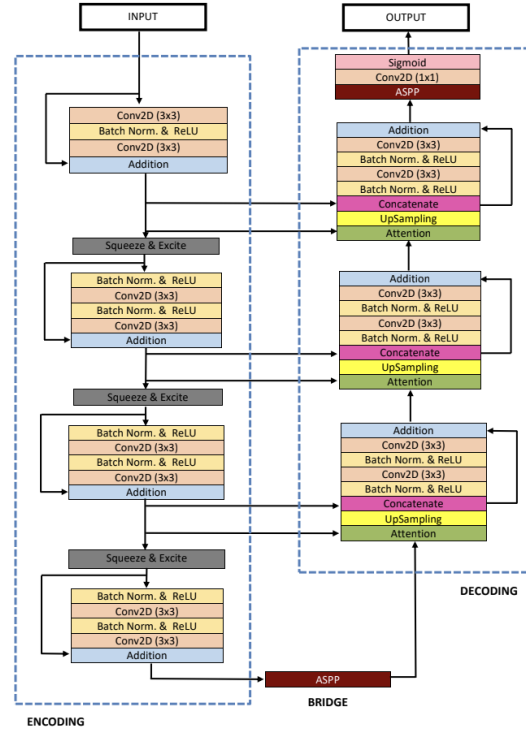


Figure 44. ResUNet++ model architecture. Source: [136].

5.5.3 FCN

FCN is a CNN model designed for semantic segmentation, commonly used in computer vision to perform pixel-wise prediction tasks (classify each pixel of an image into a specific category). Unlike traditional CNNs that use fully connected layers for classification, FCN replace these with convolutional layers, allowing the model to process input images of any size and produce output maps of corresponding dimensions. This architecture allows FCN to make predictions for every pixel in the input image, transforming the network from a classifier to a semantic segmentation model [137].

The architecture of FCN, as illustrated in Figure 45, consists of a series of convolutional and pooling layers that progressively reduce the spatial dimensions of the feature maps while increasing their depth. In the network's decoder, deconvolutional layers (also known as transposed convolutions) are used to upsample the coarse, semantically rich features back to the original image resolution. To recover fine-grained details lost during downsampling, FCN introduce skip connections that combine deep, coarse feature maps with shallow, fine feature maps.

This skip architecture, depicted as FCN-8s in the diagram, combines predictions from multiple layers (pool3, pool4, and conv7) at different strides, allowing the network to make accurate and detailed segmentations by leveraging both high-level semantic information and low-level appearance information.

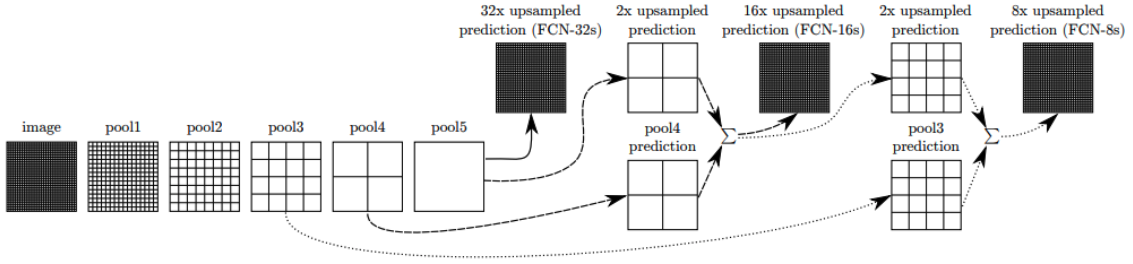


Figure 45. FCN model architecture. Source: [137].

5.5.4 BerDiff

BerDiff is a conditional Bernoulli diffusion model developed for medical image segmentation, addressing challenges such as inherent ambiguity and high uncertainty in medical images. Unlike traditional diffusion models that use Gaussian noise, BerDiff employs Bernoulli noise as the diffusion kernel, enhancing the ability of the model to produce accurate binary segmentation masks. The architecture of BerDiff includes a forward diffusion process, where Bernoulli noise is progressively added, and a reverse process that iteratively refines the segmentation masks. This approach allows BerDiff to generate multiple diverse segmentation masks from the same input image, providing valuable references for radioncologists. Additionally, BerDiff incorporates efficient sampling techniques, inspired by denoising diffusion probabilistic model and denoising diffusion implicit models, to speed up the segmentation process [138].

Figure 46 illustrates the architecture of BerDiff, beginning with the input images, from which Bernoulli noise is sampled and progressively added in the forward process to generate latent variables. These latent variables are then fed into a U-Net for Bernoulli noise estimation. The reverse process iterates through these latent variables, applying a calibration function to refine the noise and produce diverse segmentation masks. Finally, these masks are combined to create a saliency segmentation mask that highlights the most consistent regions, providing valuable references for medical practitioners.

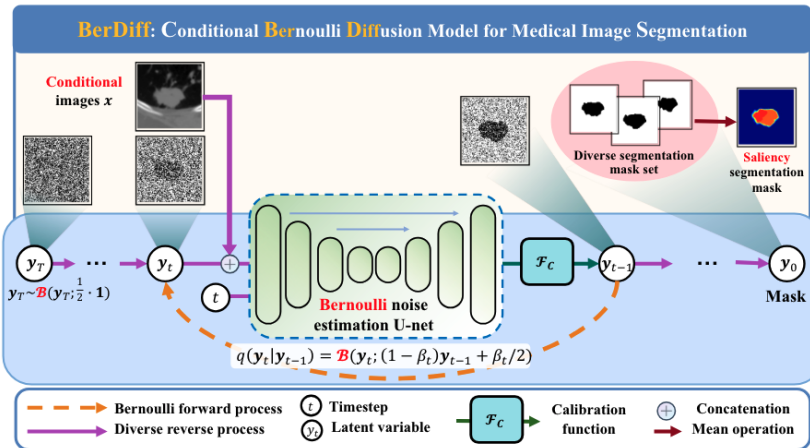


Figure 46. BerDiff model architecture. Source: [138].

5.6 Post-processing

In medical imaging, post-processing techniques are helpful procedures for analyzing model-generated images in order to enhance accuracy and overcome limitations. These techniques are crucial for objectives as improving image quality, reducing noise, increasing contrast, registering images and extracting features. Figure 47 represents the steps for the post-processing stage.

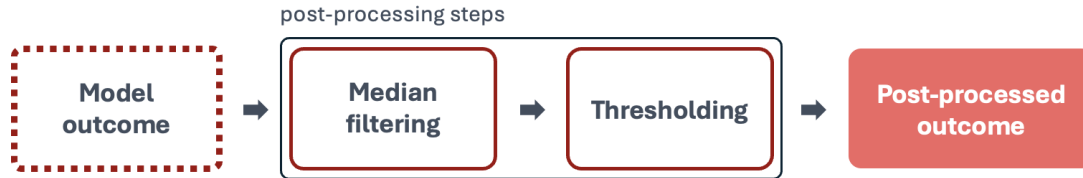


Figure 47. Post-processing flowchart.

Figure 48 presents an example of the results obtained after applying the post-processing steps defined to improve the segmentation prediction.

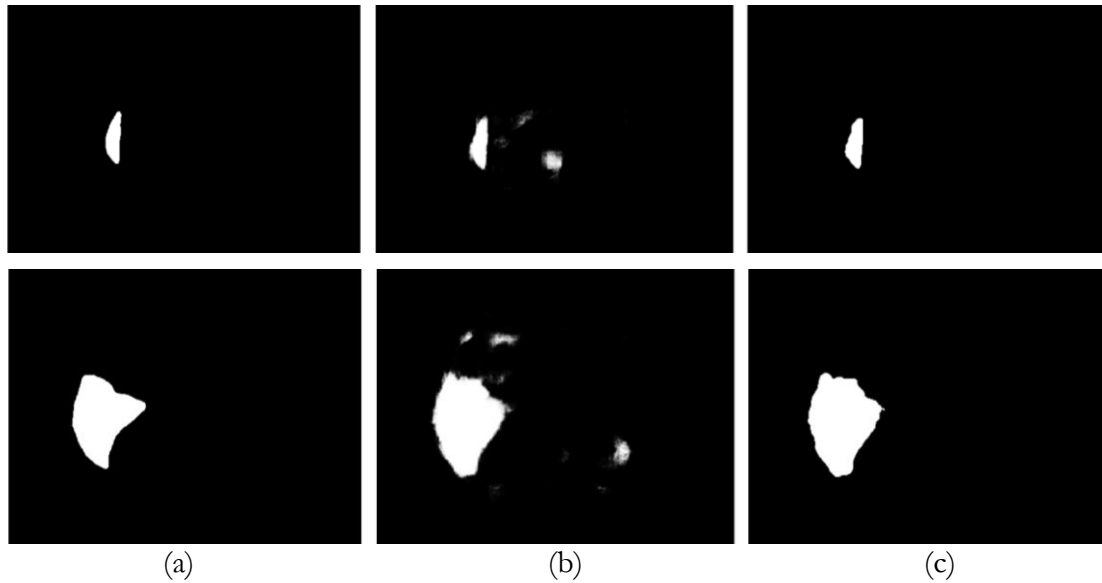


Figure 48. Post-processing comparison for two images. (a) Ground truth mask. (b) Predicted segmentation. (c) Post-processed segmentation.

5.6.1 Median Filtering

Median filtering is used to reduce noise in the images. It replaces the value of each pixel with the median value of the neighboring pixels. This technique is effective for removing salt-and-pepper noise while preserving edges. For this purpose, it was applied using `cv2.medianBlur()` with a kernel size of 5.

5.6.2 Thresholding

Thresholding converts a grayscale image to a binary image. Pixels with intensity above a certain threshold are set to the maximum value (white), and pixels below are set to the minimum value (black). This was implemented using

`cv2.threshold()`. The threshold value defined was 0.5, which determines the cut-off point for binarizing the grayscale image. This means any pixel value greater than 0.5 will be set to 1 (white), and any pixel value less than or equal to 0.5 will be set to 0 (black), as shown in Equation (8).

$$S(x) = \begin{cases} 1 & \text{if } x \in \text{segmented liver} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

5.7 Experimental Specifications

As mentioned at the beginning of the dissertation, new objectives were added to the project based on real-world challenges in medical imaging, where the capacity for adaptability of the models is evaluated using varied data sources and conditions, defining this experimental setup as *Private Data Incrementalization*. These experiments aim to simulate the gradual exposure of models to new, diverse data, a scenario common in clinical settings, providing advantages such as enhancement of the adaptability of the models, improvement of the training efficiency and assessment of the generalization capabilities.

The methodology for the experiments involves systematically introducing private data to a dataset with two public datasets (previously described in sub-chapter 5.3.1) into the training and validation processes, carefully monitoring changes in model performance across new data insertion. The experiments seek to evaluate and enhance the robustness and versatility of the segmentation models mentioned in sub-chapter 5.5.

Initially, all the experiments were intended to preserve the resolution at which the images were originally exported from the ARIA database, 512x512 pixels, however, for computational performance considerations, the images were resized to 256x256 pixels.

Table 5.5 shows the model and respective dataset used in each experiment.

Table 5.5. Configuration of the performed experiments.

Experiment	Model	Dataset
E1	U-Net	D1
E2		D2
E3		D3
E4		D4
E5		D5
E6	ResUNet++	D1
E7		D2

2D Liver Segmentation in Abdominal Computed Tomography Scans

Experiment	Model	Dataset
E8		D3
E9		D4
E10		D5
E11	FCN	D1
E12		D2
E13		D3
E14		D4
E15		D5
E16	BerDiff	D1
E17		D2
E18		D3
E19		D4
E20		D5

5.8 Models Configuration and Hyperparameters

For all experiments, the configuration included training the models for a maximum of 100 epochs, with a learning rate set to $1e-4$. The training process was monitored using an early stopping mechanism based on the validation loss to prevent overfitting. Early stopping was triggered if there was no improvement in the validation loss for 5 consecutive epochs. The batch size for training was set to 8, balancing computational efficiency and model performance.

5.8.1 U-Net

The U-Net architecture, with a total of 1,940,817 parameters, consists of a total of 9 Conv2D layers, 4 MaxPooling2D layers, 4 Conv2DTranspose layers, and 4 Dropout layers distributed across both the contracting and expansive paths. In the contracting path, features are extracted through sequential convolution operations followed by max-pooling for downsampling, while in the expansive path, transpose convolutions are used for upsampling and concatenation layers to merge information from the contracting path, preserving spatial context. The final layer uses a Conv2D operation with a sigmoid activation for binary segmentation.

5.8.2 ResUNet++

The ResUNet++ model, with a total of 4,062,673 parameters, used includes a Stem Block for initial feature extraction, followed by 4 ResNet-based blocks for deep feature learning and downsampling. At the bottleneck, an ASPP layer was incorporated. This ASPP layer uses multiple parallel dilated convolutions with different rates to capture features at various scales, enhancing the ability of the model to handle objects of different sizes. The decoding path consists of 3 attention-based Decoder Blocks, which refine the upsampled features using attention mechanisms and skip connections from earlier layers. The model concludes with an additional ASPP layer and a final convolutional layer for generating the segmentation output.

5.8.3 FCN

The FCN architecture employed, with a total of 134,264,538 parameters, is based on a VGG-like encoder-decoder structure. The encoder consists of 5 convolutional blocks, each followed by a MaxPooling2D layer, progressively reducing the spatial dimensions of the input while increasing the depth of feature maps. The final layers of the encoder are fully convolutional, with two Conv2D layers using large filters (4096 units) to capture more abstract features. In the decoder, upsampling is performed through Conv2DTranspose layers, gradually restoring the original image size. Skip connections from the encoder's corresponding layers are added at each upsampling stage to retain fine spatial details.

5.8.4 BerDiff

When training the BerDiff original model, with a total of 1,869,489 parameters, a RuntimeError: CUDA out of memory (as shown in Figure 49) was obtained, indicating that the GPU lacked sufficient memory for the task. In DL, where models and datasets are very large, efficient memory management is crucial.

```
RuntimeError: CUDA out of memory. Tried to allocate 576.00 MiB
(GPU 0; 39.42 GiB total capacity; 20.89 GiB already allocated;
403.69 MiB free; 20.90 GiB reserved in total by PyTorch)
If reserved memory is >> allocated memory try setting max_split_size_mb
to avoid fragmentation. See documentation for Memory Management and
PYTORCH_CUDA_ALLOC_CONF
```

Figure 49. RuntimeError: CUDA out of memory description.

PyTorch and CUDA facilitate accelerated training and inference but require effective memory management to avoid out-of-memory errors. PYTORCH_CUDA_ALLOC_CONF, a configuration in PyTorch, optimizes GPU memory allocation, enhancing performance by allowing fine-tuning of

caching algorithms and memory pool strategies. This tool supports dynamic memory allocation and recycling, preventing unnecessary memory waste and maximizing GPU resource efficiency. Using `PYTORCH_CUDA_ALLOC_CONF` can reduce memory allocation issues and improve overall training and inference performance.

Despite using `PYTORCH_CUDA_ALLOC_CONF` to optimize memory allocation and having subscribed to Google Colab Pro+ to access better computing resources with the objective of solving the error for the BerDiff model, it continued to occur. As a result, it was necessary to adapt and simplify the BerDiff model to work within the available GPU memory constraints.

The strategies applied to reduce the memory usage and complexity of the BerDiff model were:

- Reduction of the number of down/up sampling layers from 3 to 2.
- Reduction of the number of channels from 64 to 32.
- Use of fewer layers.
- Implementation of gradient checkpointing, which trades computation for memory by recomputing intermediate activations during the backward pass.

Addition of a mixed precision training, which uses lower precision (float16) where possible, reducing memory usage and potentially speeding up training.

The BerDiff modified model, with a total of 577,667 parameters, consists of a downsampling path with convolutional and batch normalization layers for feature extraction, followed by an intermediate section for deeper processing. In the upsampling path, `ConvTranspose2D` layers restore the spatial dimensions, and the final output is generated through a `1x1` convolutional layer. The binomial diffusion process is employed to refine the model's predictions over a series of timesteps, using a linear beta schedule to control the noise added at each step. The diffusion model is configured with epsilon prediction and a binary cross-entropy (BCE) loss function.

5.9 Evaluation

The evaluation metrics used to assess the performance of the models include loss, accuracy, precision, DSC, AOE, ASSD, and MSSD. Each of these metrics provides insights into the performance of the models. However, in a clinical application for liver segmentation in CT scans the emphasis should be placed on the latter four metrics. DSC evaluates the overall overlap accuracy, while AOE assesses precision in boundary delineation. On the other hand, ASSD and MSSD provide detailed insights into surface accuracy and the worst-case errors, respectively.

For the context of this project, these metrics were prioritized when evaluating model performance against accuracy and precision due to the fact that accuracy is highly influenced by class imbalance and can be misleading in scenarios where the liver constitutes a small part of the CT scan, leading to an overestimation of the performance. Consequently, although these two metrics are important, they must be combined with other performance measures to ensure a thorough evaluation, addressing both average performance and critical edge cases, which is essential for effective and meaningful segmentation of the liver in clinical practice.

6 RESULTS AND DISCUSSION

After the algorithms selected for this project had been executed following the proposed methodology, the results obtained by the chosen evaluation metrics were collected. In this sense, this chapter presents the results obtained by the U-Net, ResUNet++, FCN and BerDiff models for the different experiments made.

6.1 Quantitative Results

This sub-chapter presents the quantitative results obtained by the segmentation models for the different experiments.

6.1.1 Models

This sub-chapter compares the results obtained for the experiments for the U-Net, ResUNet++, FCN and BerDiff models on unseen data. The results for the training and validation datasets are available on Appendix M.

6.1.1.1 U-Net

The evaluation results of the five experiments with *Private Data Incrementalization* (E1-E5) using the U-Net model on unseen data are displayed in Table 6.1.

Table 6.1. Performance of the U-Net experiments on unseen data.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E1	0.0729	0.9969	0.9228	0.9182	0.1069	0.0113	4.4714
E2	0.1689	0.9932	0.8310	0.7915	0.2262	0.0150	4.4720
E3	0.0918	0.9962	0.9084	0.8946	0.1315	0.0126	4.4719
E4	0.0841	0.9965	0.9234	0.9015	0.1219	0.0119	4.4716
E5	0.0780	0.9967	0.9151	0.9118	0.1135	0.0118	4.4711

6.1.1.2 ResUNet++

The evaluation results of the five experiments with *Private Data Incrementalization* (E6-E10) using the ResUNet++ model on unseen data are displayed in Table 6.2.

Table 6.2. Performance of the ResUNet++ experiments on unseen data.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E6	0.0799	0.9972	0.9477	0.9179	0.1029	0.0053	4.4721
E7	0.1790	0.9935	0.8737	0.7374	0.2345	0.0060	4.4721
E8	0.0866	0.9967	0.9368	0.8932	0.1185	0.0054	4.4721
E9	0.0772	0.9970	0.9452	0.9000	0.1102	0.0053	4.4721
E10	0.0691	0.9972	0.9218	0.9449	0.0990	0.0053	4.4721

6.1.1.3 FCN

The evaluation results of the five experiments with *Private Data Incrementalization* (E11-E15) using the FCN model on unseen data are displayed in Table 6.3.

Table 6.3. Performance of the FCN experiments on unseen data.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E11	0.1197	0.9954	0.9379	0.8666	0.1647	0.0147	4.4721
E12	0.3155	0.9888	0.8386	0.6933	0.3875	0.0220	4.4721
E13	0.1912	0.9920	0.8415	0.7948	0.2540	0.0195	4.4721
E14	0.1117	0.9956	0.8921	0.8716	0.1529	0.0147	4.4721
E15	0.0920	0.9964	0.9204	0.8956	0.1271	0.0127	4.4721

6.1.1.4 BerDiff

The evaluation results of the five experiments with *Private Data Incrementalization* (E16-E20) using the BerDiff model on unseen data are displayed in Table 6.4.

2D Liver Segmentation in Abdominal Computed Tomography Scans

Table 6.4. Performance of the BerDiff experiments on unseen data.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E16	0.3446	0.9830	0.6443	0.6149	0.4250	0.0195	4.4721
E17	0.8135	0.9761	0.5739	0.2243	0.8465	0.0100	4.4721
E18	0.3743	0.9833	0.6486	0.6174	0.4565	0.0165	4.4721
E19	0.3890	0.9838	0.6833	0.6073	0.4689	0.0159	4.4721
E20	0.3696	0.9854	0.7184	0.6058	0.4469	0.0144	4.4721

6.1.2 Datasets

This sub-chapter compares the results obtained for the experiments for each dataset (1 to 5) training the U-Net, ResUNet++, FCN and BerDiff models. The results for the training and validation datasets are available on Appendix M.

6.1.2.1 Dataset 1

Table 6.5 presents the results of experiments with dataset 1 with unseen data.

Table 6.5. Comparison of the performance from each model on unseen data from the dataset 1.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E1	0.0729	0.9969	0.9228	0.9182	0.1069	0.0113	4.4714
ResUNet++	E6	0.0799	0.9972	0.9477	0.9179	0.1029	0.0053	4.4721
FCN	E11	0.1197	0.9954	0.9379	0.8666	0.1647	0.0147	4.4721
BerDiff	E16	0.3446	0.9830	0.6443	0.6149	0.4250	0.0195	4.4721

6.1.2.2 Dataset 2

Table 6.6 presents the results of experiments with dataset 2 with unseen data.

Table 6.6. Comparison of the performance from each model on unseen data from the dataset 2.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E2	0.1689	0.9932	0.8310	0.7915	0.2262	0.0150	4.4720
ResUNet++	E7	0.1790	0.9935	0.8737	0.7374	0.2345	0.0060	4.4721
FCN	E12	0.3155	0.9888	0.8386	0.6933	0.3875	0.0220	4.4721
BerDiff	E17	0.8135	0.9761	0.5739	0.2243	0.8465	0.0100	4.4721

6.1.2.3 Dataset 3

Table 6.7 shows the results of experiments with dataset 3 with unseen data.

Table 6.7. Comparison of the performance from each model on unseen data from the dataset 3.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E3	0.0918	0.9962	0.9084	0.8946	0.1315	0.0126	4.4719
ResUNet++	E8	0.0866	0.9967	0.9368	0.8932	0.1185	0.0054	4.4721
FCN	E13	0.1912	0.9920	0.8415	0.7948	0.2540	0.0195	4.4721
BerDiff	E18	0.3743	0.9833	0.6486	0.6174	0.4565	0.0165	4.4721

6.1.2.4 Dataset 4

Table 6.8 presents the results of experiments with dataset 4 with unseen data.

Table 6.8. Comparison of the performance from each model on unseen data from the dataset 4.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E4	0.0841	0.9965	0.9234	0.9015	0.1219	0.0119	4.4716
ResUNet++	E9	0.0772	0.9970	0.9452	0.9000	0.1102	0.0053	4.4721
FCN	E14	0.1117	0.9956	0.8921	0.8716	0.1529	0.0147	4.4721
BerDiff	E19	0.3890	0.9838	0.6833	0.6073	0.4689	0.0159	4.4721

6.1.2.5 Dataset 5

Table 6.9 presents the results of experiments with dataset 5 with unseen data.

Table 6.9. Comparison of the performance from each model on unseen data from the dataset 5.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E5	0.0780	0.9967	0.9151	0.9118	0.1135	0.0118	4.4711
ResUNet++	E10	0.0691	0.9972	0.9218	0.9449	0.0990	0.0053	4.4721
FCN	E15	0.0920	0.9964	0.9204	0.8956	0.1271	0.0127	4.4721
BerDiff	E20	0.3696	0.9854	0.7184	0.6058	0.4469	0.0144	4.4721

6.2 Qualitative Results

This sub-chapter presents the qualitative results from the segmentation models for the various experiments.

6.2.1 Models

This sub-chapter presents a comparative analysis of the results obtained in the training experiments for the U-Net, ResUNet++, FCN and BerDiff models, exploring the impact of *Private Data Incrementalization*. Figure 50 to Figure 53 illustrate, respectively, the inferences on unseen data for the series of experiments E1-E5 (U-Net), E6-E10 (ResUNet++), E11-E15 (FCN) and E16-E20 (BerDiff), providing a comprehensive view of the performance of each model with datasets from different sources and varying dataset size.

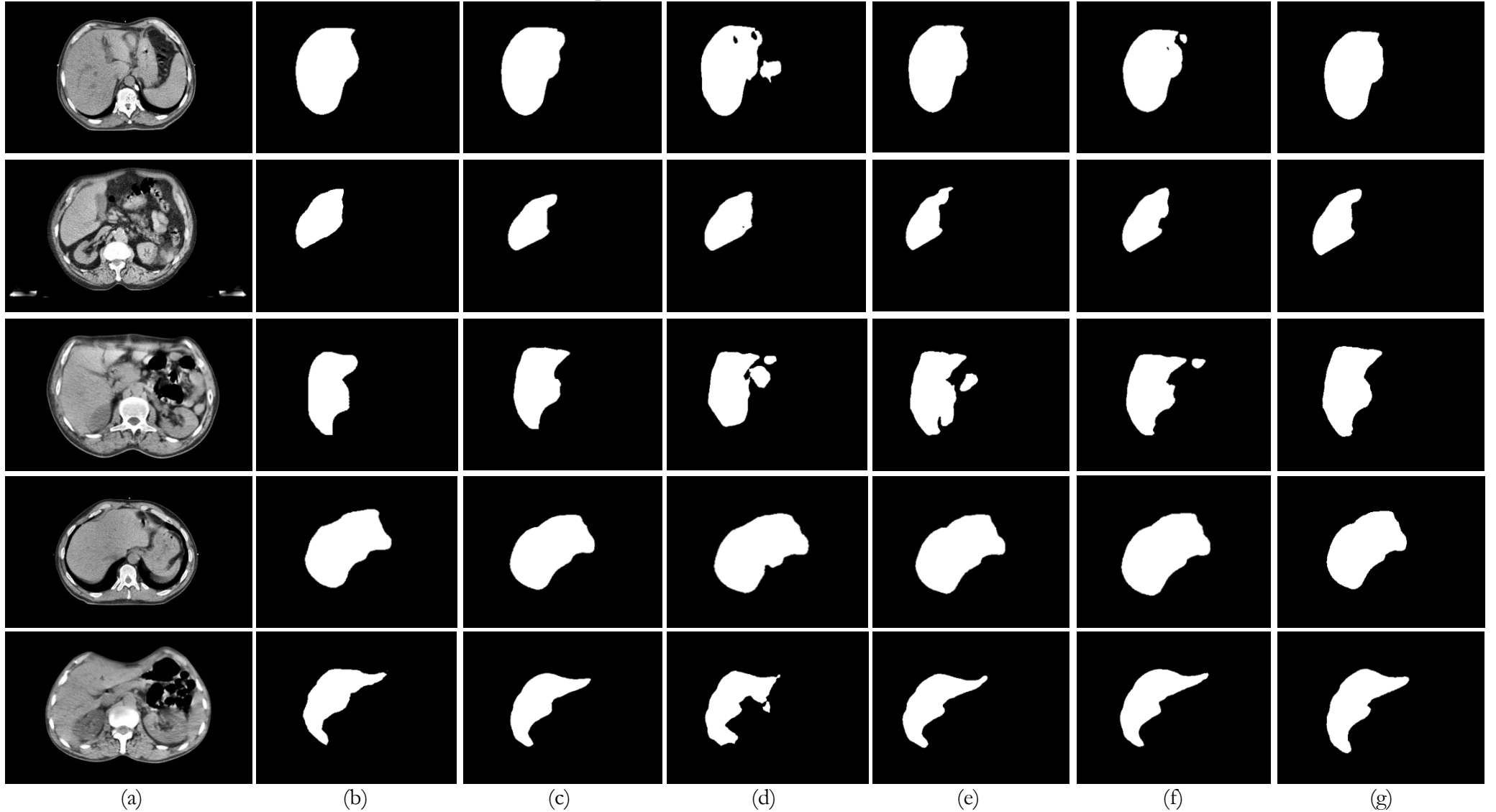


Figure 50. Performance of the U-Net experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E1. (d) Segmentation outcome with E2. (e) Segmentation outcome with E3. (f) Segmentation outcome with E4. (g) Segmentation outcome with E5.

2D Liver Segmentation in Abdominal Computed Tomography Scans

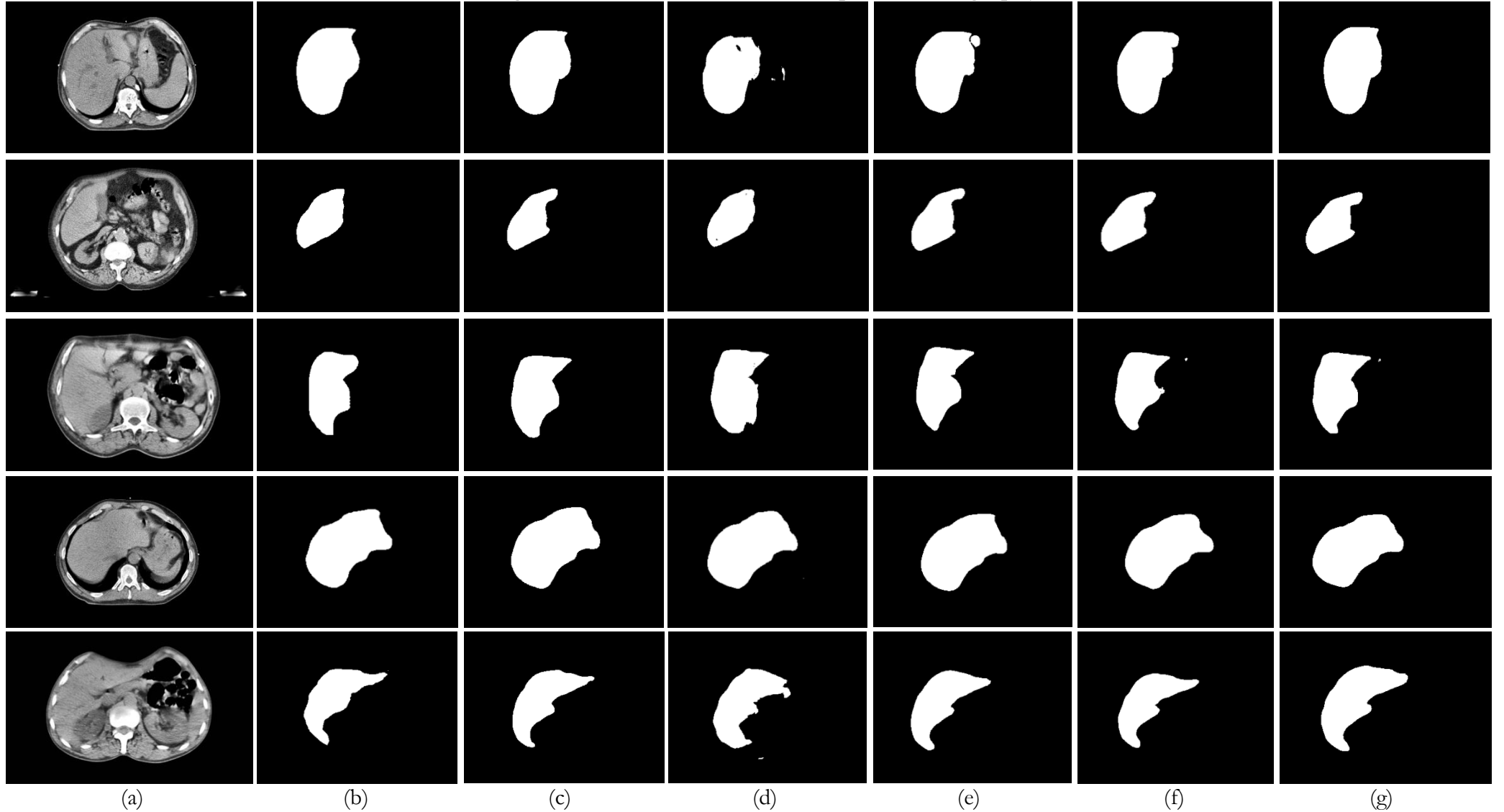


Figure 51. Performance of the ResUNet++ experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E6. (d) Segmentation outcome with E7. (e) Segmentation outcome with E8. (f) Segmentation outcome with E9. (g) Segmentation outcome with E10.

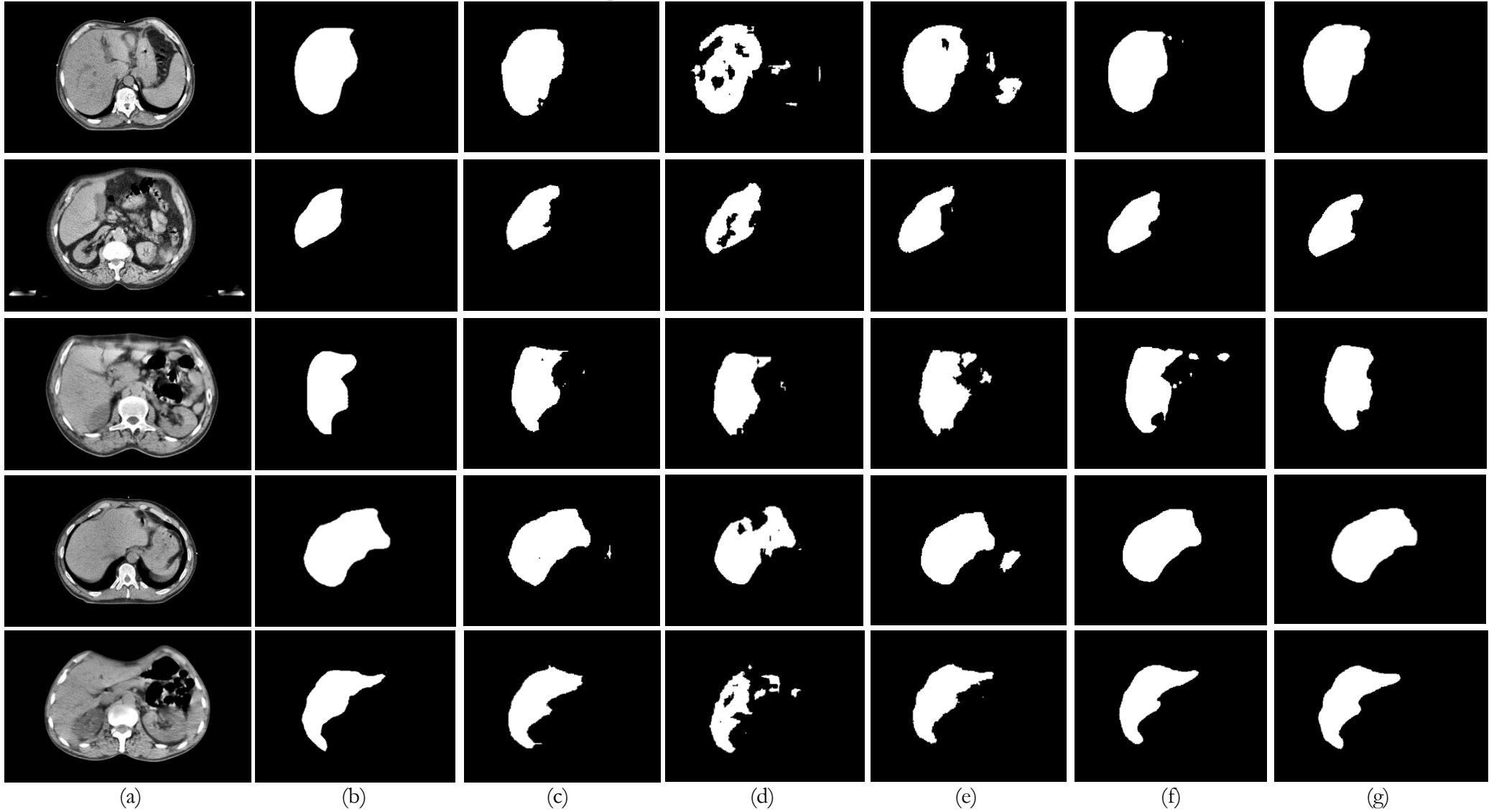


Figure 52. Performance of the FCN experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E11. (d) Segmentation outcome with E12. (e) Segmentation outcome with E13. (f) Segmentation outcome with E14. (g) Segmentation outcome with E15.

2D Liver Segmentation in Abdominal Computed Tomography Scans

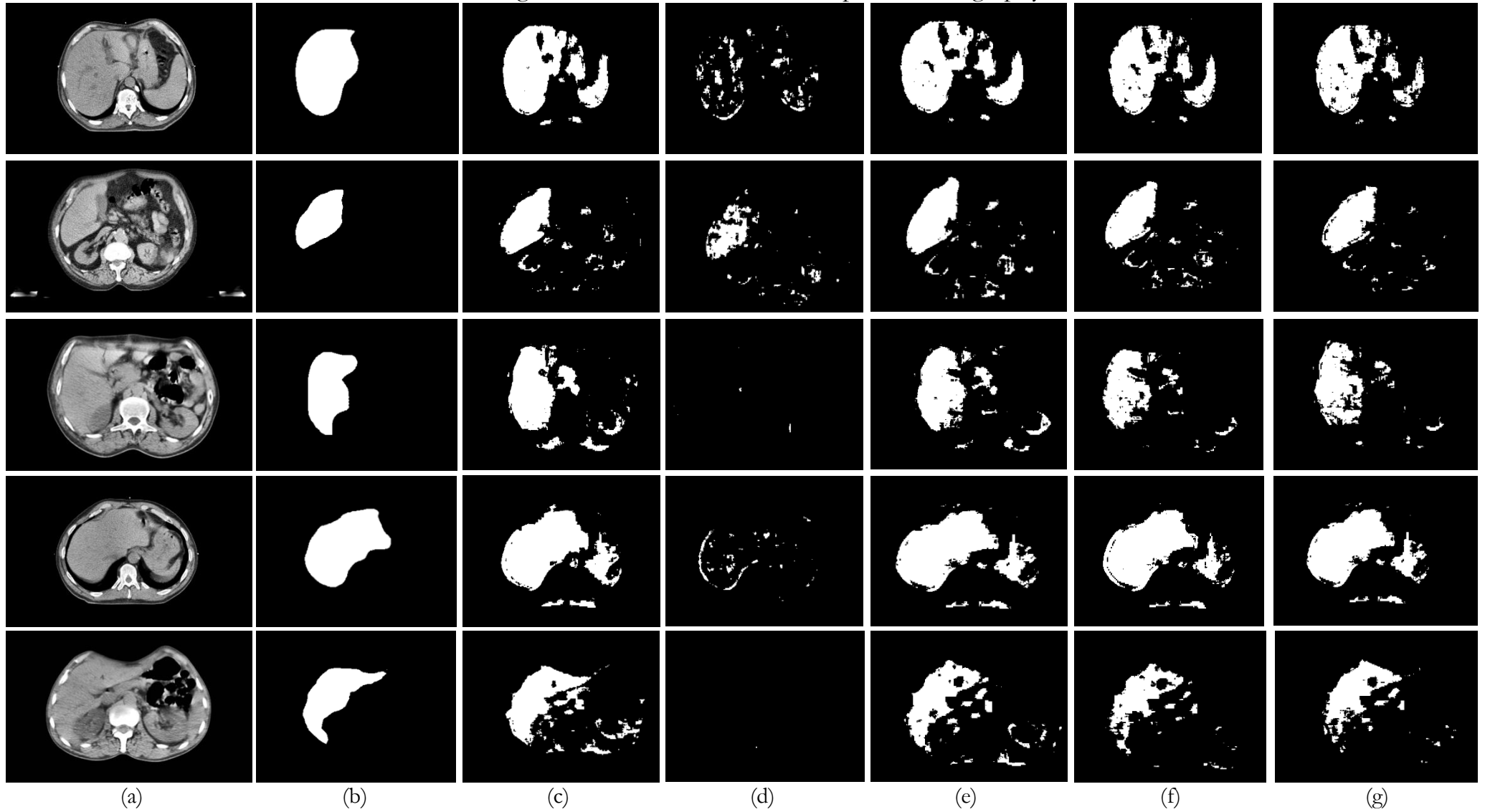


Figure 53. Performance of the BerDiff experiments on unseen data. (a) Original image. (b) Ground truth. (c) Segmentation outcome with E16. (d) Segmentation outcome with E17. (e) Segmentation outcome with E18. (f) Segmentation outcome with E19. (g) Segmentation outcome with E20.

6.2.2 Datasets

This sub-chapter presents a comparison of the results obtained for the experiments for each dataset (1 to 5) training the U-Net, ResUNet++, FCN and BerDiff models. A general overview of the values for each metric during the validation of each model across all datasets on unseen data can be seen from Figure 54 to Figure 60.

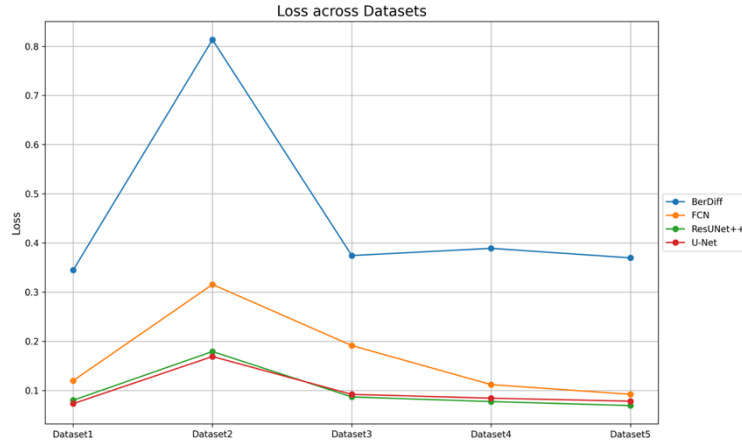


Figure 54. Loss values for each model across all datasets on unseen data.

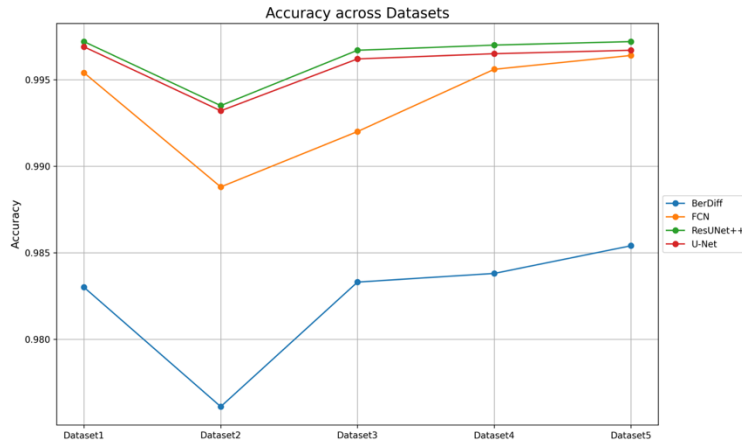


Figure 55. Accuracy values for each model across all datasets on unseen data.

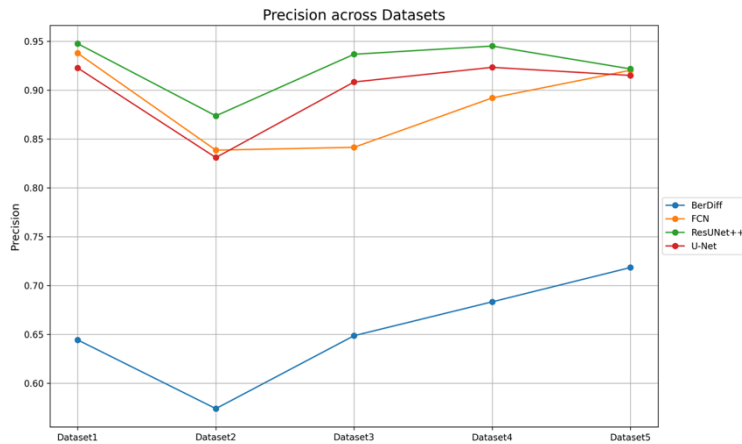


Figure 56. Precision values for each model across all datasets on unseen data.

2D Liver Segmentation in Abdominal Computed Tomography Scans

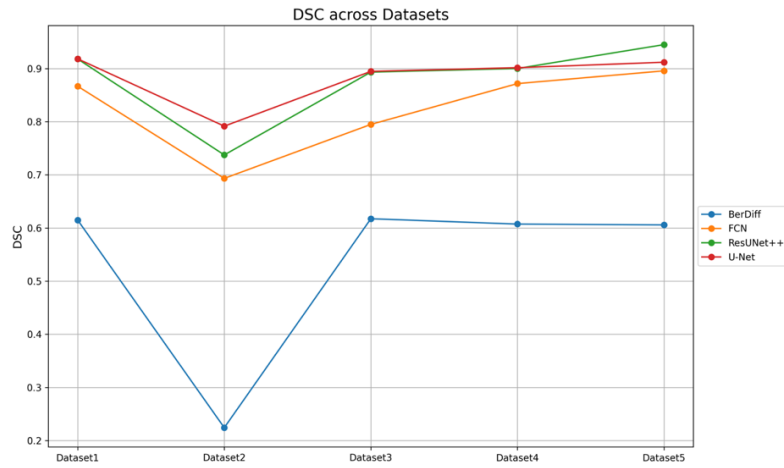


Figure 57. DSC values for each model across all datasets on unseen data.

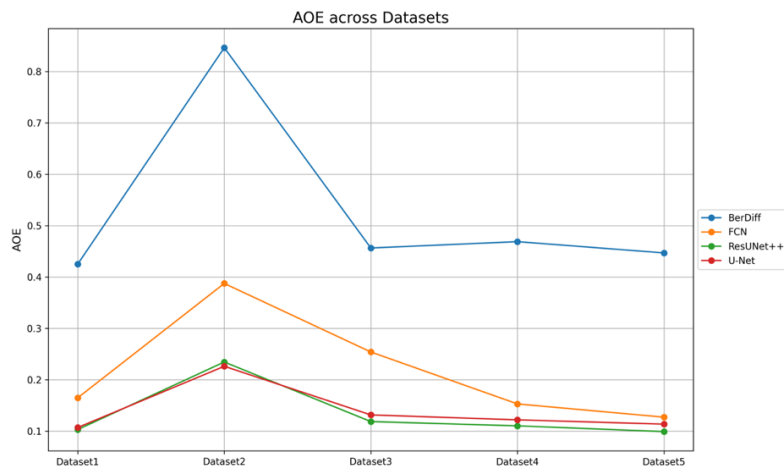


Figure 58. AOE values for each model across all datasets on unseen data.

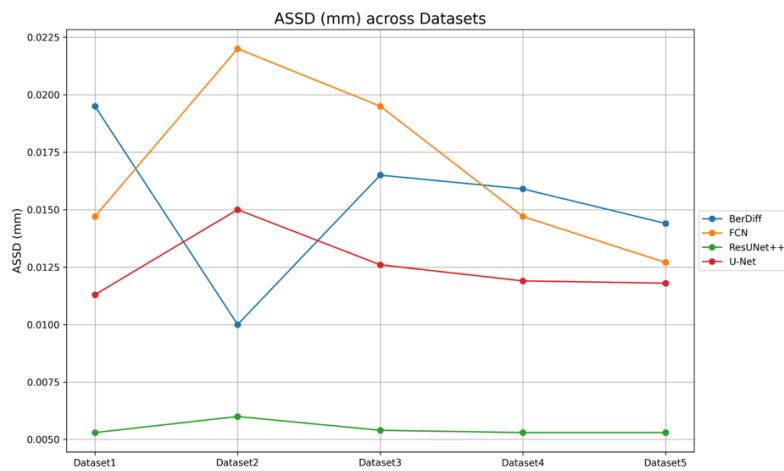


Figure 59. ASSD values for each model across all datasets on unseen data.

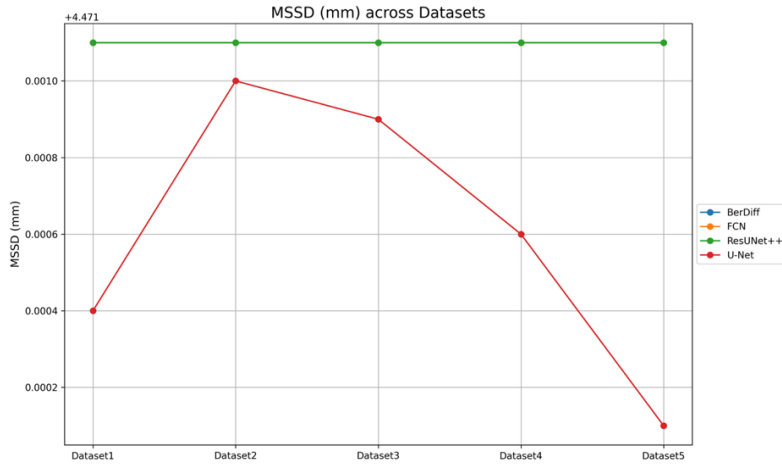


Figure 60. MSSD values for each model across all datasets on unseen data.

Figure 61 to Figure 65 illustrate, respectively, the results obtained in the experiments using datasets 1, 2, 3, 4 and 5, comparing the results from the inferences made on unseen data for the experiments conducted with the five different datasets. Each figure provides an overview of the performance achieved by the models, offering a direct comparison between the different approaches highlighting the influence of each dataset on the behavior and effectiveness of the tested models, demonstrating the robustness and adaptability of the models across different datasets.

2D Liver Segmentation in Abdominal Computed Tomography Scans

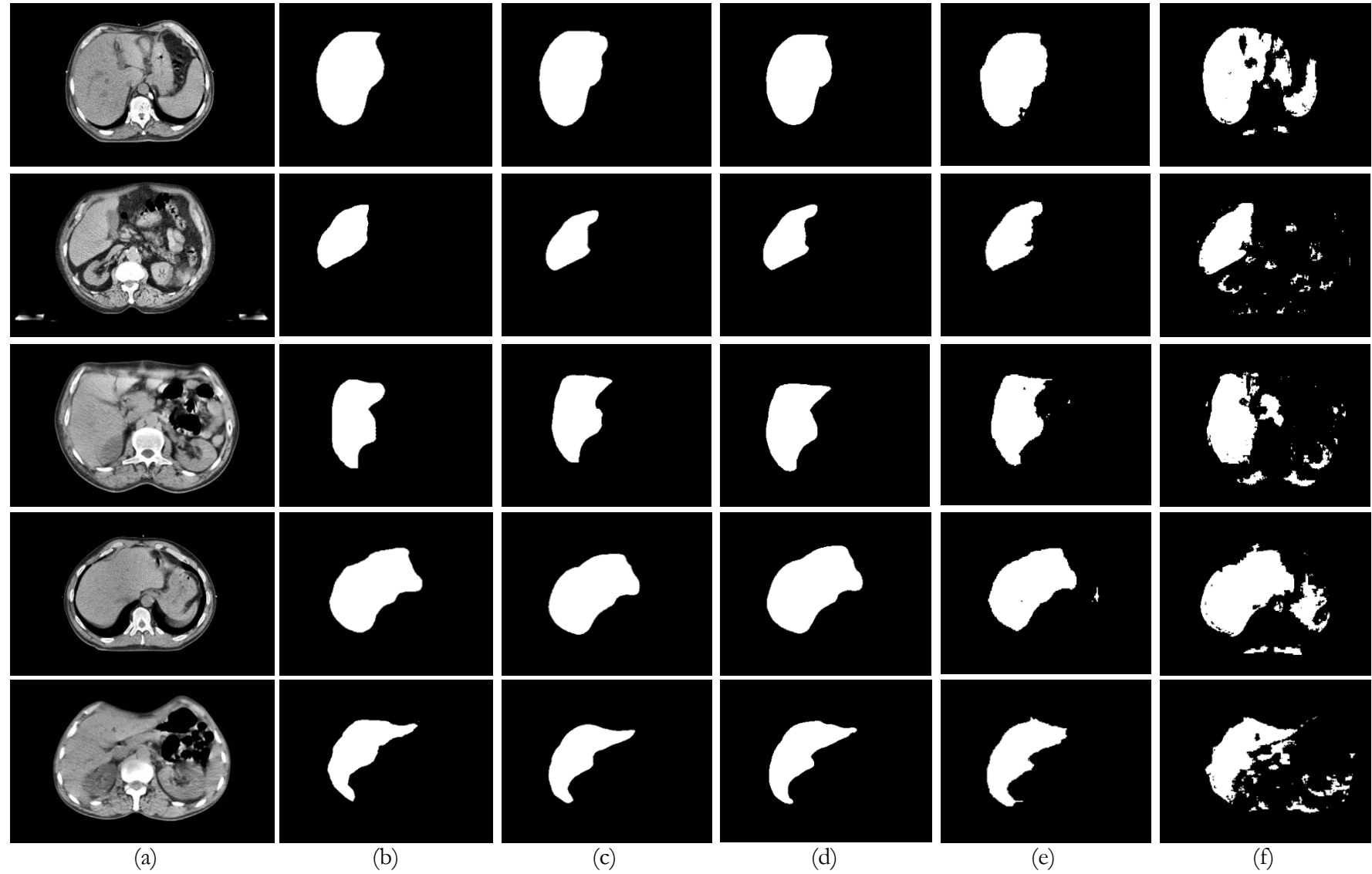


Figure 61. Comparison of the performance from each model for the dataset 1 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.

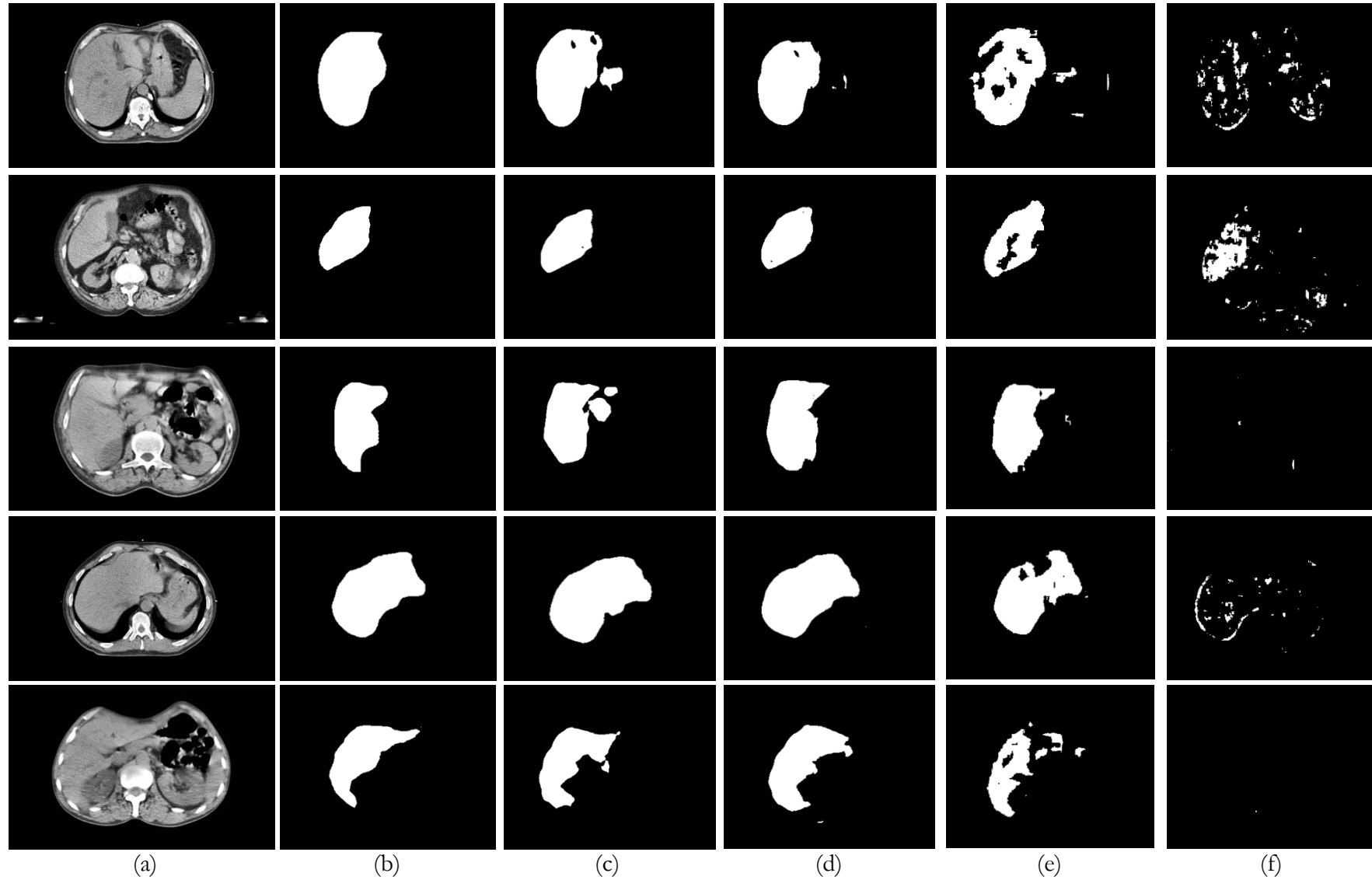


Figure 62. Comparison of the performance from each model for the dataset 2 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.

2D Liver Segmentation in Abdominal Computed Tomography Scans

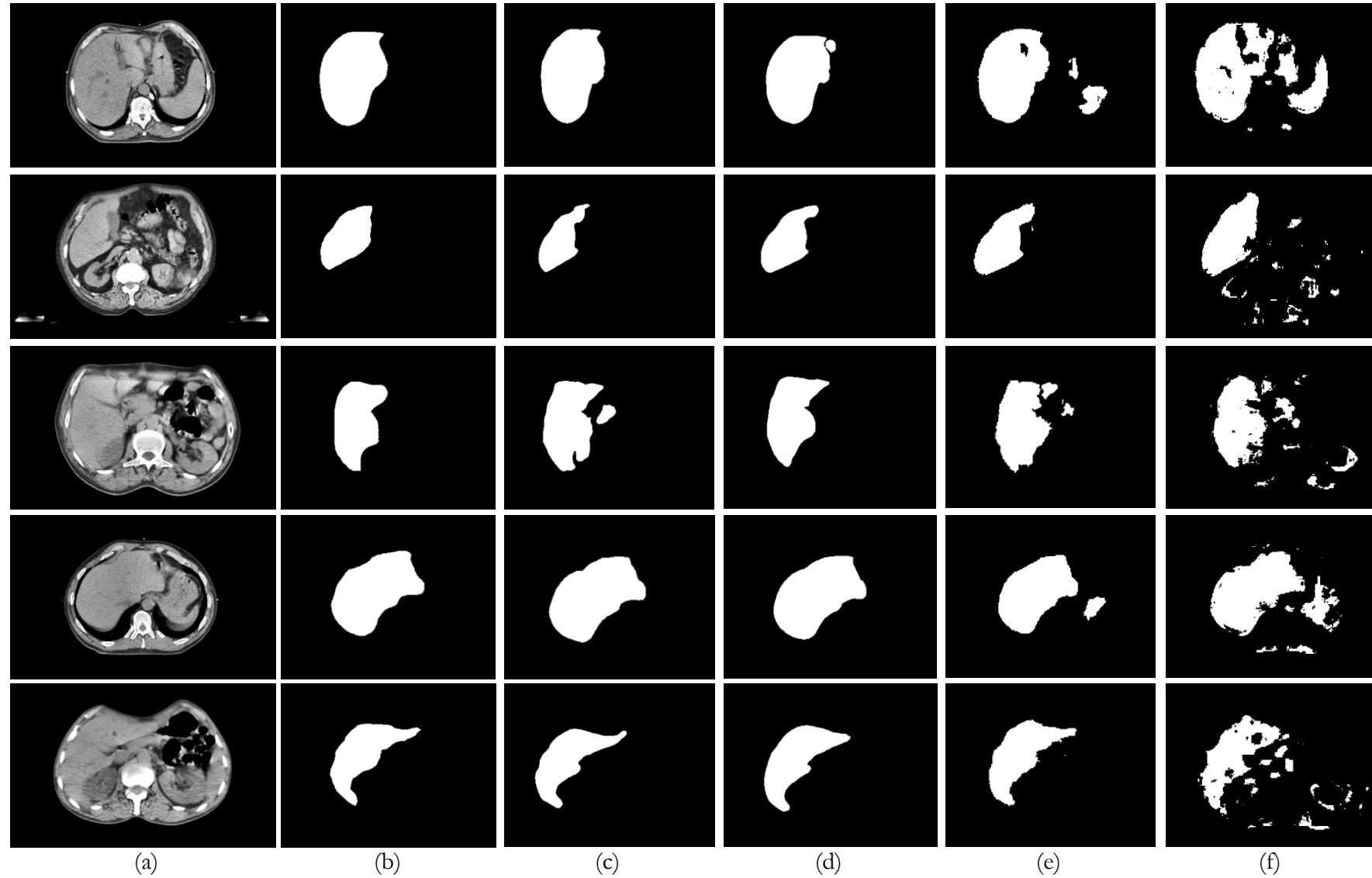


Figure 63. Comparison of the performance from each model for the dataset 3 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.

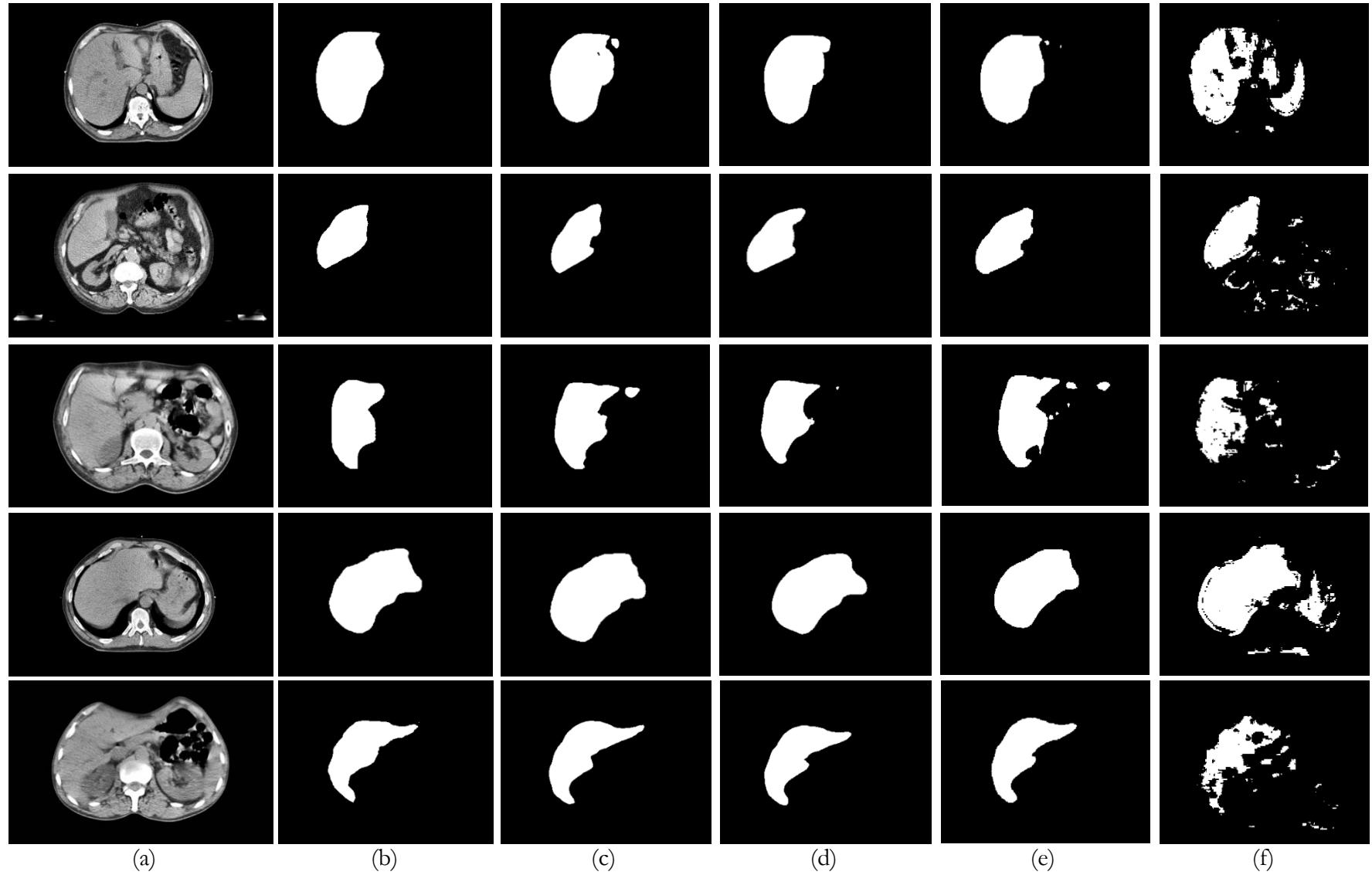


Figure 64. Comparison of the performance from each model for the dataset 4 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.

2D Liver Segmentation in Abdominal Computed Tomography Scans

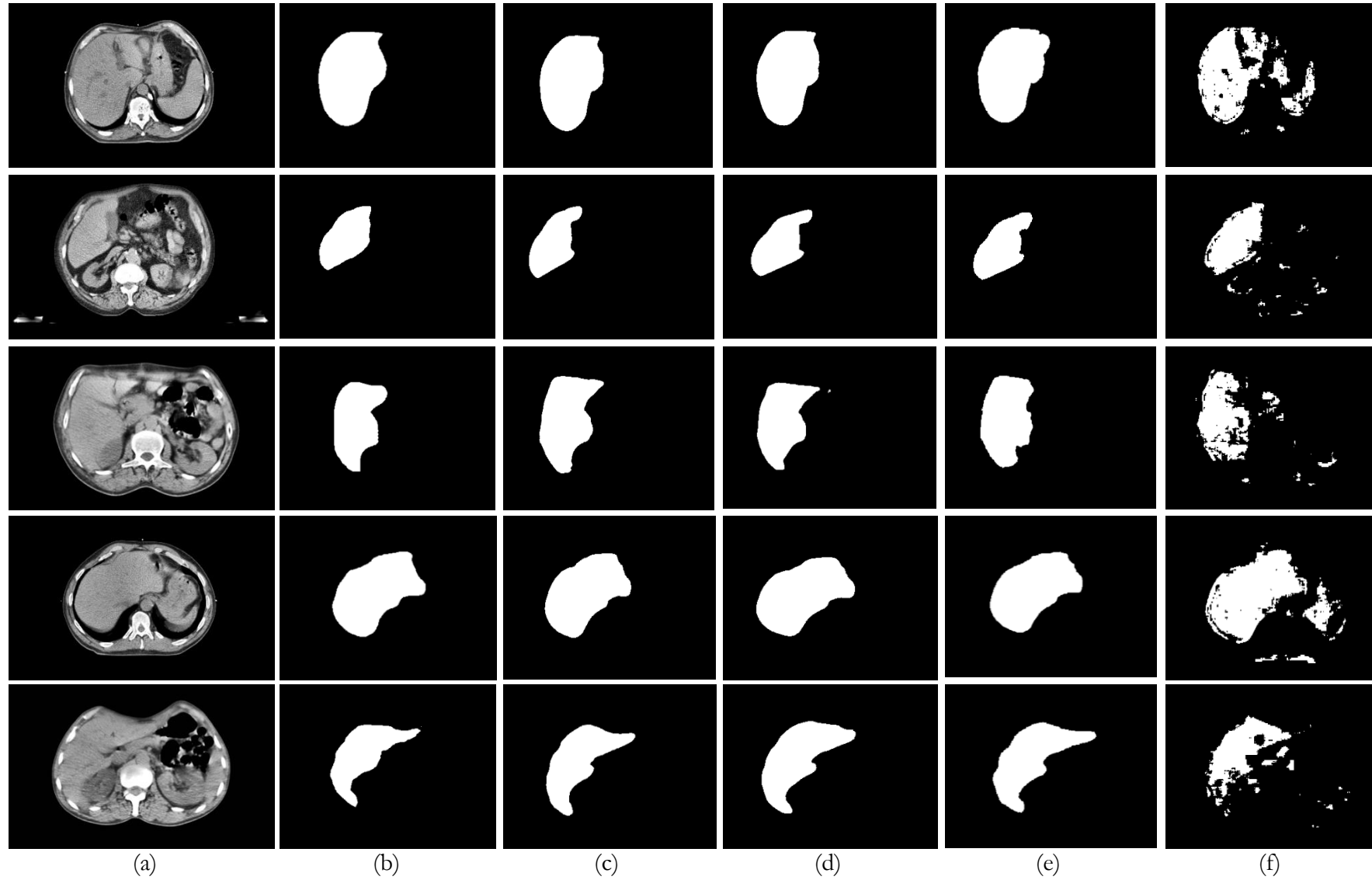


Figure 65. Comparison of the performance from each model for the dataset 5 on unseen data. (a) Original image. (b) Ground truth. (c) U-Net outcome. (d) ResUNet++ outcome. (e) FCN outcome. (f) BerDiff outcome.

6.3 Discussion

This project presented an approach to compare the capability of different segmentation models to perform liver segmentation, as well as to evaluate the generalization skills of each of the models.

This sub-chapter analyzes the performance of the four segmentation models used (U-Net, ResUNet++, FCN, and BerDiff) across the five different dataset combinations, representing a generalization test through *Private Data Incrementalization*, reminding the datasets experiments were as follows:

- Created dataset only (E1, E6, E11, E16)
- Public datasets only (E2, E7, E12, E17)
- 50 scans from the created dataset + public datasets (E3, E8, E13, E18)
- 100 scans from the created dataset + public datasets (E4, E9, E14, E19)
- 200 scans from the created dataset + public datasets (E5, E10, E15, E20)

s

6.3.1 Models Training Time

Table 6.10 shows the training and validation times for each experiment.

ResUNet++ presented the best model performance, achieving the best results despite having the longest per-epoch training time (31-33 minutes). Its total training times ranged from 3h:50m to 9h:20m, which is considerable but justified by its superior performance. The ability of the model to converge in fewer epochs (7-17) demonstrates its efficiency in learning from the data.

U-Net, the second-best performer, showed consistent speed performance with each epoch taking only 5-6 minutes. Its total training times fell between 3h:00m and 4h:05m, making it the most time-efficient among the models with the best segmentation studied within the context of this project.

FCN, while presenting lower results than ResUNet++ and U-Net, still offered reasonable results. Its training times varied considerably, from 4h:37m to 13h:18m, with each epoch lasting 15-17 minutes.

BerDiff, implemented with simplifications to the model, showed the lowest performance among the four. However, it also demonstrated the quickest training times, ranging from 2h:14m to 4h:28m, with each epoch taking about 10-12 minutes.

Based on the results presented, the impact of early stopping can be visible across all models, leading to variable total training times. This variability underscores the importance of considering both the potential for early convergence and the per-epoch training time when selecting a model for a given task.

2D Liver Segmentation in Abdominal Computed Tomography Scans

While some models might have shorter per-epoch times or fewer total epochs, this does not necessarily translate to improved performance. The efficiency of a model should be judged not just on its training time, but also on its final performance metrics and ability to generalize to unseen data, which is analysed in sub-chapter 6.3.2.

Table 6.10. Training and validation time for each experiment.

Experiment	Model	Number of epochs	Training and validation time
E1	U-Net	36	3h:30m
E2		42	4h:05m
E3		31	3h:00m
E4		36	3h:30m
E5		40	3h:53m
E6	ResUNet++	10	5h:29m
E7		16	8h:47m
E8		10	5h:29m
E9		17	9h:20m
E10		7	3h:50m
E11	FCN	37	10h:03m
E12		49	13h:18m
E13		28	7h:36m
E14		22	5h:58m
E15		17	4h:37m
E16	BerDiff	17	3h:09m
E17		21	3h:54m
E18		24	4h:28m
E19		14	2h:36m
E20		12	2h:14m

6.3.2 Models Segmentation Performance

This sub-chapter presents an analysis of the results presented by the various models, focusing on the ability of the models to adapt and generalize.

6.3.2.1 U-Net

The U-Net model presented versatility and robustness across all the datasets. Its consistently high DSC scores, even when applied to public datasets (maintaining $DSC > 0.79$), underscore its strong generalization capacity. This model exhibited a good ability to adapt and improve with the *Private Data Incrementalization*, as evidenced by steady enhancements in both DSC and AOE metrics. Perhaps most notably, U-Net's boundary delineation capabilities proved exceptionally precise and stable, maintaining a remarkably ASSD ranging from 0.0113 to 0.0150 mm across all experiments. This consistent performance in boundary precision, as well as its strong overall segmentation accuracy and adaptability, makes U-Net as a highly reliable and efficient choice for the objectives of the project.

6.3.2.2 ResUNet++

The ResUNet++ model showed the best performance in the comparative study, showcasing superior capabilities across all key performance indicators. This model demonstrated the most robust generalization capacity among all the models evaluated. A relevant aspect of ResUNet++'s performance was its unparalleled precision in boundary delineation, consistently maintaining the lowest ASSD throughout all experiments. This exceptional boundary accuracy, combined with its ability to achieve the highest DSC of 0.9449 and the lowest AOE of 0.0990 in the final experiment (E10), underscores its superior segmentation quality and precise liver localization capabilities.

6.3.2.3 FCN

The FCN model demonstrated moderate generalization capacity, showing strong capabilities when applied to the created dataset, but presented more pronounced performance degradation when transitioning to public datasets, especially in comparison to the more robust U-Net and ResUNet++ models. This suggests a higher sensitivity to variations in data sources and characteristics. However, FCN showed a commendable capacity for improvement with the *Private Data Incrementalization*, indicating its potential for enhanced performance given sufficiently varied input. A consistent observation across experiments was FCN's higher ASSD compared to U-Net and ResUNet++, pointing to less precise boundary delineation capabilities. This suggests that while FCN can produce good overall segmentations, it may struggle with the fine details of organ boundaries.

6.3.2.4 BerDiff

The BerDiff model consistently underperformed relative to the other models, presenting challenges across all experiments. Its performance was particularly problematic when generalizing to public datasets, evidenced by a decline in DSC from 0.6149 to 0.2243.

This substantial drop highlights the modified BerDiff's limited ability to adapt to varied data sources and characteristics. Interestingly, despite its overall poor performance, BerDiff demonstrated the highest relative improvement with *Private Data Incrementalization*. This observation suggests that the model may require substantially more extensive and diverse training data to achieve competitive performance levels. However, even with these improvements, BerDiff's performance remained below that of the other models in the study.

Although the only evidence of the original BerDiff model's performance can be found in the authors' original paper [138], it is possible to verify that, possibly due to the simplification of the model described in sub-chapter 5.8.4, it performed poorer than the other models. In any case, for contexts such as this project (without sufficient conditions and computational resources to train the model with the original architecture) the use of such complex and heavy models is not advantageous, since there are models that seem to have a simpler architecture, but that remain very powerful for their intended purpose.

6.3.2.5 Models Generalization Capacity

ResUNet++ demonstrated the best generalization capacity, showing the most consistent performance across different datasets and the most improvements *Private Data Incrementalization*. U-Net also showed strong generalization capabilities, maintaining high DSC scores across experiments and showing steady improvements with more data. FCN presented moderate generalization capacity, with more pronounced performance drops on public datasets but good improvements with *Private Data Incrementalization*. However, the simplified BerDiff model demonstrated poor generalization capacity, struggling with public datasets and showing limited improvement even with additional data.

6.3.2.6 Summary

Based on the results from all the experiments, it is possible to answer the question:

- **Q6 — Which is the most efficient automatic segmentation model for the case study within the chosen algorithms?**

The most efficient automatic segmentation models for this case study are U-Net and ResUNet++. However, ResUNet++ slightly edges out U-Net as the most efficient model overall. While both models performed exceptionally well, ResUNet++ showed slightly better generalization and more consistent improvements across *Private Data Incrementalization*. Also, ResUNet++ presented a balance between segmentation performance and relatively efficient training time, despite longer per-epoch duration. Its superior performance in boundary precision (as indicated by the lowest ASSD) and overall segmentation quality (highest DSC in the final experiment) makes it the most efficient model for this case study.

The choice between ResUNet++ and U-Net might depend on specific use case requirements. If computational resources are a constraint, U-Net might be preferable as it is generally less complex than ResUNet++. However, if the highest possible accuracy and generalization are the primary concerns, ResUNet++ would be the top choice.

6.3.3 Datasets Analysis

This sub-chapter presents an analysis of the results based on the dataset's selection and impact, focusing on the datasets' influence on models' performance.

6.3.3.1 Dataset 1

The created dataset (experiments E1, E6, E11, E16) on the U-Net and ResUNet++ models showed an excellent performance, both of them with accuracy above 0.99 and DSC scores above 0.91 for both training and unseen data, with high DSC indicating good overlap between predictions and ground truth, as well as a low ASSD, that suggests precise boundary delineation, as can be seen in the respective qualitative results. It is also worth mentioning that the MSSD value did not show variation for the different models, with a variation after several decimal digits. FCN performed well but slightly below U-Net and ResUNet++, with a DSC score of 0.8666 on unseen data. BerDiff struggled with this dataset, achieving a DSC score of only 0.6149 on unseen data. It is also visually evident that the performance of the other models differs from that of BerDiff through the qualitative results.

6.3.3.2 Dataset 2

For the public datasets (experiments E2, E7, E12, E17), all models showed a slight decrease in performance, indicating some challenges in generalization. U-Net and ResUNet++ maintained relatively high performance, with DSC scores above 0.73 on unseen data. FCN's performance dropped more, with a DSC score of 0.6933 on unseen data. BerDiff showed a pronounced decrease in performance, with a DSC score of only 0.2243 on unseen data, making it notable for its lower performance on metrics such as loss, which was higher than the rest of the models.

6.3.3.3 Datasets 3, 4 and 5

For the *Private Data Incrementalization* mixed datasets (experiments E3-E5, E8-E10, E13-E15, E18-E20), the U-Net and ResUNet++ models showed improved performance as more data was added, with the best results typically seen in the experiments with the datasets 4 and 5. ResUNet++ showed slightly better generalization, with more consistent improvements across *Private Data Incrementalization*, presenting the DSC value increasing from 0.8932 to 0.9449 and AOE decreasing from 0.1185 to 0.0990, and maintained the lowest ASSD throughout (0.0053-0.0054 mm), suggesting superior boundary delineation. FCN model showed improvements with *Private Data Incrementalization*, but the gains were less pronounced compared to U-Net and ResUNet++, which is consistent with the behavior of the preceding datasets and the previous experiences. The model's performance stabilized in later experiments, suggesting it might benefit from even more data. BerDiff model also presented some improvements with *Private Data Incrementalization*, but its performance remained below the other models. The BerDiff model's DSC scores on unseen data improved from 0.2243 in E17 to 0.6058 in E20, showing that it benefits from more diverse data.

Despite having initially considered performing data augmentation (whose techniques thought to be used can be found in Appendix N) during the data preparation phase, the process of generalization tests through *Private Data Incrementalization* turned out to be a process of increasing the number of input data, the difference being that it brings more benefits than a usual data augmentation process in which the variations may not be as visible as, for example, different images.

6.3.3.4 Pre-processing and Post-processing Impact

All models underwent the same pre-processing and post-processing stages, allowing for a fair comparison of their intrinsic generalization capabilities. The consistent MSSD values (approximately 4.4721 mm) across all models and experiments indicate that the post-processing stage effectively standardized the maximum boundary distances, suggesting that while different CT scan sources affect overall performance, they do not notably impact the maximum error in boundary delineation. This consistency in MSSD highlights the importance of post-processing in ensuring comparable results across different models and datasets.

The pre-processing stage likely contributed to the models' ability to handle diverse datasets, as evidenced by the improvements seen with *Private Data Incrementalization*. However, the varying degrees of performance drop when moving from the created dataset to public datasets indicate that the models have different inherent capabilities in leveraging the pre-processed data for generalization.

6.3.3.5 Private Data Incrementalization

The generalization tests through *Private Data Incrementalization* revealed that U-Net and ResUNet++ are the most promising models for the given segmentation tasks. They showed the best performance across different dataset combinations and demonstrated good generalization capabilities. FCN, while performing well, consistently lagged behind U-Net and ResUNet++, suggesting that it might benefit from architectural improvements or more extensive training. BerDiff's poor performance across all experiments indicates that it may not be suitable for these specific segmentation tasks without substantial modifications, which will probably correspond to the architecture of the original model.

All models showed variations in performance when moving from the created dataset to public datasets, indicating that the source and nature of CT scans influences model performance. For example, U-Net's DSC dropped from 0.9182 on the created dataset to 0.7915 on public datasets, showing that the characteristics of different CT scan sources can greatly affect segmentation performance.

The models generally performed better on the created dataset compared to public datasets, suggesting that CT scans from different sources (e.g., different machines, protocols, or patient populations) present generalization challenges, highlighting the importance of diverse training data that represents various CT scan sources to improve model robustness.

As more diverse CT scan data was added (combining created and public datasets), all models showed improvements in performance. An example of this behavior is the ResUNet++ model, which DSC value increased from 0.7374 on public datasets alone to 0.9449 when combining created and public datasets, indicating that a mix of CT scan sources enhances model performance.

The ASSD metric, which is sensitive to boundary precision, showed variations across different dataset combinations. This suggests that different CT scan sources may affect the models' ability to precisely delineate organ boundaries, possibly due to variations in image quality, contrast, or resolution.

Different models showed varying levels of sensitivity to CT scan sources. ResUNet++ and U-Net demonstrated better adaptability to diverse CT scan sources compared to FCN and BerDiff. This suggests that model architecture plays a role in how effectively a model can handle variations in CT scan characteristics.

6.3.3.6 Summary

Based on the results and the analysis presented above, it is possible to answer the question:

- **Q7 — How does the data source (CT scans) impact on the performance of the segmentation models?**

CT scan data source impacts segmentation model performance, affecting accuracy, generalization, and boundary precision. Models perform best when trained on diverse CT scan sources that represent the variability expected in real-world applications. This underscores the importance of creating comprehensive and diverse datasets for training robust segmentation models in medical imaging applications.

To mitigate the impact of data source variability and improve model performance it is suggested to incorporate diverse CT scan sources in training data, continuing to implement pre-processing techniques to standardize input data. Another relevant aspect can be consider domain adaptation techniques to better handle CT scans from new, unseen sources. Last but not least, regularly update and refine models with new CT scan data to maintain performance across evolving imaging technologies and protocols.

7 L+VER: SEGMENTATION APPLICATION

This chapter describes the development of a GUI, from the specification of requirements for the solution to the interface itself and its main features.

7.1 Requirements Specification

For the specification of the requirements that must be met for the development of the project, interviews and brainstorming sessions were held with physicians from HUC's Radiotherapy Department with the aim of understanding their needs and expectations, so as to answer the question Q5 initially formulated in sub-chapter 1.2:

- **Q5 — What are the best specifications and project conditions according to the HUC's temporal needs and priorities?**

To be able to develop the liver segmentation tool, a list of functional and non-functional requirements is presented below.

7.1.1 Functional Requirements

Table 7.1 outlines the functional requirements for a liver segmentation software designed for use with CT scans. The requirements ensure that the software presents a useful tool considering the HUC physician's needs and expectations.

Table 7.1. Functional requirements.

Requirement ID	Description	Acceptance criteria
FREQ-01	The system shall allow users to upload CT scans in DICOM format.	Users can successfully upload DICOM format files. Non-DICOM files are not accepted.
FREQ-02	The system visualizes the segmentation of the liver and overlays it on the original image.	Users can view the liver segmentation and its overlay on the original image.
FREQ-03	The system shall allow visually adapting the Hounsfield scale of the image.	Users can adjust the Hounsfield scale and see the changes reflected in real-time on the image.
FREQ-04	The system shall allow zooming in and out of the original image and overlay segmentation.	Users can zoom in and out of all image views (original, mask, overlay) smoothly.
FREQ-05	The system shall clear the image to allow viewing a new one.	Users can clear the current image and show a new one without errors.

Requirement ID	Description	Acceptance criteria
FREQ-06	The system shall export the segmentation result in different image formats, such as: • Image file (PNG, TIF, JPG) • Scalable Vector Graphics (SVG) • CSV from plot data	Users can export the current segmentation result in different image formats defined.
FREQ-07	The system shall change its background color (screen theme) according to the user's settings preferences.	The software's background theme changes based on the user's preferences (dark or light theme).

7.1.2 Non-functional Requirements

Table 7.2 presents the non-functional requirements to ensure that the software meets user needs effectively while maintaining high standards of performance, security, and usability.

Table 7.2. Non-functional requirements.

Requirement ID	Description	Acceptance criteria
NFREQ-01	The system shall be user-friendly and intuitive.	Users can perform tasks without extensive training.
NFREQ-02	The system shall be compatible with Microsoft Windows, MacOS and Linux.	The software installs and runs without issues on Microsoft Windows, MacOS and Linux operating systems.

7.2 Development

The development of the application involved a detailed configuration process to ensure a robust and efficient system. The technologies and tools employed throughout the process of developing the application are detailed below, with emphasis on the options chosen for both the backend and frontend development. Figure 66 shows a graphical representation of the use made of the tools chosen to develop the application.

2D Liver Segmentation in Abdominal Computed Tomography Scans

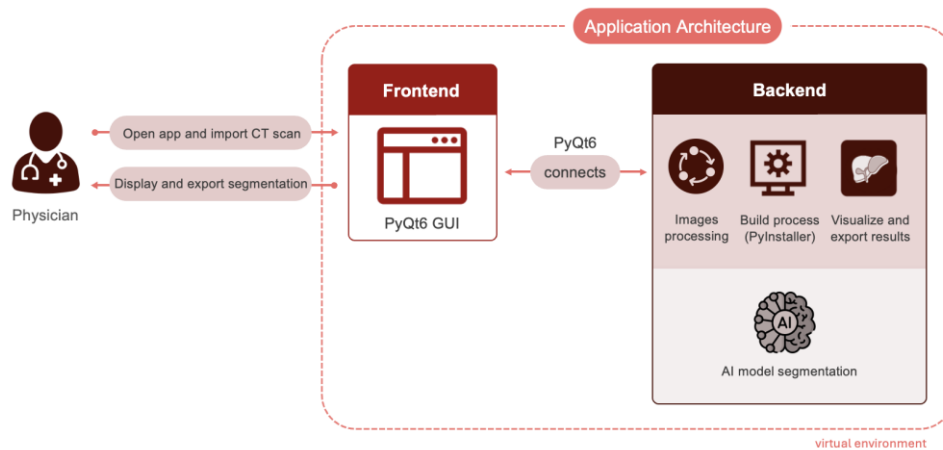


Figure 66. Application architecture diagram.

7.2.1 Environment

The GUI of the application was developed using PyQt6, a set of Python bindings for the Qt application framework. Python 3.9.19 was chosen for its compatibility with PyQt6 and its extensive libraries, facilitating the creation of a responsive and user-friendly interface.

7.2.2 Application Build Process

PyInstaller was employed to build the application. This tool converts Python applications into standalone executables, ensuring that the end-users can run the application without needing a Python interpreter. It bundles the application code, libraries, and other dependencies into a single package.

7.2.3 Automation and Requirements

A custom script was developed to automate the build process. This script handles tasks such as compiling the code, packaging the application, and running tests, ensuring a streamlined and repeatable build process.

A virtual environment was used to create the conditions and comply with the requirements for the backend and frontend in a consistent and isolated environment. Table 7.3 presents the main components of the application.

Table 7.3. Components within the application.

Frontend	Backend	AI segmentation model
Handles user interactions and communicates with the backend. PySide6 GUI packaged with PyInstaller.	Handles requests from the frontend, processes DICOM images, and interacts with the database and AI segmentation model.	AI best segmentation model (ResUNet++), dependencies, and the necessary Python scripts to run the model.

In order to be able to run the application, the libraries and packages in Appendix L must be installed.

7.3 Main Features

This sub-chapter provides a detailed overview of the main features within the application (Figure 67), underlining their functionality and importance while referencing their visual representations.

Files Export Zoom Segmentation Transformations Light Theme

Figure 67. Menu bar of the application main functionalities.

7.3.1 Menu Files

The Files menu of the application allows the user to upload individual files or entire folders (CT series). This sub-chapter describes the main options available in this menu.

7.3.1.1 Upload Single File(s)

The import “single file” feature enables users to upload single images (files) from CT scans into the application for further processing. It supports DICOM file format and ensures that the uploaded images are correctly formatted and displayed. The system includes validation checks to prevent corrupted or incompatible files from being imported, ensuring data integrity. This feature is shown in Figure 68.



Figure 68. Single CT scan image (file) uploaded.

7.3.1.2 Upload Folder (CT Series)

The import “CT series” feature enables users to upload a set (folder) of CT scan images (files) into the application for further processing. As before, it supports DICOM file and ensures that the uploaded images are correctly formatted and displayed. The system includes validation checks to prevent corrupted or incompatible files from being imported, ensuring data integrity. This feature is shown in Figure 69.

2D Liver Segmentation in Abdominal Computed Tomography Scans

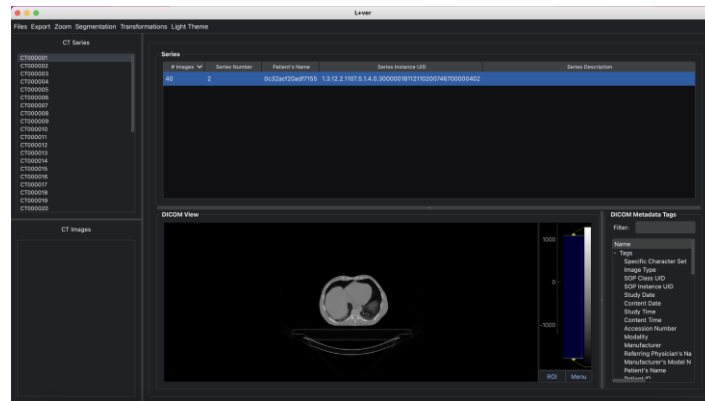


Figure 69. CT scan folder uploaded.

Whenever the application is opened, a new session is created. During the session, the application checks if a folder (CT series) has already been uploaded to prevent duplicate uploads. If the same folder is re-uploaded within the same session, the user is notified to ensure efficient folder management. However, if the folder is removed from the series list or the session ends (e.g., by closing the application), it can be uploaded again. This validation only applies locally within the current session.

7.3.1.3 Upload Multiple Folders (Series)

The app allows users to upload multiple folders for analysis, displaying essential information about each folder in the "Series" frame, as shown in Figure 70. This includes the number of images in each folder, the number of series, the patient's name (usually anonymized), the series instance UID (a DICOM tag commonly used for studying patients), and the series description. All this information is extracted by reading the metadata of the DICOM images within the folders.

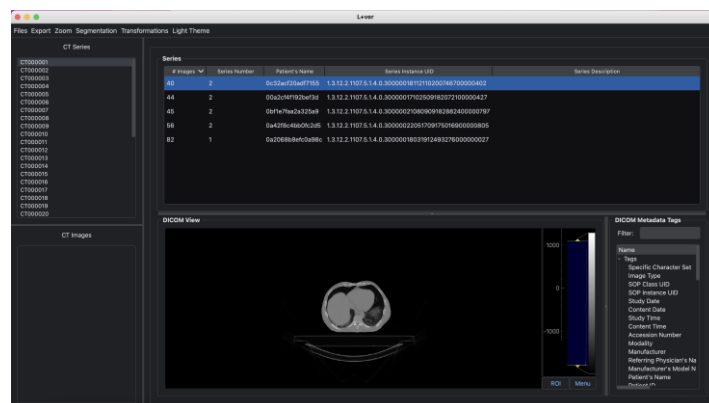


Figure 70. Multiple CT scan series visualization.

7.3.1.4 Remove Image from Visualization Frame

The app includes a feature for managing the image visualization frame ("DICOM View"), allowing users to display the original DICOM images and providing the option to remove images from the view, helping to keep the workspace organized and focused (Figure 71).

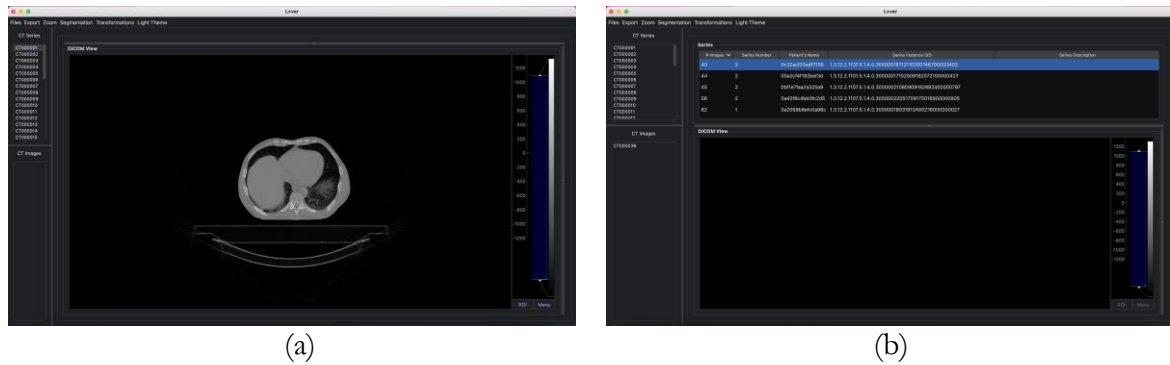


Figure 71. DICOM View frame: (a) Original image visualization. (b) Image removed from DICOM View frame.

7.3.2 Menu Export

The app offers various export options for segmentation results, allowing users to save images in different ways, copying it to the clipboard and also export it as multiple formats, including PNG, TIF, and JPG (Figure 72). Additionally, it supports exporting in Scalable Vector Graphics (SVG), CSV format for plot data, and Graphic Interchange formats, providing flexibility for different use cases and ensuring compatibility with various tools and platforms.

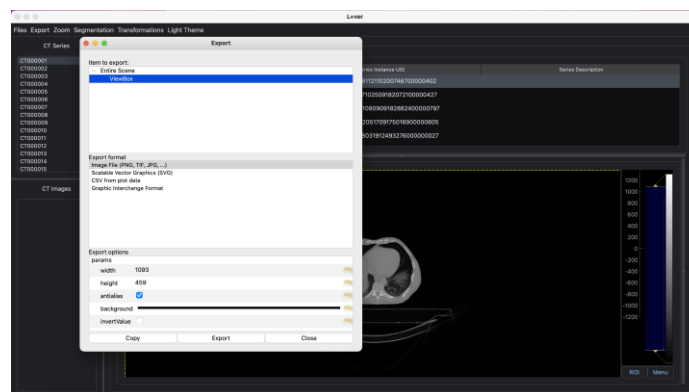


Figure 72. Export image options.

7.3.3 Menu Zoom

The app includes an option to adjust the zoom level, enabling users to zoom in and zoom out on the images for a more detailed or broader view (Figure 73), depending on their needs, always allowing them to reset to the original.

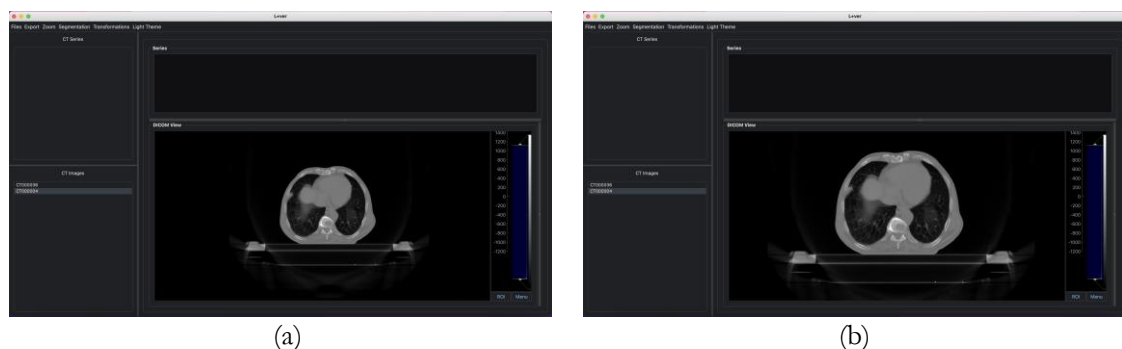
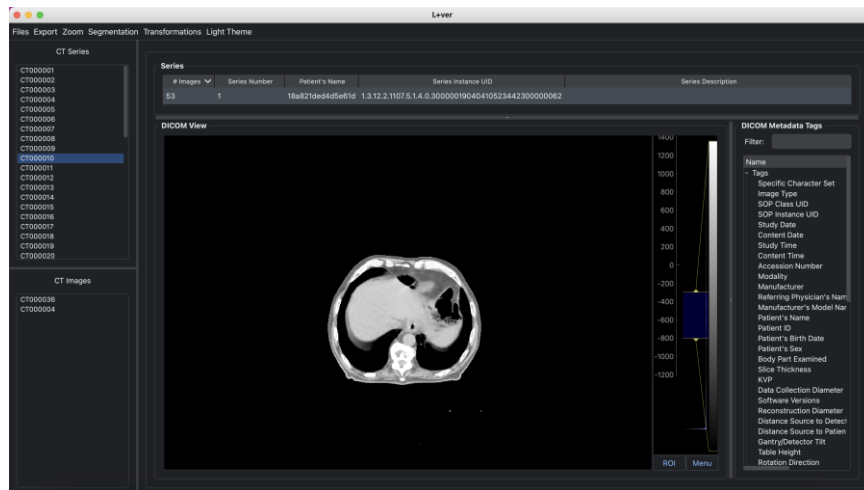


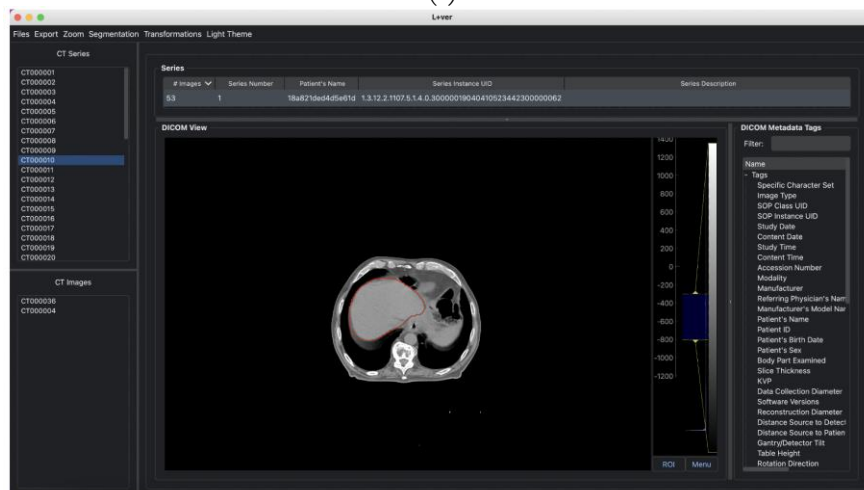
Figure 73. Zoom feature: (a) 100% zoom. (b) 200% zoom.

7.3.4 Menu Segmentation

The application features a liver segmentation tool using the ResUNet++ model, which was the best performing of the four tested models (see chapter 6), enabling users to perform liver segmentation on the current CT series or the selected file, enhancing its utility for medical image analysis (Figure 74).



(a)



(b)

Figure 74. Segmentation feature: (a) Original image visualization. (b) Liver segmentation visualization.

7.3.5 Menu Transformations

The app provides transformation options, including applying a negative (inverse) effect to the image, rotating it, and performing reflections, giving users flexibility in how they manipulate and analyze the images (Figure 75).

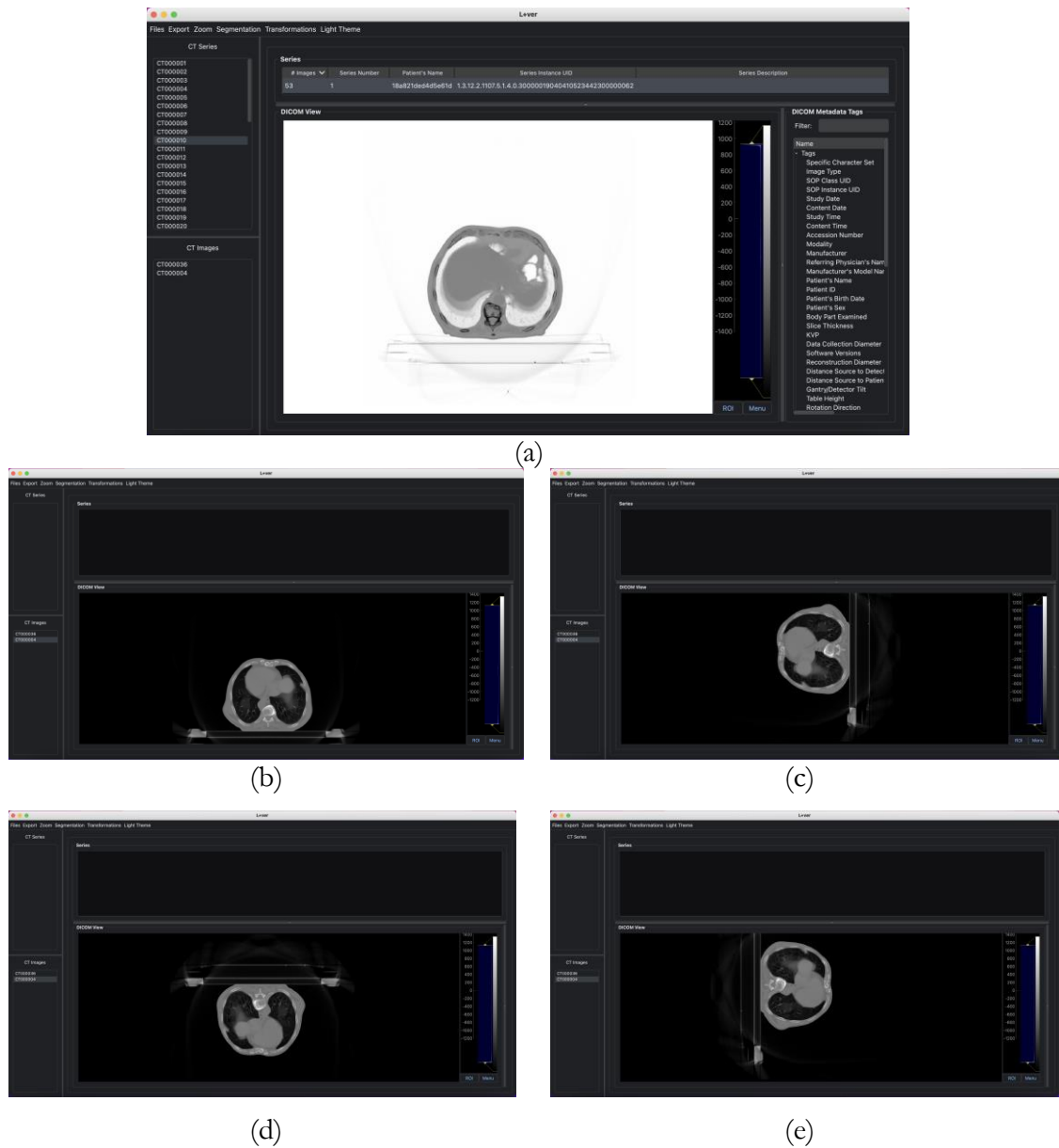


Figure 75. DICOM image transformation: (a) Negative operation. (b) Flip horizontal. (c) 90 CCW rotation. (d) Flip vertical. (e) 90 CW rotation.

7.3.6 Menu Light Theme

The app offers both light and dark themes (Figure 76), allowing users to switch between them based on their preferences for a more comfortable viewing experience.

2D Liver Segmentation in Abdominal Computed Tomography Scans



Figure 76. Examples of light screen theme.

7.3.7 Additional Features

This sub-chapter describes additional features related to layout configuration, DICOM metadata visualization, HU scale modification and ROI visualization.

7.3.7.1 Layout Configuration

The app provides the option to customize the layout according to the user's preferences, allowing for different configurations and views, as shown in the Figure 77. This flexibility enhances the user experience by enabling them to arrange the interface in a way that suits their workflow. This customization only applies locally within the current session.

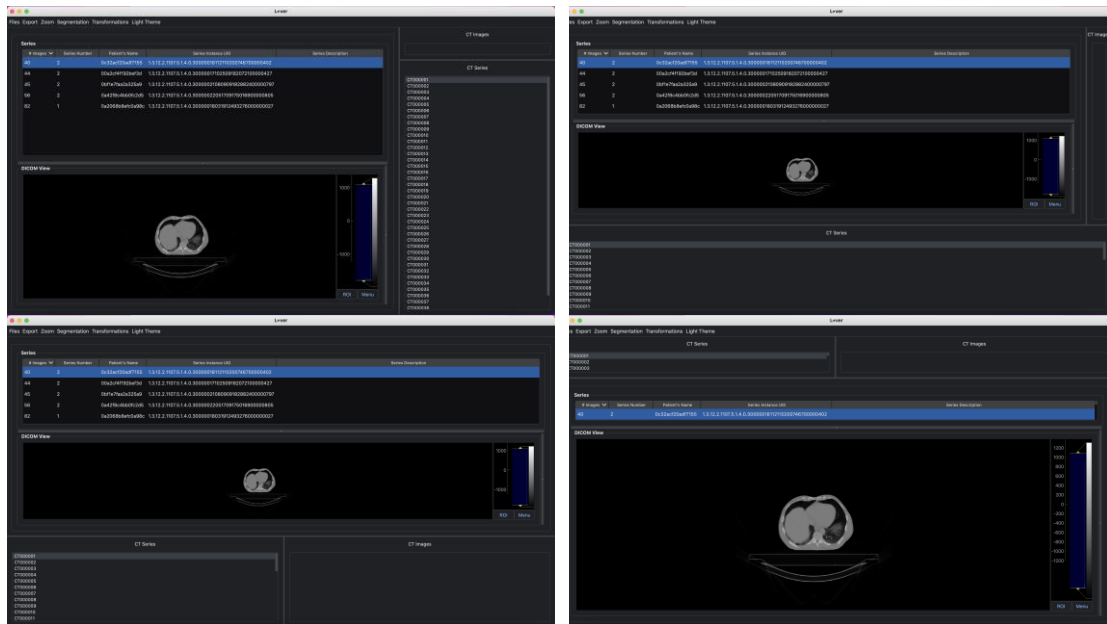


Figure 77. Some of the application layout options.

7.3.7.2 DICOM Metadata Visualization

The app provides a dedicated frame for accessing DICOM metadata tags, enabling users to search/filter for specific values or tags. This functionality allows for easy verification of important information within the DICOM files (Figure 78).

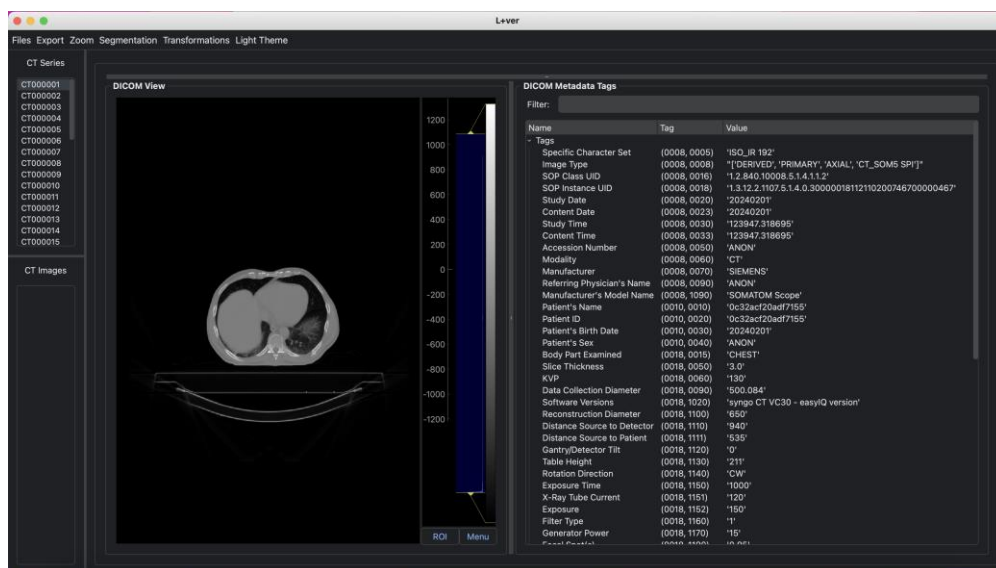


Figure 78. DICOM metadata.

7.3.7.3 Hounsfield Units

The app enables users to adjust the HU range of the images to be visualized, allowing for enhanced visualization of different structures and tissues within the CT scan images, being possible to see how the application of HU can impact the image quality (Figure 79).

2D Liver Segmentation in Abdominal Computed Tomography Scans

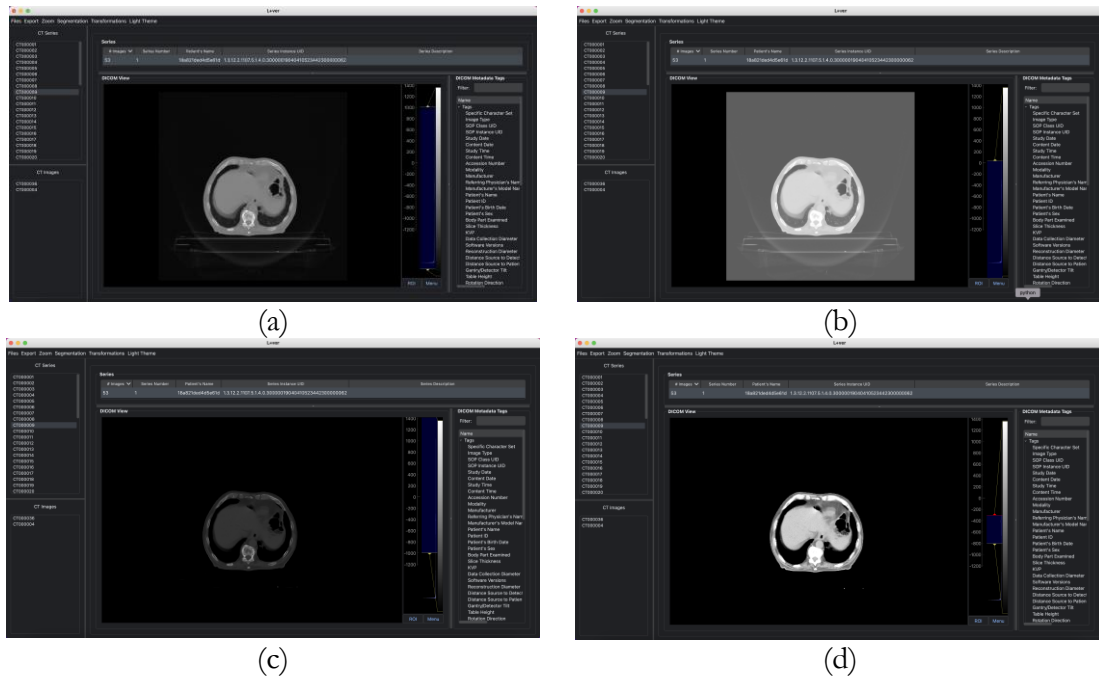


Figure 79. HU modifications: (a) Original image. (b) Lower values of HU. (c) Higher values of HU. (d) Smaller range of values of HU.

7.3.7.4 Region of Interest Visualization

The app displays an ROI plot that is based on the HU values, which are used to quantify the density of tissues in CT images. In the visualization, once an ROI is selected on the DICOM image, the plot below dynamically represents the variation in HU values across the region (Figure 80). This allows users to analyze the intensity of the selected area in terms of tissue density, helping to identify and differentiate structures within the image more effectively.

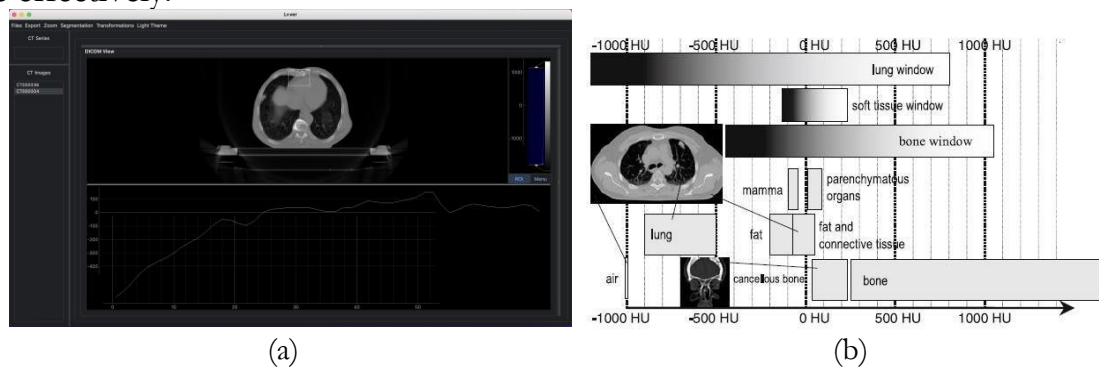


Figure 80. HU application: (a) ROI visualization based on HU values. (b) HU scale for different structures and tissues. Source: [137].

8 CONCLUSIONS

This project addressed a comprehensive evaluation of four different segmentation models (U-Net, ResUNet++, FCN, and BerDiff) across various datasets to assess their generalization capabilities. The results offer valuable insights into both the practical application of liver segmentation and the performance characteristics of different ANN architectures in medical imaging tasks.

Through this project an intuitive tool was developed allowing HUC's physicians to segment one of their main organs of study at no cost. This reduces the expense typically associated with the segmentation technique, crucial for treatment planning and monitoring the progression of oncological pathologies. By freeing the time of the physicians allows them to focus on more critical and complex tasks. The tool offers some flexibility in terms of its GUI, allowing it to be customized according to the preferences of the physicians. It easily and quickly imports the scan intended for liver segmentation and features adjustable scan display settings. This includes the ability to modify the HU for better visualization, giving access to scan segmentation with a considerably good performance compared to other methods reviewed in the literature.

Taking an in-depth look at the timetable for carrying out the tasks initially presented in Table 1.1, although the project was completed in a reasonable timeframe, it was initially intended that the T1 completion period would be shorter, as the project was started at an early date. Due to logistical reasons and permissions to access HUC data, the T3 process ended up being delayed in both its start and its duration, as it took a considerable period of time to learn how to submit a project to the HUC Ethics Committee, as well as to prepare and review all the necessary documentation. The support of the HUC was essential for this process, and a strong support was received from the HUC throughout T3. T5 turned out to be more time-consuming as it involved learning about new concepts and techniques. As an advantage and a very positive aspect of these changes in the schedule, there has been a series of publications and scientific contributions within the scope of the project.

8.1 Strengths

The project provides a comprehensive comparison of different segmentation models, which conducted across various dataset combinations, offers deep insights into the generalization capabilities of these models. Particularly noteworthy is the assessment of *Private Data Incrementalization*, which evaluated how models perform with the progressive addition of diverse data. This approach provided valuable insights into the importance of dataset diversity in medical imaging tasks. These results not only validate the effectiveness of the

chosen models but also contribute to the broader understanding of model behavior in medical image segmentation tasks, also considering the substantial number of scientific publications submitted during the development of the project.

Through this project, a cost-effective and intuitive tool for liver segmentation was developed for HUC's physicians. This tool reduces the expenses associated with segmentation techniques crucial for treatment planning and monitoring oncological pathologies. The application features a user-friendly interface with a flexible GUI that can be customized to physicians' preferences, allowing for easy importation of scans and offering adjustable display settings, including modifiable HU for enhanced visualization. By automating the segmentation process with good accuracy compared to methods reviewed in the literature, this tool frees up valuable time for physicians to focus on more critical tasks.

8.2 Limitations

Despite the project's successes, several limitations were encountered that are important to acknowledge. An important constraint was the limited computational resources, which restricted the ability to test more robust ANN architectures. This limitation was particularly evident in the performance of the simplified BerDiff model, which consistently underperformed compared to other models, indicating that some complex architectures may require substantial resources to be effective.

The inability to publicly share the dataset created for this project due to ethical considerations stipulated by HUC's Ethics Committee, which limits the reproducibility of the exact results and hinders broader validation by the scientific community. Additionally, there were technical barriers in developing automatic export functions and direct integration with the hospital's ARIA server due to strict security measures in the hospital's IT infrastructure. These challenges represent the complexities of implementing new technologies in sensitive healthcare environments.

Furthermore, due to time constraints and hospital authorization processes, it was not possible to conduct usability tests in the hospital departments. This limitation means that while the tool shows promise in controlled settings, its performance and usability in real-world clinical environments remain to be fully validated.

8.3 Future Work

For the future, there are numerous avenues that could enhance the impact and capabilities of the liver segmentation tool. One interesting task is the implementation of an automatic integration not only with the CT scans from the Radiotherapy Department, but more broadly of all the scans archived in the

hospital's PACS, which can improve the workflow integration of the application. Alongside this, implementing a feature that allows physicians to modify the automated segmentation results would provide valuable flexibility, allowing the tool to better align with clinical judgment.

To further validate and improve the tool, conducting comprehensive usability tests in hospital departments is a relevant task. This real-world feedback would be invaluable for refining the GUI and ensuring the tool meets the practical needs of healthcare professionals. Additionally, there is potential in adapting the algorithm to accurately segment other organs, which could greatly expand the HUC's capabilities across a wider range of anatomical structures.

From a technical perspective, several enhancements could push the boundaries of the performance of the models. Implementing more advanced data augmentation techniques could artificially increase dataset diversity, potentially improving model performance, especially for models like FCN and modified BerDiff that showed more sensitivity to data variability. Exploring transfer learning techniques could also yield benefits, particularly in improving performance on smaller or more specialized datasets. Further optimization of the ResUNet++ and U-Net models, focusing on maintaining their high performance while reducing computational requirements, could make these models more accessible in resource-constrained environments.

Advanced ML techniques also offer promising directions for future work. Investigating domain adaptation methods could enhance the ability of the models to generalize across different CT scan sources and protocols, a critical factor in real-world medical imaging scenarios. A longitudinal study assessing how the models perform over time as they are exposed to new, real-world data could provide insights into strategies for continuous learning and model updating. Incorporating explainable AI techniques could offer valuable insights into the decision-making processes of the models, potentially increasing trust and adoption among medical professionals. Lastly, exploring ensemble methods that combine the strengths of different models could push the boundaries of segmentation accuracy and robustness even further.

REFERENCES

- [1] Francque, S. “The Liver and the Cardiovascular System: Two of a Kind?” *J. Am. Heart Assoc.* 2021, 10, e020286
- [2] Trefts, E.; Gannon, M.; Wasserman, D.H. “The liver”. *Curr. Biol.* 2017, 27, R1147–R1151
- [3] Sorensen, A., “Magnetic resonance as a cancer imaging biomarker”. *Journal of clinical oncology* 2006, 24(20):3274–3281
- [4] Mendes, B.; Domingues, I.; Silva, A.; Santos, J. “Prostate Cancer Aggressiveness Prediction Using CT Images”. *Life* 2021, 11, 1164
- [5] Pereira, G.; Domingues, I.; Martins, P.; Abreu, P.H.; Duarte, H.; Santos, J. “Registration of CT with PET: A Comparison of Intensity-Based Approaches”. *International Workshop on Combinatorial Image Analysis (IWCIA)*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 134–149
- [6] Zhang, H.; Luo, K.; Deng, R.; Li, S.; Duan, S. “Deep Learning-Based CT Imaging for the Diagnosis of Liver Tumor”. *Comput. Intell. Neurosci.* 2022, 2022, 1–7
- [7] Vernuccio, F.; Cannella, R.; Bartolotta, T.V.; Galia, M.; Tang, A.; Brancatelli, G. “Advances in liver US, CT, and MRI: moving toward the future”. *Eur. Radiol. Exp.* 2021, 5, 52
- [8] Domingues, I.; Cardoso, J.S. “Using Bayesian surprise to detect calcifications in mammogram images”. *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, USA, 26–30 August 2014; pp. 1091–1094
- [9] Mendes, B.; Domingues, I.; Santos, J. “Multi-class Semantic Segmentation for Prostate Cancer Radiotherapy Treatment Optimization”. *International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMA2SC)*, Porto, Portugal, 27–29 June 2022
- [10] Bechar, M.; Settouti, N.; Domingues, I. “Deep Learning vs. Super Pixel Classification for Breast Masses Segmentation”. *Deep Learning for Biomedical Applications*; CRC Press: Boca Raton, FL, USA, 2021; pp. 121–156
- [11] Oliveira, A.; Domingues, I.; Duarte, H.; Santos, J.; Abreu, P. “Going Back to Basics on Volumetric Segmentation of the Lungs in CT: A Fully Image Processing Based Technique”. *Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*; Springer:

Berlin/Heidelberg, Germany, 2019; Volume 11868 LNCS, pp. 322–334

- [12] Carbone, I.; Martins, P.; Teixeira, A.; Silva, A. “A Vocal Tract Segmentation and Analysis over a European Portuguese MRI Database”. *Electrónica e Telecomunicações* 2008, 4, 1050–1053
- [13] Zhou, L.Q.; Wang, J.Y.; Yu, S.Y.; Wu, G.G.; Wei, Q.; Deng, Y.B.; Wu, X.L.; Cui, X.W.; Dietrich, C.F. “Artificial intelligence in medical imaging of the liver”. *World J. Gastroenterol.* 2019, 25, 672–682
- [14] Chlebus, G.; Schenk, A.; Moltz, J.H.; van Ginneken, B.; Hahn, H.K.; Meine, H. “Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing”. *Sci. Rep.* 2018, 8, 15497
- [15] Christ, P.F.; Ettlinger, F.; Grün, F.; Elshaera, M.E.A.; Lipkova, J.; Schlecht, S.; Ahmaddy, F.; Tataavarty, S.; Bickel, M.; Bilic, P.; et al. “Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks”. *arXiv* 2017, arXiv: 1702.05970
- [16] Ansari, M.Y.; Abdalla, A.; Ansari, M.Y.; Ansari, M.I.; Malluhi, B.; Mohanty, S.; Mishra, S.; Singh, S.S.; Abinahed, J.; Al-Ansari, A.; et al. “Practical utility of liver segmentation methods in clinical surgeries and interventions”. *BMC Med. Imaging* 2022, 22, 97
- [17] Le, D.C.; Chinnasarn, K.; Chansangrat, J.; Keeratibharat, N.; Horkaew, P. “Semi-automatic liver segmentation based on probabilistic models and anatomical constraints”. *Sci. Rep.* 2021, 11, 6106
- [18] Hoffman, J.M., “New advances in brain tumor imaging”. *Current opinion in oncology* 2001, 13(3):148–153
- [19] Boxerman, J. L.; Mosher, T. J.; McVeigh, E. R.; Atalar, E.; Lima, J.; Bluemke, D. A., “Advanced mr imaging techniques for evaluation of the heart and great vessels”. *Radiographics* 1998, 18(3):543–564
- [20] Riederer, S. J., “Recent technical advances in mr imaging of the abdomen”. *Journal of Magnetic Resonance Imaging* 1996, 6(5):822–832
- [21] Niño, S.B.; Bernardino, J.; Domingues, I. “Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective”. *Preprints Sensors* 2024, 24, 1752. <https://doi.org/10.3390/s24061752>
- [22] Batista, S.; Domingues, I.; Couceiro, M.; Filipe, R. “AI-Driven Approaches for Radiation Dose Prediction in Computed Tomography

2D Liver Segmentation in Abdominal Computed Tomography Scans

- Scans”. Portuguese Conference on Pattern Recognition. 2023, 60, 24–25
- [23] Niño, S.B.; Bernardino, J.; Domingues, I. “Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective”. *Sensors* 2024, 24, 1752. doi: 10.20944/preprints202402.0464.v1
- [24] Niño, S.B.; Bernardino, J.; Domingues, I. “Algorithms for Liver Segmentation in Computed Tomography Scans”. *Encyclopedia* 2024. <https://encyclopedia.pub/entry/56271>
- [25] Instituto Superior de Engenharia de Coimbra. “Jornadas de Engenharia Biomédica”, 2024. May 15, 2024. Retrieved from <https://www.isec.pt/PT/eventos/JEB2024>
- [26] ICMASC'24, “International Conference on Mathematical Analysis and Applications in Science and Engineering”, 2024. Porto, Portugal. June 20-22, 2024. Retrieved from <https://www2.isep.ipp.pt/icmasc>
- [27] U.S. Food & Drug (FDA), “Medical Imaging”, 2018. Retrieved October 16, 2023 from <https://www.fda.gov/radiation-emitting-products/radiation-emitting-products-and-procedures/medical-imaging>
- [28] National Institute of Biomedical Imaging and Bioengineering, “Computed Tomography (CT)”, 2024. Retrieved October 16, 2023 from <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct#:~:text=The%20term%20%E2%80%9Ccomputed%20tomography%2C%E2%80%9D,images%2C%20or%E2%80%9Cslices.%E2%80%9D>
- [29] Two Views. “Abdomen or Abdominal CT scans”, 2024. Retrieved October 16, 2023 from <https://www.two-views.com/ct/abdomen-abdominal-scan.html>
- [30] U.S. Food & Drug (FDA). “What is Computed Tomography?”, 2020. Retrieved October 16, 2023 from <https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/what-computed-tomography>
- [31] Brazier, Y. “How does a CT or CAT scan work?”, 2018. Retrieved October 17, 2023 from <https://www.medicalnewstoday.com/articles/153201>
- [32] Prince, J.L.; Links, J.M., “Medical Imaging: Signals and Systems”, 2006. Pearson Prentice Hall.

- [33] Tisch, C., et al. “Hounsfield Unit Values of Liver Pathologies in Unenhanced Post-Mortem Computed Tomography.” *International Journal of Legal Medicine*, vol. 133, no. 6, 2019, pp. 1861-1867
- [34] Medical Imaging Technology Association, “DICOM Standard”, 2024. Retrieved October 23, 2023 from <https://www.dicomstandard.org>
- [35] National Electrical Manufacturers Association, “DICOM Standards Committee”, 2024. Retrieved October 23, 2023 from <https://dicom.nema.org/medical/dicom/current/output/html/part01.html>
- [36] Strickland NH, “PACS: filmless radiology”, *Arch Dis Child*. 2000 Jul;83(1):82-6
- [37] Ebnehoseini, Z.; Tabesh, H.; Deldar, K.; Mostafavi, S.M.; Tara, M., “Determining the Hospital Information System (HIS) Success Rate: Development of a New Instrument and Case Study”, *Open Access Maced J Med Sci*. 2019 May 14;7(9):1407-1414
- [38] IHE International, “Integrating the Healthcare Enterprise”, 2024. Retrieved October 23, 2023 from <https://www.ihe.net>
- [39] Conze, P.-H.; Kavur, A. E.; Gall, E. C.-L.; Gezer, N. S.; Meur, Y. L.; Selver, M. A.; Rousseau, F., “Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks”, *Artificial Intelligence in Medicine*, 117:102109, July 2021
- [40] Furtado, P. “Loss, post-processing and standard architecture improvements of liver deep learning segmentation from computed tomography and magnetic resonance”, *Informatics in Medicine Unlocked*, 24:100585, 2021
- [41] Shuang, Y.; Wang, Z., “A novel approach for automatic and robust segmentation of the 3D liver in computed tomography images”, *Measurement Science and Technology*, 31(11):115701, September 2020
- [42] Kalra, A.; Yetiskul, E.; Wehrle, C.J.; et al., “Physiology, Liver”, 2023. Retrieved October 23, 2023 from <https://www.ncbi.nlm.nih.gov/books/NBK535438>
- [43] National Cancer Institute, “What is Liver Cancer”, 2024. Retrieved October 23, 2023 from <https://www.cancer.gov/types/liver/what-is-liver-cancer>
- [44] Nishikawa, H.; Osaki, Y., “Liver Cirrhosis: Evaluation, Nutritional Status, and Prognosis”, *Mediators Inflamm*. 2015;2015:872152

- [45] Marinho, R., “Conselhos para o bem-estar do fígado”, 2016. Retrieved March 20, 2024 from <https://www.future-healthcare.pt/news2/sp/doenca.html>
- [46] Verissimo, S., “Encefalopatia hepática: o que é, sintomas e tratamento”, 2021. Retrieved March 20, 2024 from <https://www.maisquecuidar.com/encefalopatia-hepatica-o-que-e-sintomas-e-tratamento>
- [47] Directorate General of Health, “National Health Plan 2012-2016”, 2012.
- [48] Wu, E.M.; Wong, L.L.; Hernandez, B.Y.; Ji, J.F.; Jia, W.; Kwee, S.A.; Kalathil, S., “Gender differences in hepatocellular cancer: disparities in nonalcoholic fatty liver disease/steatohepatitis and liver transplantation”, *Hepatoma Res.* 2018;4:66
- [49] Sachdev, P., “What is Liver Failure?”, 2023. Retrieved October 28, 2023 from <https://www.webmd.com/digestive-disorders/digestive-diseases-liver-failure>
- [50] Mayo Clinic, “Liver Disease”, 2024. Retrieved October 29, 2023 from <https://www.mayoclinic.org/diseases-conditions/liver-problems/diagnosis-treatment/drc-20374507>
- [51] Ohri, N.; Dawson, L.A.; Krishnan, S.; et al. “Radiotherapy for Hepatocellular Carcinoma: New Indications and Directions for Future Study”, *J Natl Cancer Inst.* 2016. p. 108
- [52] Tanguturi, S.K.; Wo, J.Y.; Zhu, A.X.; Dawson, L.A.; Hong, T.S., “Radiation therapy for liver tumors: ready for inclusion in guidelines?”, *Oncologist.* 2014;19:868–879
- [53] Heimann, T.; Van, B.; Styner, M.; Arzhaeva, Y.; Aurich, V.; Bauer, C.; Beck, A.; Becker, C.; Beichel, R.; et al., “Comparison and evaluation of methods for liver segmentation from CT datasets”, *IEEE Transactions on Medical Imaging* 2009, 28(8):1251 –1265
- [54] Mharib, A.; Ramli, A.; Mashohor, S.; Mahmood, R., “Survey on liver ct image segmentation methods”, *Artificial Intelligence Review*, 2012, 37(2):83–95
- [55] Punia, R.; Singh, S., “Review on machine learning techniques for automatic segmentation of liver images”, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2013, 3(4):666–670
- [56] Vakalopoulou, M.; Christodoulidis, S.; Burgos, N.; et al., “Deep Learning: Basics and Convolutional Neural Networks (CNNs)”,

Machine Learning for Brain Disorders 2023, New York, NY: Humana; 2023. Chapter 3

- [57] Yang, J.; et al., “A review of action recognition based on Convolutional Neural Network”, J. Phys.: Conf. Ser. 2021, 1827
- [58] Sakib, S.; Ahmed, N.; Kabir, A. J.; Ahmed, H. “An Overview of Convolutional Neural Network: Its Architecture and Applications”, Preprints 2018, 2018110546
- [59] Murphy, J., “An overview of convolutional neural network architectures for deep learning”, Microway Inc, 2016: 1-22
- [60] Adi, Yossi; et al., “Turning your weakness into a strength: Watermarking deep neural networks by backdooring”, 27th USENIX Security Symposium (USENIX Security 18), 2018
- [61] Domino AI, “Ground Truth”, 2024. Retrieved October 29, 2023 from <https://domino.ai/data-science-dictionary/ground-truth>
- [62] Encord, “7 Ways to Improve Medical Imaging Dataset”, 2022. Retrieved October 29, 2023 from <https://encord.com/blog/improve-medical-imaging-dataset-machine-learning>
- [63] Anonymous, “Dataset vs Ground-Truth Dataset”, 2024. Retrieved October 29, 2023 from <https://blog.workaround.vercel.app/blog/dataset-vs-ground-truth-dataset>
- [64] Yamashita, R.; Nishio, M.; Do, R.K.G.; et al., “Convolutional neural networks: an overview and application in radiology”, Insights Imaging 2018, 9, 611–629
- [65] Encord, “Medical Image Segmentation: A Complete Guide”, 2023. Retrieved October 30, 2023 from <https://encord.com/blog/medical-image-segmentation>
- [66] Sakshi, K.V, “Image Segmentation Techniques: Statistical, Comprehensive, Semi-Automated Analysis and an Application Perspective Analysis of Mathematical Expressions”, Archives of Computational Methods in Engineering 2023, 30, 457-495
- [67] Platero, C.; Tobar, M.C., “A multiatlas segmentation using graph cuts with applications to liver segmentation in CT scans”, Comp and mathematical methods in medicine, 2014 (2014)
- [68] Mostafa, A; Fouad, A.; et al., “CT liver segmentation using artificial bee colony optimisation”, Procedia Computer Science, 2015, 1622-1630
- [69] Dou, Q.; Chen, H.; et al., “3D deeply supervised network for automatic liver segmentation from CT volumes”, 19th MICCAI, (2016), 149-157

- [70] Chartrand, G.; Cresson, T.; Chav, R.; Gotra, A.; Tang, A.; et al, “Liver Segmentation on CT and MR Using Laplacian Mesh Optimization,” *IEEE Transactions on Biomedical Engineering*, 2017, vol. 64
- [71] Zhang, Y.; He, Z.; Zhong, C., “Fully Convolutional Neural Network with Post-processing Methods for Automatic Liver Segmentation from CT”, *Chinese Automation Congress (CAC)*, 2017
- [72] Rafiei, S.; et al, “Liver Segmentation in CT Images Using Three Dimensional to Two Dimensional Fully Convolutional Network”, *25th IEEE International Conference on Image Processing (ICIP)* (2018): 2067-2071
- [73] Hiraman, A., “Liver segmentation using 3D CT scans”, *University of Kwazulu-Natal*, 2018
- [74] Ahmad, M.; et al., “Deep belief network modeling for automatic liver segmentation”, *IEEE Access*, 2019;7:20585-95
- [75] Wang, K.; Mamidipalli, A.; et al., “Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network”, *Radiology: AI*, 1(2) (2019), 180022
- [76] Lebre, M., et al., “A robust multi-variability model-based liver segmentation algorithm for CT-scan and MRI modalities”, *Computerized Medical Imaging and Graphics*, 2019
- [77] Almotairi, S; Kareem, G.; et al., “Liver tumor segmentation in CT scans using modified SegNet”, *Sensors*, 20(5) (2020), 1516
- [78] Ayalew, Y.A; Fante, K.A; Mohammed, M.A, “Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method”, *BMC Biomedical Engineering*, 3(1) (2021), 4
- [79] Pettit, R.; Marlatt, B.B.; Corr, S. J.; Havelka, J.; Rana, A., “nnU-Net Deep Learning Method for Segmenting Parenchyma and Determining Liver Volume from Computed Tomography Images,” *Annals of Surgery Open: perspectives of surgical history, education, and clinical approaches*, 3(2), 2022
- [80] Sabir, M., et al., “Segmentation of Liver Tumor in CT Scan Using ResU-Net,” *Applied Sciences* 2022, Vol. 12, Page 8650, vol. 12, no. 17
- [81] Scicluna, D., “Automatic segmentation of healthy liver in abdominal computed tomography scans”, *University of Malta*, 2022
- [82] Olivares, J.; Satpute, N.; Gómez-Luna, J., “Cross-modality-guided contrast enhancement on liver segmentation”, *18th Iberian Conference on Information Systems and Technologies (CISTI)*, 2023
- [83] Ezzat, K.A; Omran, L.N; El Seddawy, A., “Automatic liver segmentation in computed tomography scans using deep semantic

- segmentation”, *Bulletin of Electrical Engineering and Informatics*, 12(1) (2023), 250-256
- [84] Shao, J; Luan, S.; et al., “Attention Connect Network for Liver Tumor Segmentation from CT and MRI Images”, *Technology in Cancer Research and Treatment*, 23, 2024
 - [85] Ying-Hwey, N.; Bernice, W.; Nadya, L.; O'Doherty, S.; et al., “Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset”, *Computers in Biology and Medicine*, Volume 134, 2021, 104497
 - [86] Huynh, N., “Understanding Evaluation Metrics in Medical Image Segmentation”, 2023. Retrieved October 30, 2023 from https://medium.com/@nghihuyh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a373a3f
 - [87] Vorontsov, E.; Cerny, M.; Régnier, P.; et al., “Deep Learning for Automated Segmentation of Liver Lesions at CT in Patients with Colorectal Cancer Liver Metastases”, *Journal Reading*, 2019. Power point presentation
 - [88] Yeghiazaryan, V.; Irina, V., “An Overview of Current Evaluation Methods Used in Medical Image Segmentation”. University of Oxford, Department of Computer Science, 2015
 - [89] Halevi, G.; Moed, H.; Bar-Ilan, J., “Suitability of Google Scholar as a source of scientific information and as a source of data for 612 scientific evaluation - Review of the Literature”, *Journal of Informetrics* 2017, 11, 823–834
 - [90] Al-Saeed, Y.; Gab-Allah, W.; Elmogy, M., “Hepatic tumors diagnosis system based on fuzzy c-means using computed tomography images”, *Research Square Preprint* 2022
 - [91] Soler, L.; Hostettler, A.; Agnus, V.; Charnoz, A.; Fasquel, J.; Moreau, J.; Osswald, A.; Bouhadar, M.; Marescaux, J., “3D Image Reconstruction for Comparison of Algorithm Database: A Patient Specific Anatomical and Medical Image Database”, *Tech. rep*, IRCAD, Strasbourg, France, 2010
 - [92] Kavur, A.E.; Gezer, N.S.; Bars, M.; Aslan, S.; Conze, PH.; Groza, V.; Pham, D.D.; Chatterjee, S.; Ernst, P.; Ozkan, S.; et al., “CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation”, *Medical Image Analysis* 2021, 69, 101950
 - [93] MICCAI-Sliver07, “Segmentation of the Liver Competition 2007 (SLIVER07)”, 2007. Retrieved November 27, 2023 from <https://sliver07.grand-challenge.org>

- [94] Codalab, “LiTS - Liver Tumor Segmentation Challenge”, 2017. Retrieved November 27, 2023 from <https://competitions.codalab.org/competitions/17094>
- [95] Bae, K.T.; Giger, M.L.; Chen, C.T.; Kahn, J., “Automatic segmentation of liver structure in CT images”, *Medical physics* 1993, 20, 71-78
- [96] Gao, L.; Heath, D.G.; Kuszyk, B.S.; Fishman, E.K., “Automatic liver segmentation technique for three-dimensional visualization of CT data”, *Radiology* 1996, 201, 359-364
- [97] Soler, L.; Malandain, G.; Montagnat, J.; Delingette, H.; Ayache, N.; Clément, J.M.; Roy, C.; Russier, Y.; Tassetti, V.; Marescaux, J., “Automatic Segmentation of Portal Vein in CT-Scans of the Liver”, *Proceedings of the World Congress on Medical Physics and Biomedical Engineering (MPBE)*, 1997, p. 788
- [98] Yoo, S.W.; Cho, J.S.; Noh, S.M.; Shin, K.S.; Park, J.W., “Advanced Liver Segmentation by Using Pixel Ratio in Abdominal CT Image”, *Proceedings of the IEEK Conference. The Institute of Electronics and Information Engineers*, 2000, pp. 39-42
- [99] Pan, S.; Dawant, B.M., “Automatic 3D segmentation of the liver from abdominal CT images: a level-set approach”, *Proceedings of the Medical Imaging: Image Processing. SPIE*, 2001, Vol. 4322, pp. 128-138
- [100] Saitoh, T.; Tamura, Y.; Kaneko, T., “Automatic segmentation of liver region through blood vessels on multi-phase CT”, *Proceedings of the International Conference on Pattern Recognition. IEEE*, 2002, Vol. 1, pp. 735-738
- [101] Masumoto, J.; Hori, M.; Sato, Y.; Murakami, T.; Johkoh, T.; Nakamura, H.; Tamura, S., “Automated liver segmentation using multislice CT images”, *Systems and Computers in Japan* 2003, 34, 71-82
- [102] Lim, S.J.; Jeong, Y. Y.; Lee, C.W.; Ho, Y.S., “Automatic segmentation of the liver in CT images using the watershed algorithm based on morphological filtering”, *Proceedings of the Medical Imaging: Image Processing. SPIE*, 2004, Vol. 5370, pp. 1658-1666
- [103] Liu, F.; Zhao, B.; Kijewski, P.K.; Wang, L.; Schwartz, L.H., “Liver segmentation for CT images using GVF snake”, *Medical physics* 2005, 32, 3699-3706
- [104] Lim, S.J.; Jeong, Y.Y.; Ho, Y.S., “Automatic liver segmentation for volume measurement in CT Images”, *Journal of Visual Communication and Image Representation* 2006, 17, 860-875

- [105] Beichel, R.; Bauer, C.; Bornik, A.; Sorantin, E.; Bischof, H., “Liver segmentation in CT data: A segmentation refinement approach”; Proceedings of the 3D Segmentation in The Clinic: A Grand Challenge, 2007, pp. 235-245
- [106] Massoptier, L.; Casciaro, S., “A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans”, European radiology 2008, 18, 1658-1665
- [107] Akram, M.U.; Khanum, A.; Iqbal, K. “An automated system for liver CT enhancement and segmentation”, ICGST Journal of Graphics, Vision and Image Processing (ICGST-GVIP) 2010, 10, 17-22
- [108] Oliveira, D.A.; Feitosa, R.Q.; Correia, M.M., “Segmentation of liver, its vessels and lesions from CT images for surgical planning”, Biomedical engineering online 2011, 10, 1-23
- [109] Linguraru, M.G.; Richbourg, W.J.; Jianfei Liu.; Watt, J.M.; Pamulapati, V.; Shijun Wang.; Summers, R.M., “Tumor Burden Analysis on Computed Tomography by Automated Liver and Tumor Segmentation”, IEEE Transactions on Medical Imaging 2012, 31, 1965-1976
- [110] Li, X.; Luo, S.; Li, J., “Liver Segmentation from CT Image Using Fuzzy Clustering and Level Set”, Journal of Signal and Information Processing 2013, 04, 36-42
- [111] Jena, B.; Krishna Nayak, G.; Saxena, S., “Comprehensive Review of Abdominal Image Segmentation using Soft and Hard Computing Approaches”, Proceedings of the International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020, pp. 1-5
- [112] Maurício, J.; Domingues, I.; Bernardino, J., “Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review”, Applied Sciences 2023, 13
- [113] Priyadarsini, S.; Selvathi, D., “Survey on segmentation of liver from CT images”, Proceedings of the IEEE international conference on advanced communication control and computing technologies (ICACCCI), 2012, pp. 234-238
- [114] Khan, F., “Automated segmentation of CT liver images: a review”, Journal of Communications Technology, Electronics and Computer Science 2018, 19, 5-9
- [115] Moghbel, M.; Mashohor, S.; Mahmud, R.; Saripan, M.I.B., “Review of liver segmentation and computer assisted detection/diagnosis methods

2D Liver Segmentation in Abdominal Computed Tomography Scans

- in computed tomography”, *Artificial Intelligence Review* 2018, 50, 497-537
- [116] Vanmore, S.V.; Chougule, S.R., “Survey on automatic liver segmentation techniques from abdominal CT images”, *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICCS)*. IEEE, 2019, pp. 1030-1035
- [117] Krizhevsky, A.; Sutskever, I.; Hinton, G.E., “Imagenet classification with deep convolutional neural networks”, *Advances in neural information processing systems* 2012, 25
- [118] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al., “ImageNet Large Scale Visual Recognition Challenge”, *International Journal of Computer Vision (IJCV)* 2015, 115, 211-252
- [119] Ronneberger, O.; Fischer, P.; Brox, T., “U-Net: Convolutional Networks for Biomedical Image Segmentation”, *Proceedings of 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 234-241, 2015
- [120] Mir, R.; Kelly, S.M.; Xiao, Y.; Moore, A.; Clark, C.H.; Clementel, E.; Corning, C.; Ebert, M.; Hoskin, P.; Hurkmans, C.W.; et al., “Organ at risk delineation for radiation therapy clinical trials: Global Harmonization Group consensus guidelines”, *Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology* 2020, 150, 30-39
- [121] Sarria, G.R.; Kugel, F.; Roehner, F.; Layer, J.; Dejonckheere, C.; Scafa, D.; Koeksal, M.; Leitzen, C.; Schmeel, L.C., “Artificial Intelligence-Based Autosegmentation: Advantages in Delineation, Absorbed Dose-Distribution, and Logistics”, *Advances in Radiation Oncology* 2024, 9, 101394
- [122] Therapanacea, “Annotate AI-based automatic delineation”, 2024. Retrieved March 26, 2024 from <https://www.therapanacea.eu/our-products/annotate/#annotate>
- [123] U.S. Food & Drug (FDA), “TheraPanacea 510(k) Premarket Notification”, 2021. Retrieved June 10, 2024 from <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K202700>
- [124] MVision AI, “Contour+ Guideline-Based AI Segmentation”, 2024. Retrieved March 26, 2024 from <https://mvision.ai/contour>
- [125] U.S. Food & Drug (FDA), “Contour+ 510(k) Premarket Notification”, 2022. Retrieved June 10, 2024

<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K212915>

- [126] Limbus AI, “Limbus Contour Automatic Contouring for Radiation Therapy”, 2024. Retrieved March 26, 2024 from <https://limbus.ai>
- [127] U.S. Food & Drug (FDA), “Limbus AI 510(k) Premarket Notification”, 2023. Retrieved June 10, 2024 <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K230575>
- [128] Frączek, Marcin, et al. “The usefulness of free dicom viewers in rhinologic practice.” *Wiadomosci lekarskie* (Warsaw, Poland : 1960) vol. 69,3 pt 2 (2016): 597-601
- [129] Hamer, P., “Laparoscopic Cholecystectomy”, 2024. Retrieved March 29, 2024 from <https://www.drpeterhamer.com.au/general-surgery/laparoscopic-cholecystectomy/>
- [130] VARIAN, “Sistema de informação de oncologia ARIA”, 2024, Retrieved March 29, 2024 from <https://www.varian.com/pt-pt/products/software/digital-oncology/oncology-management-systems/aria-oncology-information-system>
- [131] Schmid, S., “Visualization of Medical Images: Adjusting Contrast through Windowing”, 2023. Retrieved May 20, 2024 from <https://medium.com/@susanne.schmid/visualization-of-medical-images-adjusting-contrast-through-windowing-c2dd9abb1d5>
- [132] Virdee, A., “A Matter of Grayscale: Understanding Dicom Windows”, 2020. Retrieved May 20, 2024 from <https://towardsdatascience.com/a-matter-of-grayscale-understanding-dicom-windows-1b44344d92bd>
- [133] scikit-image, “skimage.exposure”, n.d. Retrieved May 22, 2024 from <https://scikit-image.org/docs/stable/api/skimage.exposure.html#>
- [134] Voronyansky, Y., “Digital Image Processing | Histogram Calculation, Equalization and Normalization”, 2023. Retrieved May 22, 2024 from <https://vrnsky.medium.com/digital-image-processing-histogram-calculation-equalization-and-normalization-ad09b0bba5b1>
- [135] Desternes, J., “Data augmentation for medical image analysis in deep learning”, 2022. Retrieved May 22, 2024 from <https://www.imaio.com/en/resources/blog/ai-for-medical-imaging-data-augmentation>
- [136] Jha, Debesh, et al. "ResUNet++: An Advanced Architecture for Medical Image Segmentation." *arXiv preprint arXiv:1911.07067* (2019)

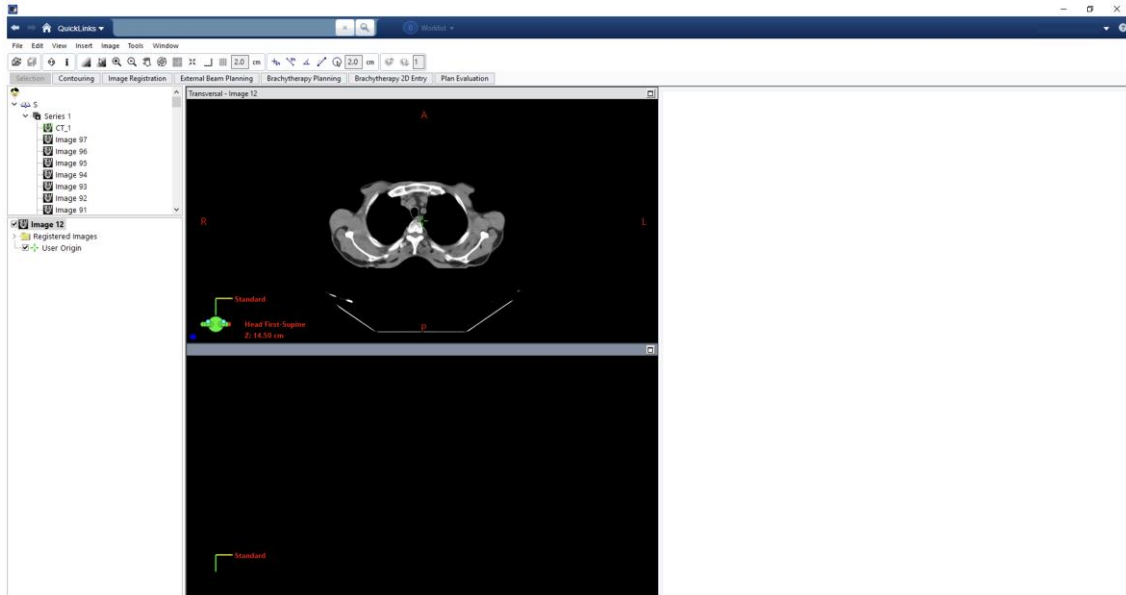
- [137] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440
- [138] Chen, T.; Chenhui, W.; Hongming, S., "BerDiff: Conditional Bernoulli Diffusion Model for Medical Image Segmentation", Medical Image Computing and Computer Assisted Intervention – MICCAI, 2023, pp. 491–501
- [139] Kareliotis, G.; "Study of kVp and mAs Effect on Radiation Dose and Image Quality in Computed Tomography." 2015

APPENDICES

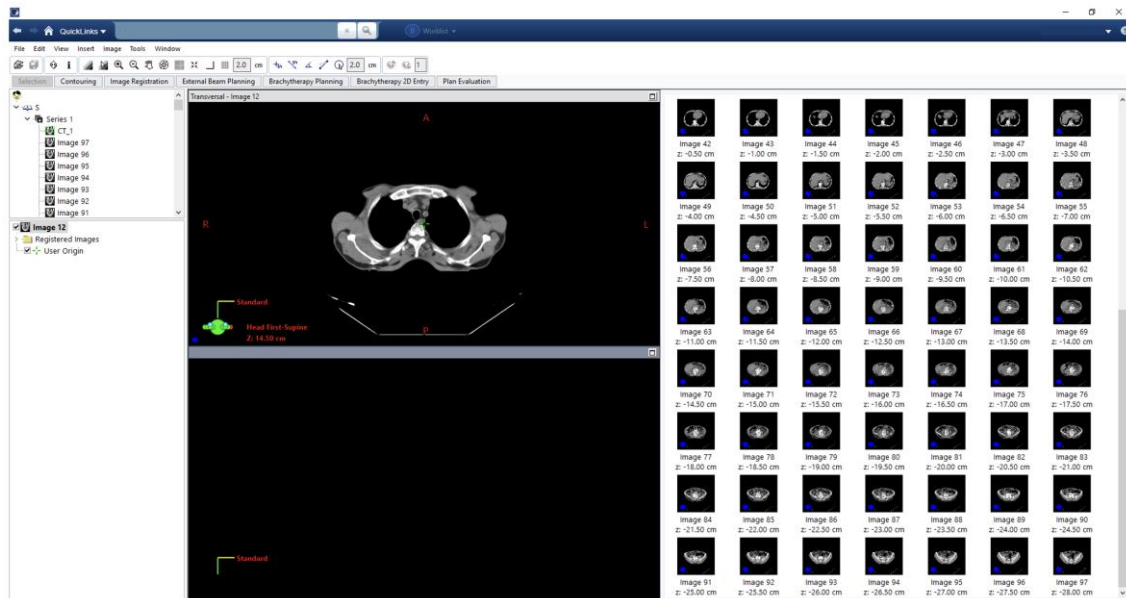
Appendix A– Scans Exportation Strategy

The steps of the strategy chosen for the export process are presented below:

1. Search for the patient ID in the table of all scans that can be used
2. Select scan

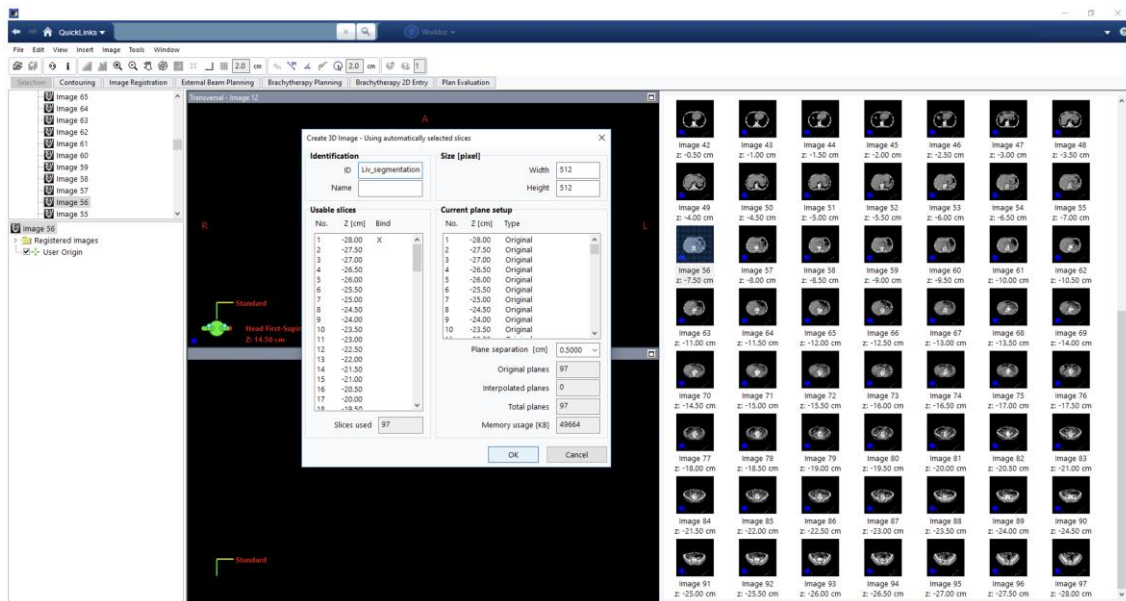


3. Drag the scan to the right



4. Select one of the images on the right
5. Click **Insert > New 3D image**

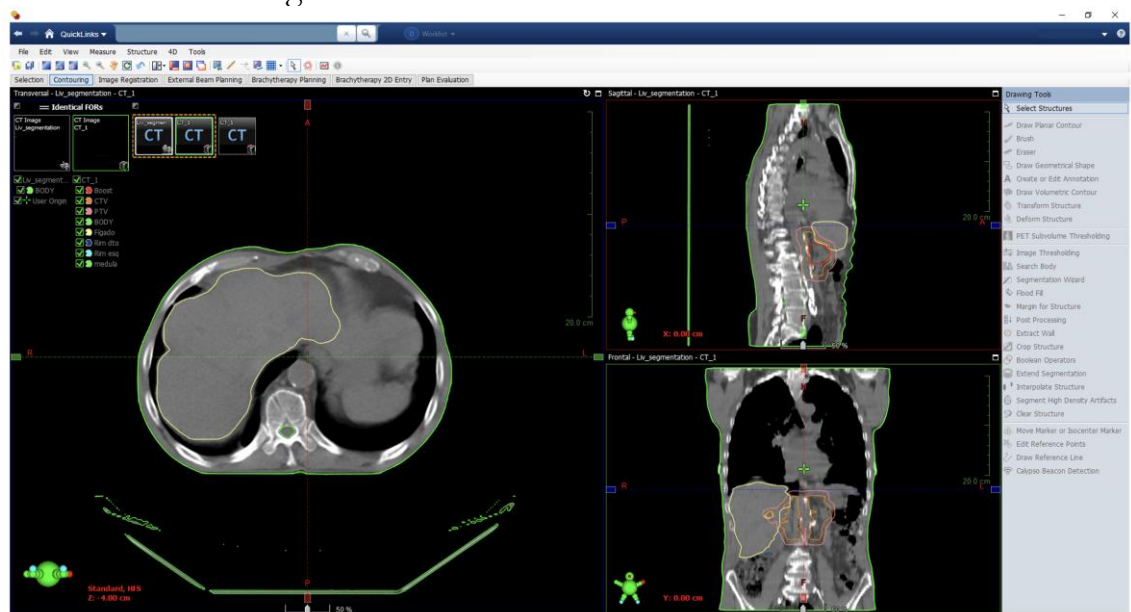
6. Define ID to identify new 3D image



7. Click **Contouring** at the top

8. Confirm left side has the new 3D image created

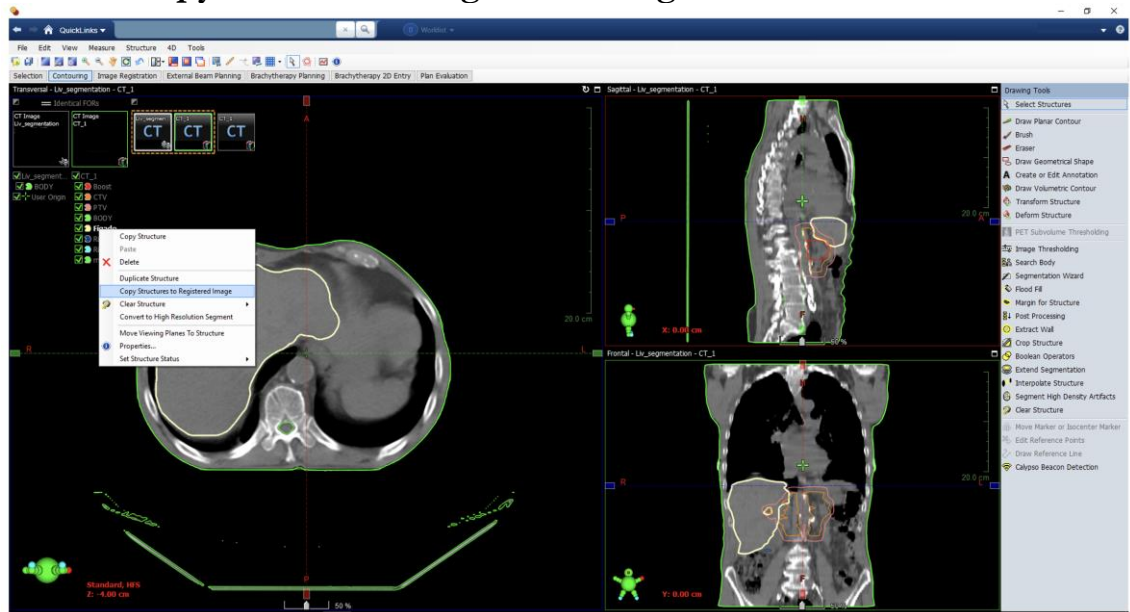
9. Choose the 3D image with structures identified



10. Right-click on **Liver** in the image structures

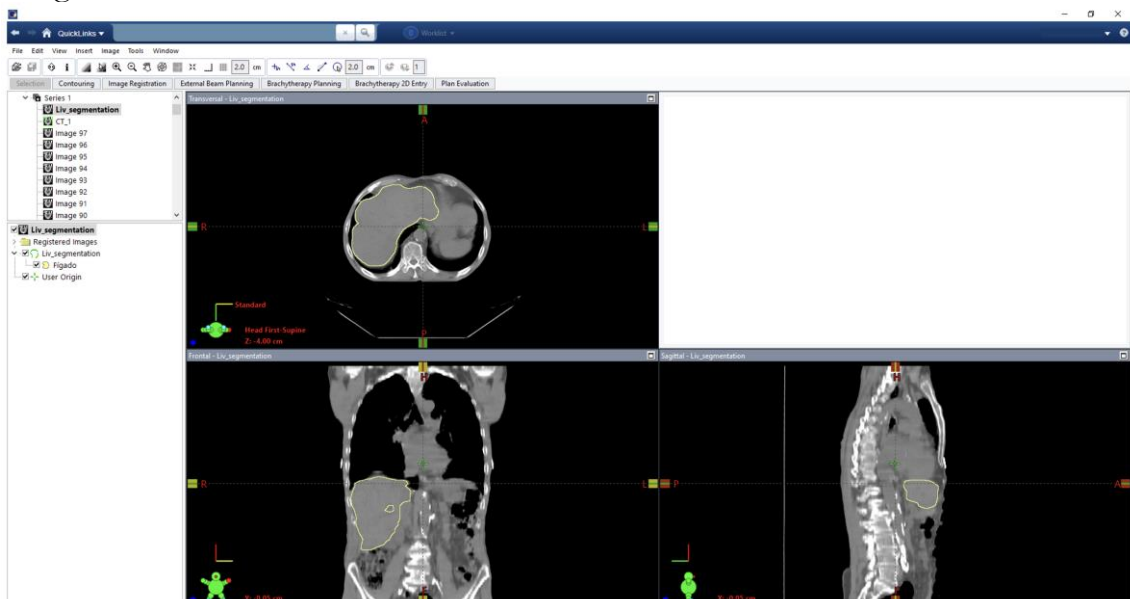
2D Liver Segmentation in Abdominal Computed Tomography Scans

11. Click on **Copy structures to registered image**



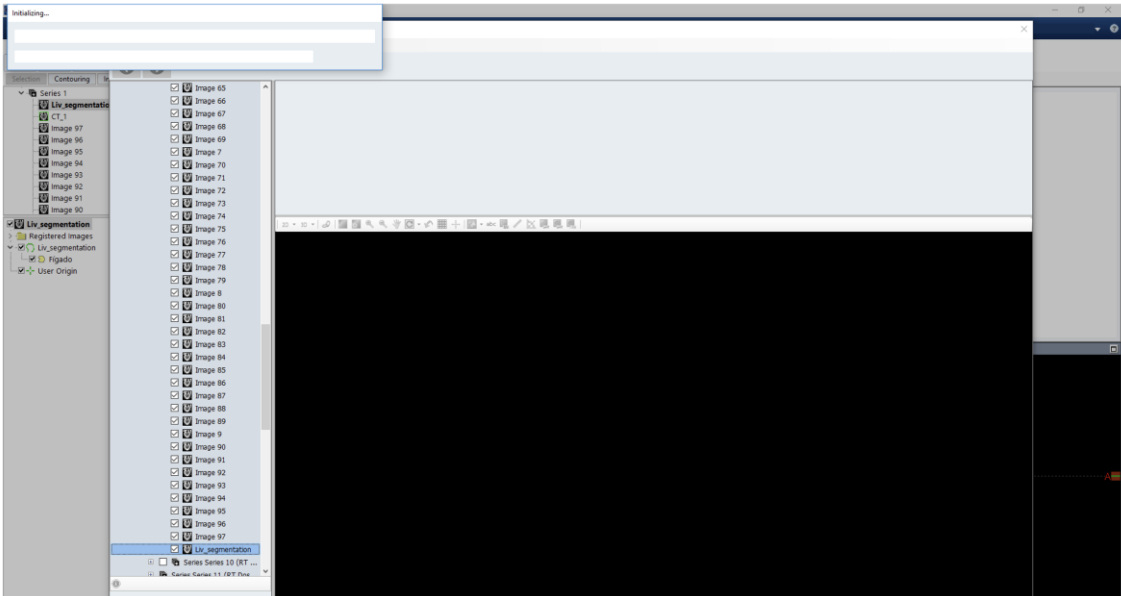
12. Right-click on the **Body** in the new 3D image on the right to delete this structure

13. Click **Selection** and **Save** to save the liver contouring for the new 3D image

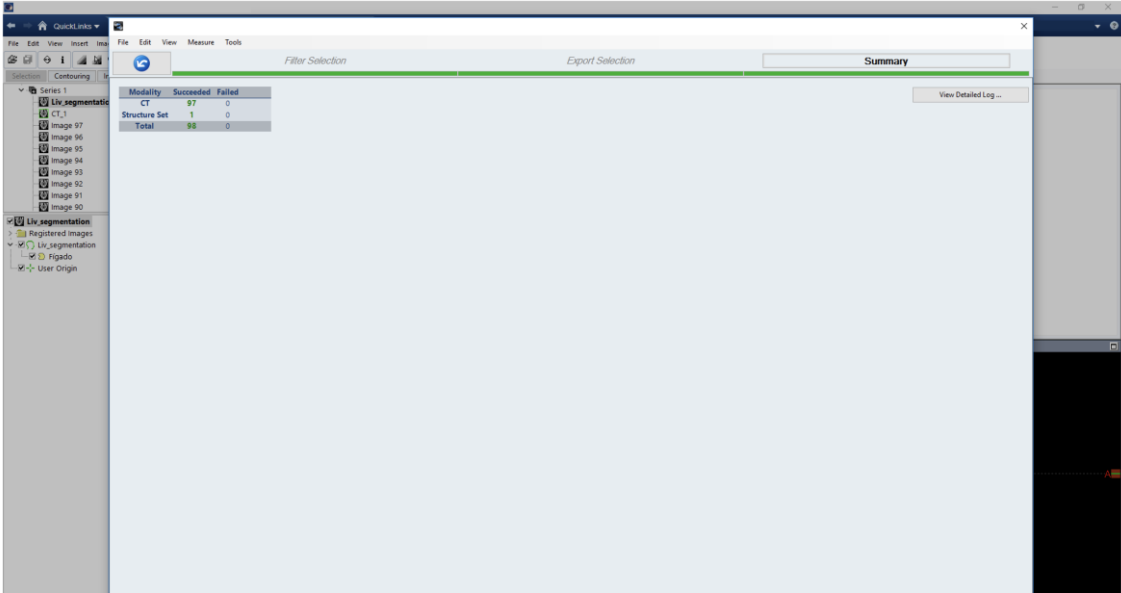


14. Click **File, Export** and choose folder

15. Confirm selection



16. Export selection



Appendix B – SQL Query for Scans Retrieval

The first step for creating the dataset involved defining the Structured Query Language (SQL) query for ARIA's MSSQL database presented below:

```
SELECT
TOP 1000
pat.PatientId
,c.[CourseID]
,c.[StartDateTime]
FROM
[VARIAN].[dbo].[Course] c
JOIN
[VARIAN].[dbo].Patient as pat
on c.PatientSer = pat.PatientSer
WHERE
c.CourseID LIKE '%[eE][sS][tT][oO]%' OR c.CourseID LIKE
'%[Pp][âÂaA][nN]%' OR c.CourseID LIKE '%[Ff][iIí][gG]%' OR
c.CourseID LIKE '%[Ll][iI][vV][eE]%' OR c.CourseID LIKE
'%[Hh][eE][pP]%'
OR c.CourseID LIKE '%[Bb][Ii][lL][iI]%' OR c.CourseID LIKE
'%[Ee][sS][oOóÓ][fF]%'
OR c.CourseID LIKE '%[Rr][Ii][mM]%' OR c.CourseID LIKE
'%[pP][uU][lL][mM]%'
ORDER BY
c.PatientSer DESC
```


Appendix C – Script for Scans Anonymization

```
#!/bin/bash

#

#
https://dicom.nema.org/medical/dicom/current/output/chtml/part15/chapter_E.html

ID=`openssl rand -hex 8`

DATE="20240201"
TIME="123947.318695"

dcm_modify () {
dcmmodify -nb -imt -ma "(0008,0050)=ANON" -ea "(0008,0012)" -ea
"(0008,0013)" \
    -ea "(0008,0015)" -ea "(0008,0021)" -ea "(0008,0022)" -ea "(0008,0024)" -ea
"(0008,0025)" \
    -ea "(0008,002A)" -ea "(0008,0031)" -ea "(0008,0032)" -ea "(0008,0034)" -ea
"(0008,0035)" \
    -ea "(0008,0054)" -ea "(0008,0055)" -ea "(0008,0080)" -ea "(0008,0081)" -ea
"(0008,0082)" \
    -ea "(0008,0096)" -ea "(0008,009D)" -ea "(0008,0201)" -ea "(0008,1010)" -ea
"(0008,1030)" \
    -ea "(0008,103E)" -ea "(0008,1040)" -ea "(0008,1041)" -ea "(0008,1048)" -ea
"(0008,1049)" \
    -ea "(0008,1050)" -ea "(0008,1052)" -ea "(0008,1060)" -ea "(0008,1062)" -ea
"(0008,1070)" \
    -ea "(0008,1072)" -ea "(0008,1080)" -ea "(0008,1084)" -ea "(0008,1088)" -ea
"(0008,2111)" \
    -ea "(0010,0021)" -ea "(0010,0032)" -ea "(0010,0050)" -ea "(0010,0101)" -ea
"(0010,0102)" \
    -ea "(0010,1000)" -ea "(0010,1001)" -ea "(0010,1002)" -ea "(0010,1010)" -ea
"(0010,1020)" \
    -ea "(0010,1030)" -ea "(0010,1100)" -ea "(0010,2160)" -ea "(0010,2180)" -ea
"(0010,21B0)" \
}
```

-ea "(0010,2203)" -ea "(0010,2297)" -ea "(0010,2299)" -ea "(0010,4000)" -ea
"(0012,0051)" \
-ea "(0012,0071)" -ea "(0012,0072)" -ea "(0012,0082)" -ea "(0012,0086)" -ea
"(0012,0087)" \
-ea "(0016,002B)" -ea "(0016,004B)" -ea "(0016,004D)" -ea "(0016,004E)" -
ea "(0016,004F)" \
-ea "(0016,0050)" -ea "(0016,0051)" -ea "(0016,0070)" -ea "(0016,0071)" -ea
"(0016,0072)" \
-ea "(0016,0073)" -ea "(0016,0074)" -ea "(0016,0075)" -ea "(0016,0076)" -ea
"(0016,0077)" \
-ea "(0016,0078)" -ea "(0016,0079)" -ea "(0016,007A)" -ea "(0016,007B)" -ea
"(0016,007C)" \
-ea "(0016,007D)" -ea "(0016,007E)" -ea "(0016,007F)" -ea "(0016,0080)" -ea
"(0016,0081)" \
-ea "(0016,0082)" -ea "(0016,0083)" -ea "(0016,0084)" -ea "(0016,0085)" -ea
"(0016,0086)" \
-ea "(0016,0087)" -ea "(0016,0088)" -ea "(0016,0089)" -ea "(0016,008A)" -ea
"(0016,008B)" \
-ea "(0016,008C)" -ea "(0016,008D)" -ea "(0016,008E)" -ea "(0018,0027)" -ea
"(0018,0035)" \
-ea "(0018,1000)" -ea "(0018,1004)" -ea "(0018,1005)" -ea "(0018,1007)" -ea
"(0018,1008)" \
-ea "(0018,1009)" -ea "(0018,100A)" -ea "(0018,1012)" -ea "(0018,1014)" -ea
"(0018,1030)" \
-ea "(0018,1042)" -ea "(0018,1043)" -ea "(0018,1072)" -ea "(0018,1073)" -ea
"(0018,1078)" \
-ea "(0018,1079)" -ea "(0018,1200)" -ea "(0018,1201)" -ea "(0018,1204)" -ea
"(0018,1205)" \
-ea "(0018,1400)" -ea "(0018,5011)" -ea "(0018,700A)" -ea "(0018,700C)" -ea
"(0018,700E)" \
-ea "(0018,9185)" -ea "(0018,9373)" -ea "(0018,937B)" -ea "(0018,937F)" -ea
"(0018,9424)" \
-ea "(0018,9516)" -ea "(0018,9517)" -ea "(0018,9937)" -ea "(0018,A002)" -ea
"(0018,A003)" \
-ea "(0020,0027)" -ea "(0020,4000)" -ea "(0020,9158)" -ea "(0032,1060)" -ea
"(0032,1066)" \

2D Liver Segmentation in Abdominal Computed Tomography Scans

-ea "(0032,1067)" -ea "(0038,0010)" -ea "(0038,0011)" -ea "(0038,0014)" -ea "(0038,0060)" \

-ea "(0038,0061)" -ea "(0038,0062)" -ea "(0038,0064)" -ea "(003A,0329)" -ea "(003A,032B)" \

-ea "(0040,0007)" -ea "(0040,0009)" -ea "(0040,0244)" -ea "(0040,0245)" -ea "(0040,0250)" \

-ea "(0040,0251)" -ea "(0040,0253)" -ea "(0040,0254)" -ea "(0040,0275)" -ea "(0040,0280)" \

-ea "(0040,0310)" -ea "(0040,051A)" -ea "(0040,0555)" -ea "(0040,0600)" -ea "(0040,0602)" \

-ea "(0040,1002)" -ea "(0040,100A)" -ea "(0040,1102)" -ea "(0040,1103)" -ea "(0040,1104)" \

-ea "(0040,A032)" -ea "(0040,A033)" -ea "(0040,A078)" -ea "(0040,A07A)" -ea "(0040,A07C)" \

-ea "(0044,0105)" -ea "(0050,001B)" -ea "(0050,0020)" -ea "(0050,0021)" -ea "(006A,0006)" \

-ea "(0070,0082)" -ea "(0070,0083)" -ea "(0070,0086)" -ea "(0088,0200)" -ea "(0100,0420)" \

-ea "(0400,0310)" -ea "(0400,0402)" -ea "(0400,0403)" -ea "(0400,0404)" -ea "(0400,0561)" \

-ea "(0400,0600)" -ea "(2200,0002)" -ea "(2200,0005)" -ea "(3002,0121)" -ea "(3002,0123)" \

-ea "(3006,0004)" -ea "(3006,0006)" -ea "(3006,0028)" -ea "(3006,002D)" -ea "(3006,002E)" \

-ea "(3006,0038)" -ea "(3006,004D)" -ea "(3006,004E)" -ea "(3006,0085)" -ea "(3006,0088)" \

-ea "(3008,0054)" -ea "(3008,0056)" -ea "(3008,0105)" -ea "(3008,0250)" -ea "(3008,0251)" \

-ea "(300A,0003)" -ea "(300A,0004)" -ea "(300A,0006)" -ea "(300A,0007)" -ea "(300A,000E)" \

-ea "(300A,0016)" -ea "(300A,0072)" -ea "(300A,00B2)" -ea "(300A,00C3)" -ea "(300A,00DD)" \

-ea "(300A,0196)" -ea "(300A,01A6)" -ea "(300A,01B2)" -ea "(300A,0216)" -ea "(300A,02EB)" \

-ea "(300A,0676)" -ea "(300A,078E)" -ea "(300A,0792)" -ea "(300A,0794)" -ea "(300A,079A)" \

```

-ea "(300C,0113)" -ea "(300E,0008)" -ea "(3010,0036)" -ea "(3010,0037)" -ea
"(3010,004C)" \
-ea "(3010,004D)" -ea "(3010,0056)" -ea "(3010,0061)" -ea "(3010,0077)" -ea
"(3010,0085)" \
-ea "(FFFA,FFFA)" -ea "(FFFC,FFFC)" \
-ma      "(0008,0020)=$DATE"      -ma      "(0008,0023)=$DATE"      -ma
"(0008,0030)=$TIME" \
-ma      "(0008,0033)=$TIME"      -ma      "(0008,0090)=ANON"      -ma
"(0008,009C)=ANON" \
-ma      "(0008,0106)=ANON"      -ma      "(0008,0107)=ANON"      -ma
"(0010,0010)=$ID" \
-ma      "(0010,0020)=$ID"      -ma      "(0010,0030)=$DATE"      -ma
"(0010,0040)=ANON" \
-ma      "(0012,0010)=ANON"      -ma      "(0012,0020)=ANON"      -ma
"(0012,0021)=ANON" \
-ma      "(0012,0030)=ANON"      -ma      "(0012,0031)=ANON"      -ma
"(0012,0040)=ANON" \
-ma      "(0012,0042)=ANON"      -ma      "(0012,0050)=ANON"      -ma
"(0012,0060)=ANON" \
-ma      "(0012,0081)=ANON"      -ma      "(0018,0010)=ANON"      -ma
"(0018,11BB)=ANON" \
-ma      "(0018,1203)=ANON"      -ma      "(0018,9074)=ANON"      -ma
"(0018,9151)=ANON" \
-ma      "(0018,9367)=ANON"      -ma      "(0018,9369)=ANON"      -ma
"(0018,936A)=ANON" \
-ma      "(0018,9371)=ANON"      -ma      "(0018,9623)=ANON"      -ma
"(0018,9701)=ANON" \
-ma      "(0018,9804)=ANON"      -ma      "(0018,9919)=ANON"      -ma
"(0020,0010)=ANON" \
-ma      "(0034,0001)=ANON"      -ma      "(0034,0002)=ANON"      -ma
"(0034,0005)=ANON" \
-ma      "(0034,0007)=ANON"      -ma      "(003A,0314)=ANON"      -ma
"(0040,0512)=ANON" \
-ma      "(0040,0513)=ANON"      -ma      "(0040,0551)=ANON"      -ma
"(0040,0562)=ANON" \
-ma      "(0040,0610)=ANON"      -ma      "(0040,1101)=ANON"      -ma
"(0040,2016)=ANON" \

```


2D Liver Segmentation in Abdominal Computed Tomography Scans

-ma "(0040,2017)=ANON" "(0040,A030)=ANON" \	-ma "(0040,A027)=ANON" -ma
-ma "(0040,A073)=ANON" "(0040,A082)=ANON" \	-ma "(0040,A075)=ANON" -ma
-ma "(0040,A088)=ANON" "(0040,A121)=ANON" \	-ma "(0040,A120)=ANON" -ma
-ma "(0040,A122)=ANON" "(0040,A13A)=ANON" \	-ma "(0040,A123)=ANON" -ma
-ma "(0040,A730)=ANON" "(0044,0104)=ANON" \	-ma "(0042,0011)=ANON" -ma
-ma "(0068,6226)=ANON" "(006A,0005)=ANON" \	-ma "(0068,6270)=ANON" -ma
-ma "(0070,0001)=ANON" "(0072,000A)=ANON" \	-ma "(0070,0084)=ANON" -ma
-ma "(0072,005E)=ANON" "(0072,0061)=ANON" \	-ma "(0072,005F)=ANON" -ma
-ma "(0072,0063)=ANON" "(0072,0066)=ANON" \	-ma "(0072,0065)=ANON" -ma
-ma "(0072,0068)=ANON" "(0072,006B)=ANON" \	-ma "(0072,006A)=ANON" -ma
-ma "(0072,006C)=ANON" "(0072,006E)=ANON" \	-ma "(0072,006D)=ANON" -ma
-ma "(0072,0070)=ANON" "(0400,0105)=ANON" \	-ma "(0072,0071)=ANON" -ma
-ma "(0400,0115)=ANON" "(0400,0563)=ANON" \	-ma "(0400,0562)=ANON" -ma
-ma "(0400,0564)=ANON" "(2100,0140)=ANON" \	-ma "(0400,0565)=ANON" -ma
-ma "(3006,0002)=ANON" "(3006,0009)=\$TIME" \	-ma "(3006,0008)=\$DATE" -ma
-ma "(3006,00A6)=ANON"	-ma "(3008,0024)=ANON" \
-ma "(3008,0025)=ANON" "(3008,0164)=ANON" \	-ma "(3008,0162)=ANON" -ma
-ma "(3008,0166)=ANON" "(300A,0002)=ANON" \	-ma "(3008,0168)=ANON" -ma
-ma "(300A,022C)=ANON" "(300A,0608)=ANON" \	-ma "(300A,022E)=ANON" -ma

```

-ma "(300A,0611)=ANON" -ma "(300A,0615)=ANON" -ma
"(300A,0619)=ANON" \
-ma "(300A,0623)=ANON" -ma "(300A,062A)=ANON" -ma
"(300A,067C)=ANON" \
-ma "(300A,067D)=ANON" -ma "(300A,0734)=ANON" -ma
"(300A,0736)=ANON" \
-ma "(300A,073A)=ANON" -ma "(300A,0741)=ANON" -ma
"(300A,0742)=ANON" \
-ma "(300A,0760)=ANON" -ma "(300A,0783)=ANON" -ma
"(300C,0127)=ANON" \
-ma "(300E,0004)=$DATE" -ma "(300E,0005)=$TIME" -ma
"(3010,000F)=ANON" \
-ma "(3010,0017)=ANON" -ma "(3010,001B)=ANON" -ma
"(3010,002D)=ANON" \
-ma "(3010,0033)=ANON" -ma "(3010,0034)=ANON" -ma
"(3010,0035)=ANON" \
-ma "(3010,0038)=ANON" -ma "(3010,0043)=ANON" -ma
"(3010,0054)=ANON" \
-ma "(3010,005A)=ANON" -ma "(3010,005C)=ANON" -ma
"(3010,007A)=ANON" \
-ma "(3010,007B)=ANON" -ma "(3010,007F)=ANON" -ma
"(3010,0081)=ANON" \
$1
}

```

```
# -ma "(3006,0026)=ANON"
```

```

BASE_DIR=`dirname $1`
STORE_DIR="$BASE_DIR/$ID"

```

```

if [ -d $1 ]; then
    mv $1 $STORE_DIR
    cd $STORE_DIR
fi

```

2D Liver Segmentation in Abdominal Computed Tomography Scans

```
SAVEIFS=$IFS
```

```
IFS=$(echo -en "\n\b")
```

```
for file in `ls`; do
```

```
# echo "$file"
```

```
  dcm_modify "$file"
```

```
done
```

```
# restore $IFS
```

```
IFS=$SAVEIFS
```


Appendix D – RECPAD 2023 Publication

Poster presentation at *Portuguese Conference on Pattern Recognition (RECPAD 2023)* with the initial project proposal (**Artificial Intelligence-driven Approaches for Radiation Dose Prediction in Computed Tomography Scans**):

AI-Driven Approaches for Radiation Dose Prediction in Computed Tomography Scans

Stephanie Batista¹
stephanieabn@gmail.com

Inês Domingues^{1,2}
inesdomingues@gmail.com

Miguel Couceiro^{1,3,4}
couceiro@isec.pt

Ricardo Filipe⁵
ricardo.angelo.filipe@gmail.com

¹ Polytechnic Institute of Coimbra, Coimbra Institute of Engineering,
Rua Pedro Nunes - Quinta da Nora, 3030-199 Coimbra, Portugal.

² IPO Porto Research Center (CIPOP): Medical Physics,
Radiobiology & Radiation Protection Research Group, Portugal

³ Institute of Applied Research (I2A), Portugal

⁴ Laboratory for High Performance Computing (LaCED), Portugal

⁵ Altice Labs, S.A, Portugal

Abstract

Optimizing radiation doses for computed tomography (CT) scans is essential to ensure patient safety by minimizing potential health risks associated with ionizing radiation exposure. The necessity arises from the key role of CT scans in disease diagnosis and monitoring, balanced against the inherent radiation hazards they pose. To address this need, an innovative project is underway, proposing the application of Artificial Intelligence (AI) models to predict ionizing radiation doses within CT scans, effectively refining radiological practices and minimizing associated risks. The project aims to enhance healthcare quality by aligning doses with Diagnostic Reference Levels (DRLs) in Portugal. The approach employed is personalized, adapting imaging protocols based on individual patient attributes, following the ALARA (As Low as Reasonably Achievable) principle, aimed at achieving the minimal radiation exposure necessary for diagnosis without compromising accuracy. By leveraging technology, the project empowers healthcare professionals with efficient data analysis, facilitating precise diagnoses and effective treatments. This initiative contributes to compliance with radiological protection guidelines, contributing to Portugal's commitment to radiological protection, while healthcare professionals diligently adhere to medical protocols. These collaborative actions prioritize patient welfare and improve healthcare safety by mitigating the risks of ionizing radiation and promoting a technologically advanced healthcare environment.

1 Introduction

Radiology exams play a crucial role in the diagnosis and monitoring of diseases, but they also expose patients to ionizing radiation, which can pose health risks if not properly controlled. In Portugal, medical radiological exposures constitute the main source of artificial ionizing radiation exposure for citizens, making it essential to ensure radiological protection of patients during CT scans. To this end, strict guidelines and regulations have been adopted, including [3], [4], [6] and [7], which emphasize the importance of protecting patients from ionizing radiation in medical radiological exposures.

When it comes to the application of Medical Physics in Portugal, especially in CT scans for diagnosing and monitoring diseases, it is regulated by international regulations and national legislation. Organizations such as the International Commission on Radiological Protection (ICRP), the International Atomic Energy Agency (IAEA) and the European Commission establish fundamental guidelines for radiological protection and the safety of ionizing radiation equipment [11][12]. Portuguese legislation establishes a comprehensive legal framework for radiological protection in Portugal, with specific ordinances detailing technical standards and requirements. This regulatory framework guarantees the safety, efficacy and quality of radiological procedures, protecting workers, patients and the public from the risks associated with exposure to ionizing radiation [1].

The proposed project aims to apply Artificial Intelligence (AI) models to predict the dose of ionizing radiation to which patients are exposed during CT scans, providing relevant information to improve radiological practices and minimize the risks associated with CT scans. With this project, the goal is to compare the performance of various models (decision trees, neural networks, support vector machines, random forest, etc.) in dose prediction, considering that currently the dose is determined manually by healthcare professionals or automatically by the equipment during CT scans [5]. Therefore, this project can be of great utility in validating the radiation dose that is defined to be administered to the patient.

By predicting the radiation dose, this project can significantly contribute to the improvement of the quality of healthcare services provided to patients in Portugal, optimizing radiation doses and ensuring they are within the DRLs. This allows for the adaptation of imaging protocols to each patient, considering factors such as age, weight, and medical history. This personalized approach ensures that patients receive

the lowest possible radiation dose without compromising the quality of image diagnosis, following the ALARA principle [10].

Additionally, it provides healthcare professionals with greater efficiency in data analysis and interpretation, resulting in more accurate diagnoses and more effective treatments. Thus, the project represents a significant technological advancement in the healthcare field in Portugal, transforming how healthcare professionals deal with the evaluation of data related to ionizing radiation dose in CT scans.

The structure of the paper is composed in section 2 by an overview of the proposal that will be implemented. Section 3 presents some of the fundamental dosimetry quantities that might be used in the project, which are essential for establishing radiological protection principles and systems for both external and internal exposure of biological tissues to radiation. In section 4, it is presented the conclusions of the proposal and, lastly, the main references used to prepare the article are provided.

2 Approach

This study aims to develop a strategy for predicting ionizing radiation doses in CT scans by applying AI models. At an early stage, models such as Support Vector Machines (SVM), Random Forest, Artificial Neural Networks and Decision Trees will be explored, bearing in mind that a selection of models will subsequently be made based on their ability to handle complex medical data and make accurate predictions in the context of the project.

The strategy will be based on a large set of data from CT scans (examination example shown in Figure 1) stored in the Picture Archiving and Communication System (PACS) from a Portuguese hospital [2][8]. To ensure compliance with the General Data Protection Regulation (GDPR), an anonymization procedure will be applied during a pre-processing stage of the data, selecting those features that might allow patients to be identified, with the purpose of guaranteeing patient protection and safety in all circumstances.



Figure 1: Chest CT: lung window (a) and soft tissue window (b) [13].

As a first approach to dose prediction, features such as age, weight, and medical history, along with imaging parameters, will be used to estimate the radiation dose that would be delivered during the scan, also analyzing the relationship between different features, as well as avoiding redundancy of information in features that present the same information, to allow the recognition of patterns in patient medical information.

Once trained with a vast set of known data, these models will be able to accurately estimate radiation doses for new CT scans, helping to optimize imaging protocols, personalize patient care and minimize risks associated with exposure to ionizing radiation. The most appropriate models will be selected later, considering the specific objectives of the study and the clinical context (Figure 2).

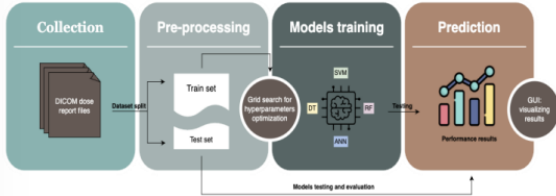


Figure 2: System approach overview.

In a conclusive phase, the theoretical values of several fundamental quantities in the field of Medical Physics are calculated to complement the prediction value of the models (see Section 3), which are finally compared with the previously defined DRLs.

DRLs are predetermined values used to assess radiation dose in radiological examinations, like CT scans. They optimize procedures, ensuring diagnostic quality while keeping patient exposure within acceptable limits. DRLs are crucial for radioprotection, monitoring and controlling radiation dose to minimize risks.

In the context of CT, DRLs are established based on protocols considering aspects such as technique, body region, and patient age. Standardization improves diagnoses and reduces dose variability across hospitals. Continuous monitoring, comparing dose data with established values, ensures compliance and patient safety.

By applying these AI models, the project seeks to provide healthcare professionals with a reliable tool for estimating radiation doses in CT scans. This enables them to optimize imaging protocols, personalize patient care, and minimize the potential health risks associated with ionizing radiation exposure.

3 Dosimetric Quantities for Radiation Measurement and Dose Assessment

Dosimetric physical quantities are essential for establishing principles and systems of radiological protection in both external and internal exposure of biological tissues to radiation [9]. External radiation fields can be described solely by physical quantities, while internal fields depend on biokinetic, anatomical, and physiological parameters, making them extremely difficult to estimate.

Kinetic Energy Released per unit Mass (*Kerma*) is defined as the average energy transferred ($E_{T,R}$) by neutral particles (photons and neutrons) to charged particles (electrons and protons) in the medium, per unit mass (m). The SI unit of *Kerma* is Gray (Gy) and is defined by equation (1).

$$K = \frac{dE_{T,R}}{dm} \quad (1)$$

Absorbed dose (*D*) is the fundamental quantity in radiological protection and is defined as the energy deposited by ionizing radiation (E) per unit mass of tissue (m). The SI unit of absorbed dose is Gray (Gy) and is defined by equation (2).

$$D = \frac{dE}{dm} \quad (2)$$

Dose equivalent (H_T) in an organ or tissue T corresponds to the average absorbed dose in that tissue or organ ($D_{T,R}$) deposited by radiation R , which is defined by the type and energy of the interacting radiation and W_R corresponds to the radiation weighting factors that have been stipulated based on the relative biological effectiveness (RBE) of the

various radiations. It is calculated using equation (3), and its SI unit is Sievert (Sv).

$$H_T = \sum W_R D_{T,R} \quad (3)$$

Effective dose (E) is defined as the weighted sum of equivalent doses in different tissues (equation (4)), where W_T is the tissue weighting factor, and $\sum W_T = 1$. This parameter considers the different radiosensitivity of each exposed tissue or organ, with tissues having higher radiosensitivity associated with higher values of W_T . Its SI unit is Sievert (Sv).

$$E = \sum_T W_T H_T = \sum_T W_T \sum_R W_R D_{T,R} \quad (4)$$

The values of W_T were initially established in 1977 by the ICRP and revised in 1990 to consider the fatality of different neoplasms and hereditary effects. Based on various epidemiological studies, they were revised again, resulting in changes that apply to all age groups and both genders. It is worth noting that over the years, there has been a decrease in W_T values, indicating a decrease in the radiosensitivity associated with different tissues. Consequently, the radiation hazard to biological tissues has been reduced.

4 Conclusions

This study intends to achieve several objectives related to importing, processing and analyzing data from CT images. By developing models for predicting the appropriate radiation dose, identifying potential adverse effects of radiation exposure, analyzing the factors that influence the dose received by patients and validating diagnostic reference levels, the aim is to ensure adherence to the principles of justification, optimization and limitation of radiation dose.

By achieving these results, the project will contribute to the radiological protection of patients, improving the quality of imaging exams and maximizing clinical benefits, ensuring that each patient receives the appropriate radiation dose and minimizing the associated risks.

References

- [1] Brenner, D. J., & Hall, E. J. (2007). Computed Tomography — An Increasing Source of Radiation Exposure. *New England Journal of Medicine*, 357(22), 2277–2284.
- [2] Choplin, R. H., Boehme, J., & Maynard, C. (1992). Picture archiving and communication systems: an overview. *Radiographics*, 12(1), 127–129.
- [3] Council Directive 2013/59/Euratom L 13/1 (12/05/2013). <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32013L0059>
- [4] Council Directive 97/43/Euratom L 180/22 (06/30/1997). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31997L0043&qid=1693062785587>
- [5] Crouch, G., & Crouch, J. (2012). Generating an estimate of patient radiation dose from medical imaging scans. Patent No. US 2012/0148132 A1.
- [6] Decreto-Lei n.º 81/2022. Diário da República n.º 234/2022, Série I (12/06/2022). <https://diariodarepublica.pt/dr/detalhe/decreto-lei/81-2022-204379871>
- [7] Decreto-Lei n.º 108/2018. Diário da República n.º 232/2018, Série I (03/12/2018). <https://diariodarepublica.pt/dr/legislacao-consolidada/decreto-lei/2018-117202808>
- [8] DICOM. (2014). DICOM Strategic Document. Arlington: DICOM.
- [9] Greening, J. (1981). *Fundamentals of Radiation Dosimetry*, Adam Hilger, Bristol.
- [10] Hendee, W., & Edwards, F. (1986). *ALARA and an integrated approach to radiation protection*. *Semin. Nucl. Med.* 16(2): 142–150.
- [11] IAEA. (2014). *Dosimetry in Diagnostic Radiology: an International Code of Practice*. Igarss 2014, New York.
- [12] ICRP. (2007). *The 2007 Recommendations of the International Commission on Radiological Protection*. ICRP Publication 103.
- [13] Murphy, A. (2020). *CT chest non-contrast (protocol)*. <https://radiopaedia.org/articles/ct-chest-non-contrast-protocol-1?lang=gb> (accessed Aug. 12, 2023).

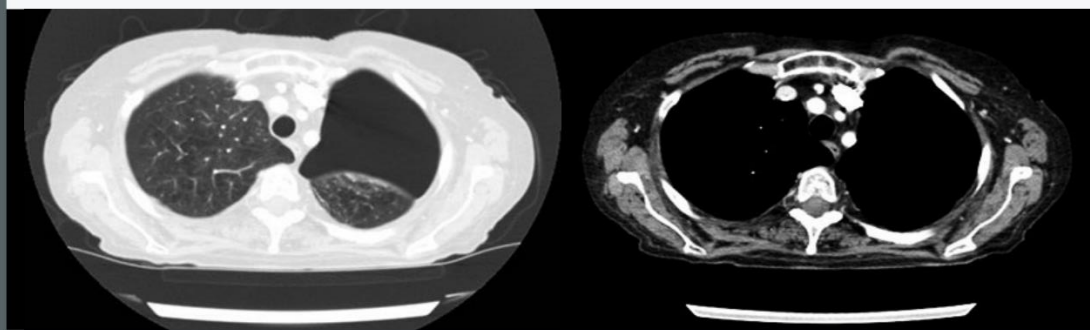
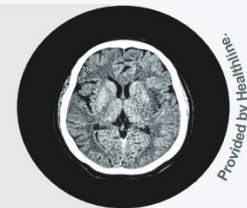
AI-Driven Approaches for Radiation Dose Prediction in Computed Tomography Scans

Stephanie Batista, Inês Domingues, Miguel Couceiro, Ricardo Filipe

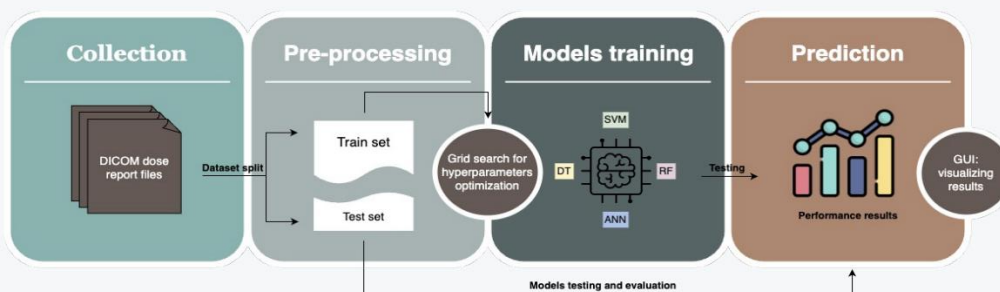
 *protecting lives, one scan at a time* 

Description

This project aims to develop an AI-based strategy for predicting ionizing radiation doses in Computed Tomography (CT) scans. Using a large dataset from CT scans stored in a Portuguese hospital's Picture Archiving and Communication System (PACS), this study will apply an anonymization procedure to protect patient privacy in compliance with GDPR. Various AI models will be explored. The selection of models will be based on their ability to handle complex medical data and accurately predict radiation doses.



System approach overview



Main characteristics

- Enables healthcare professionals to analyze data efficiently for **accurate diagnoses and treatments**.
- Align doses with **Diagnostic Reference Levels (DRLs)** in Portugal.
- Approach based on individual attributes following the **As Low as Reasonably Achievable** principle.

Appendix E – MDPI Sensors Journal Publication

Publication of a paper (**Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective**) in *MDPI Sensors* journal, more specifically in the *Biomedical Sensing and Bioinformatics Processing Special Issue*:



sensors



Review

Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective

Stephanie Batista Niño ¹ , Jorge Bernardino ^{1,2} and Inês Domingues ^{1,3,*}

¹ Polytechnic Institute of Coimbra, Coimbra Institute of Engineering, Rua Pedro Nunes-Quinta da Nora, 3030-199 Coimbra, Portugal; a2019114900@isec.pt (S.B.N.); jorge@isec.pt (J.B.)

² Centre for Informatics and Systems, University of Coimbra (CISUC), Pólo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal

³ Medical Physics, Radiobiology and Radiological Protection Group, Research Centre of the Portuguese Institute of Oncology of Porto (CI-IPOP), 4200-072 Porto, Portugal

* Correspondence: inesdomingues@gmail.com

Abstract: Oncology has emerged as a crucial field of study in the domain of medicine. Computed tomography has gained widespread adoption as a radiological modality for the identification and characterisation of pathologies, particularly in oncology, enabling precise identification of affected organs and tissues. However, achieving accurate liver segmentation in computed tomography scans remains a challenge due to the presence of artefacts and the varying densities of soft tissues and adjacent organs. This paper compares artificial intelligence algorithms and traditional medical image processing techniques to assist radiologists in liver segmentation in computed tomography scans and evaluates their accuracy and efficiency. Despite notable progress in the field, the limited availability of public datasets remains a significant barrier to broad participation in research studies and replication of methodologies. Future directions should focus on increasing the accessibility of public datasets, establishing standardised evaluation metrics, and advancing the development of three-dimensional segmentation techniques. In addition, maintaining a collaborative relationship between technological advances and medical expertise is essential to ensure that these innovations not only achieve technical accuracy, but also remain aligned with clinical needs and realities. This synergy ensures their applicability and effectiveness in real-world healthcare environments.

Keywords: artificial intelligence; computed tomography; hepatic pathologies; liver segmentation



Citation: Niño, S.B.; Bernardino, J.; Domingues, I. Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective. *Sensors* **2024**, *24*, 1752. <https://doi.org/10.3390/s24061752>

Academic Editors: Wenfeng Zhao, Min Shi, Yan Luo and Yu Tian

Received: 5 February 2024

Revised: 2 March 2024

Accepted: 5 March 2024

Published: 8 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As one of the most important organs in the digestive system, the liver performs critical functions such as breaking down nutrients, producing bile and eliminating toxic substances. However, liver-related diseases, particularly oncological diseases, pose health risks, and liver cancer is a leading cause of cancer-related mortality worldwide [1,2]. Computed tomography (CT) has become an integral part of diagnosis, treatment planning and monitoring the progress of oncological diseases [3,4], providing detailed cross-sectional images for accurate visualisation of internal structures, including liver tumours [5,6].

With the advancement of technology and artificial intelligence (AI) in medicine, there is a growing need to optimise the identification of oncological diseases [7]. Medical image segmentation is emerging as a fundamental step in the pipeline [8–11]. Liver segmentation in CT scans has emerged as a critical area, requiring accurate identification and delimitation of the liver region for treatment planning and progress monitoring, as well as for early detection of liver lesions and metastases to other organs [12–14]. However, accurate liver segmentation on CT scans is challenging due to factors such as artefacts, varying soft tissue densities and the complexity caused by adjacent organ proximity [15,16].

The aim of this paper is to provide an overview of the state of the art in the application of AI, as well as traditional methods and techniques, for liver segmentation in CT scans to enable an understanding of which factors most influence the performance of the models and methods used by the selected studies and lead them to perform differently for the same objective.

Some of the aspects that will be compared include datasets considerations, algorithms used, robustness, performance and evaluation. Specifically, we aim to answer the following research questions (RQs):

- RQ1—What are the challenges and limitations associated with accurate liver segmentation in CT scans?
- RQ2—How does the choice of the method impact the accuracy and efficiency of liver segmentation in CT scans?
- RQ3—What are the evaluation metrics commonly used to assess the performance of AI models and traditional methods for liver segmentation in CT scans?

A systematic review of the literature was conducted to address these research questions. Google Scholar was used for document retrieval, and the most representative documents from each year were selected. It is worth noting that only a few studies were found that provided an intuitive and visual approach for healthcare professionals to manipulate and interpret the segmentation results.

This paper reviews studies on liver segmentation from CT scans, organised according to the methods used. The paper's selection methodology is detailed in Section 2, followed by a categorisation of AI models and medical image processing methods in Section 3. Section 4 summarises the main findings, while Section 5 provides a discussion of the results of the methods used. The paper concludes in Section 6 with the conclusions and directions for future work.

2. Methodology

This section outlines the approach taken in conducting the literature review, which involved synthesising existing knowledge, critically assessing methodologies and analysing the results to compare the performance of each AI model and traditional methods for liver segmentation in CT scans.

2.1. Data Sources

Google Scholar (<https://scholar.google.com/>, accessed on 20 January 2024) has developed over the years and has become a robust database for the scientific literature [17]. It was therefore chosen as the research tool for the present study.

2.2. Search Queries

A search performed on 30 January 2024, with the query “intitle:Liver + intitle:segmentation + (intitle: CT OR + intitle:tomography)” returned approximately 980 results in just 0.04 s.

2.3. Inclusion Criteria

The most relevant papers from each year have been included in this historical overview. In the context of Google Scholar, relevance refers to the degree to which the search results match the criteria or the context of the query. The sorting algorithm takes into account several factors to determine the order of the results, including the presence of search terms and citation counts. In addition, the review papers were all included in the current state of the art, which amounted to a further five documents.

2.4. Exclusion Criteria

Papers that did not meet the criteria defined in Section 2.3 were not included. Papers written in languages other than English, Portuguese and Spanish were also excluded. As a final exclusion criterion, papers were excluded if the full document was not publicly available.

2.5. Characterisation of Selected Papers

The earliest paper dates to 1990, and thus, a potential total of 35 papers could be selected (the most relevant per year between the years of 1990 and 2024). The distribution of the total number of papers retrieved per year is given in Figure 1. As can be seen, some years do not have any papers, and thus a total of 30 documents were finally included in the historical review. It is also clear that interest in the topic has increased over the years.

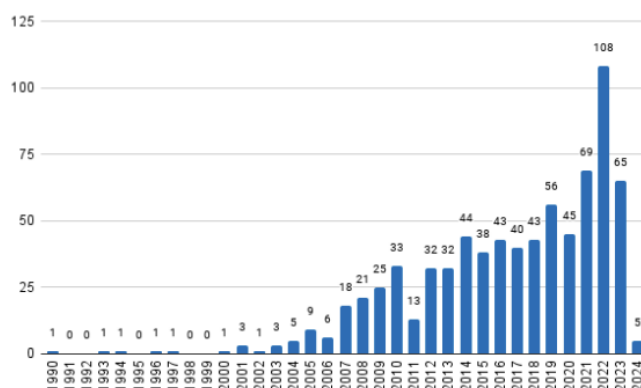


Figure 1. Distribution of the retrieved papers over the years.

A word cloud constructed from the titles of the selected papers is shown on the left side of Figure 2. On the right side is a pie chart showing the number of journals, conference proceedings and reports.

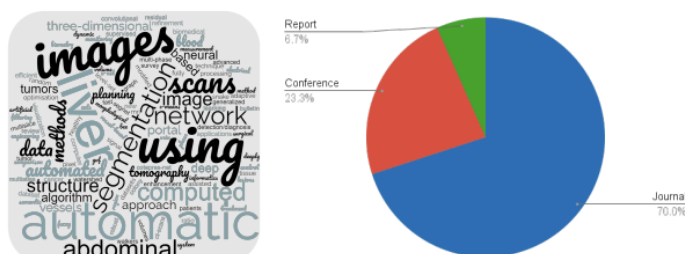


Figure 2. Characterisation of the selected papers.

Several taxonomies can be used to characterise the reviewed papers. We decided to adopt the taxonomy introduced in a recent paper by Sakshi and Kukreja (2023) [18], published in the reputable journal Archives of Computational Methods in Engineering (Impact Factor of 9.7). Figure 3 shows the selected papers categorised by image segmentation technique according to this taxonomy.

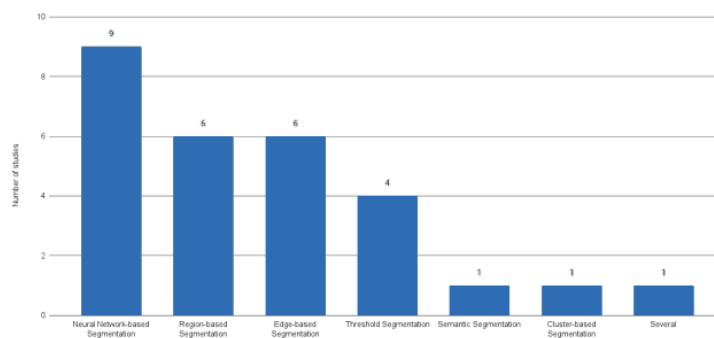


Figure 3. Selected paper distribution according to the categories of image segmentation techniques.

3. Literature Review

Based on the selected papers, this literature review provides a historical overview of different methods and approaches for liver segmentation in medical imaging (Section 3.1). It covers neuronal-network-based segmentation, region-based segmentation, edge-based segmentation, threshold segmentation, semantic segmentation and cluster-based segmentation techniques. This review includes a comparison of the results and discusses the potential applications and strengths of each method. Section 3.2 compares this review with other existing summaries of the state of the art.

3.1. Historical Overview

As mentioned in Section 2, one paper per year between the years of 1990 and 2024 was selected, summing to a total of 35 documents. For five years, the search retrieved no results (see Figure 1), resulting in 30 documents. Unfortunately, the full texts of two works were not available. The remaining 28 documents are briefly described below.

The oldest work found to tackle liver segmentation is the one by Bae et al. (1993) [19], presenting a similar sequential image-by-image segmentation technique using a reference image, where the liver occupies a significant portion of the abdomen cross-section. Image processing techniques, including grey-level thresholding, Gaussian smoothing and connectivity tracking, are employed to extract the liver boundaries. The resulting boundaries are then smoothed using mathematical morphology techniques and B-splines. This study focuses on a living-donor liver transplant program, and the computer-determined boundaries are compared with those drawn by a radiologist, showing agreement within 10% of the calculated areas.

Gao et al. (1996) [20] focus on facilitating 3D visualisations for surgical planning. Their method employs a global histogram analysis, morphologic operations and a parametrically deformable contour model to delineate the liver boundary. Ten cases were used to validate the approach and promising results were found with minimal operator intervention required.

Soler et al. (1997) [21] propose an automatic method for segmenting the portal vein, with the primary objective of achieving accurate segmentation with detailed branching and topological information, facilitating the localisation of liver tumours concerning Couinaud's anatomical segmentation. This approach involves the initial detection of liver contours using 3D deformable models, followed by limiting the CT images to a liver mask containing hepatic tissue, vascular trees and potential tumours. Classification of anatomical structures is performed using Gaussian curves fitted to an intensity histogram. The vascular trees and tumours are segmented through a hysteresis thresholding technique based on a distance map, considering the Gaussian parameters. An isotropic image is obtained through shape-based

interpolation and the portal vein is reconstructed using skeletonization, eliminating short branches and correcting errors. Results demonstrate that the algorithm automatically extracts the first three main bifurcations of the portal vein, comparable to manual segmentation.

Yoo et al. (2000) [22] focus on the use of pixel ratios. By analysing the grey value range of a normal liver in CT images, a binary image is generated and then processed into four mesh images based on hole ratios to eliminate noise. A template representing the general outline of the liver is generated from the union image of these mesh images and subtracted from the binary image to accurately represent the organ boundary. The pixel ratio, which takes into account the distribution of organ pixels, was used to discriminate between the organ and noise, especially in cases where organs have similar grey value ranges. The proposed method reduced the processing time compared to existing methods and was validated against manual segmentation by medical experts.

Pan and Dawant (2001) [23] introduce a level-set approach, which addresses the challenge of defining appropriate speed functions for contour propagation. A speed function is proposed to stop the propagation of the contour at organ boundaries with weak edges by incorporating the accumulative speed based on the path of the contour, enhancing the robustness of segmentation in noisy images. The method also leverages a priori anatomical information to improve the accuracy. Tested on five CT datasets, including cases with abnormal livers, this method demonstrates good agreement with manual delineations.

Saitoh et al. (2002) [24] present an automated method for segmenting the liver region from the third phase of abdominal CT scans. Their approach involves the extraction of blood vessels using a threshold, followed by morphological dilation to define an approximate liver region useful for the removal of adjacent organs. The final liver region is then extracted using a threshold. The method is thus based on mathematical morphology and thresholding techniques, using the unique characteristics of blood vessels to functionally identify the liver region. The experiments performed on eight CT datasets show a good agreement between the automatically and manually detected liver regions.

Masumoto et al. (2003) [25] use multislice CT images. Their method uses two time-varying images acquired during the contrast medium circulation phase, highlighting the liver region through CT value changes. The proposed scheme involves generating a liver likelihood image by analysing CT value changes and subsequently extracting the liver region while considering the geometric characteristics of blood vessels and tumours. The evaluation, based on receiver operating characteristic (ROC) analyses, demonstrates the superiority of the proposed method over other approaches, especially when using information from both phases.

The scheme proposed by Lim et al. (2004) [26] uses an ROI approach to optimise computational efficiency. Morphological filters, incorporating a priori knowledge of liver location and intensity, detect the initial boundary. The algorithm then generates a gradient image using the weighted initial boundary and employs an immersion-based watershed algorithm for segmentation. Post-processing includes region merging based on statistical information to refine the segmentation.

Liu et al. (2005) [27] present a gradient vector flow (GVF) snake-based method for the semi-automatic segmentation of liver volumes in contrast-enhanced CT images. The algorithm follows a stepwise approach, starting with the computation of an initial edge map using the Canny edge detector and the estimation of a liver template. The edge map is then modified to suppress edges within the liver using the liver template, and a concavity removal algorithm is applied to refine the liver boundary. The GVF field is computed based on the modified edge map, and the initial liver contour is determined by considering the candidate initial contour and the computed GVF field. The final liver contour is obtained by deforming the initial contour using the snake. The method was evaluated on 20 contrast-enhanced volumetric liver images, and the results were compared with a radiologist's manual delineation. The median difference ratio between the computer-generated results and manual results is 5.3%, with a range of 2.9% to 7.6%.

A three-stage approach is used by Lim et al. (2006) [28]. The first stage involves image simplification as preprocessing, where an ROI is identified and thresholds are determined using multilevel thresholding. The second stage detects a search range using multiscale morphological filtering, region labelling, and partition clustering. The third stage uses a contour-based segmentation approach with a labelling-based search algorithm to refine the initial liver boundary. The effectiveness of the algorithm is demonstrated through experimental results on contrast-enhanced abdominal CT images, with an average segmentation accuracy of 96%. Volume measurement is performed based on the segmented liver regions, with an average error rate of 3%.

Beichel et al. (2007) [29] introduce a two-step process. First, initial segmentation is performed using graph cuts, overcoming challenges such as the high variability in liver shape and grey-value appearance. Second, an interactive refinement step is introduced, allowing users to correct segmentation errors in a 3D environment. The refinement is facilitated by a hybrid desktop/virtual reality (VR) user interface. This approach is demonstrated on ten contrast-enhanced liver CT scans, demonstrating robustness to variations in patient data. The results also indicate an improved segmentation quality with low interaction times.

Massotier and Casciaro (2008) [30] present a fully automated method that uses a statistical-model-based approach to distinguish liver tissue from other abdominal organs. An active contour technique using gradient vector flow is used for smoother segmentation of the liver surface. Automatic classification is performed to isolate hepatic lesions from liver parenchyma. The method was evaluated on 21 datasets and demonstrated robust and efficient liver and lesion segmentations close to the ground truth, with an average processing time of 11.4 s per 512×512 pixel slice. The volume overlap for liver surface segmentation is 94.2%, and the accuracy is 3.7 mm. Tumour detection achieved a sensitivity and specificity of 82.6% and 87.5%, respectively.

Heimann et al. (2009) [31] focus on the comparison and evaluation of different methods. The image data, acquired from different CT scanners, consisted of contrast-dye-enhanced scans showing pathological conditions like tumours and cysts. Radiology experts manually delineated the liver contours in transversal slices to create reference segmentations. A total of 40 images were divided into training and test sets for algorithm evaluation. Evaluation measures included volumetric overlap, relative volume difference, and surface distances. Fully automated and interactive segmentation methods were employed, with the former showing discernible performance differences. The best-performing automated approaches used statistical shape models. Interactive methods achieved higher scores with more user interaction. A combined approach using majority voting from the best-performing methods outperformed individual automated and interactive results.

A three-step procedure is outlined by Akram et al. (2010) [32]. Firstly, a pre-processing step involves converting the image to greyscale and applying a 3×3 median filter to reduce noise. The second step focuses on liver segmentation, with a global threshold and morphological operations to obtain the final segmented liver region. Finally, post-processing steps include adaptive histogram equalisation, Gaussian smoothing, and grey-level transformations to enhance the segmented liver region. Experimental tests on 100 CT images demonstrate the accuracy of the proposed method by comparing automated segmentation results with images manually segmented by hepatologists and oncologists.

The approach of Oliveira et al. (2011) [33] involves a sequence of four steps. First, the liver is segmented using level sets with parameters optimised by a genetic algorithm (GA). A Gaussian fit is employed to define the speed image for level set propagation. Secondly, vessels and nodules are segmented using a Gaussian mixture model, focusing on adipose nodules. A region-growing method with information from the Gaussian model is applied. Thirdly, vessels are classified into portal veins or hepatic veins using a vein tracking method. Finally, a geometric approach based on the identified veins is used to segment the liver into different Couinaud regions. Liver segmentation is based on the assumption that the liver parenchyma homogeneity and veins are mainly inside the liver. The parameters are estimated using a GA, and the fitness evaluation involves comparing the segmentation with

a reference using five disparity metrics. The proposed method shows good performance, ranking among the top methods in the MICCAI-SLiver07 conference evaluation.

The method developed by Linguraru et al. (2012) [34] uses a robust parametrisation of 3D surfaces for point-to-point correspondence, overcoming challenges such as inconsistent contrast enhancements and imaging artefacts. A shape descriptor that is invariant under rotation and scale is used to compare the local shape features of organs. Initial liver segmentation is refined using a shape-driven geodesic active contour, and hepatic tumours are detected and segmented using graph cuts and support vector machines (SVMs). This technique is evaluated on a dataset of 101 CT scans and shows improvements in the liver segmentation accuracy, particularly in cases with large tumours and segmentation errors. Furthermore, the method identifies liver tumours with a low rate of false positives.

Li et al. (2013) [35] discuss a method that makes use of fuzzy clustering and level set techniques. The Fuzzy C-Means (FCM) clustering algorithm is employed, which assigns pixels to different categories based on fuzzy memberships, considering both the grey level intensity and spatial information. The FCM algorithm is iteratively optimised by minimising a cost function, allowing the fuzziness of the resulting partition. To overcome the limitations of standard FCM, a spatial FCM algorithm is introduced that incorporates spatial information into fuzzy membership functions. This paper also introduces the level set method, a continuous deformable model for segmentation. Distance-Regularized Level Set Evolution (DRLSE) is proposed to address reinitialisation issues and improve the efficiency. The proposed method is evaluated using accuracy, sensitivity, and specificity metrics and demonstrates a high performance in liver segmentation, especially in cases with unclear boundaries. A comprehensive review of abdominal image segmentation using soft and hard computing approaches is provided in [36].

Platero et al. (2014) [37] integrate a multi-atlas segmentation approach with graph cuts. Their method includes several steps: (1) obtaining an initial solution using low-level operations to define the ROI around the liver; (2) constructing a fast probabilistic atlas for the ROI and computing a coarse binary segmentation using segmentation-affine registration; (3) ranking the atlases based on segmentation similarity and propagating selected atlases to the target image; (4) improving the segmentation accuracy through label fusion, minimising the discrete energy function; and (5) evaluating the approach using a public liver segmentation database. The experimental results show a high accuracy, competitive with human expert segmentation.

Artificial Bee Colony (ABC) optimisation is used by Mostafa et al. (2015) [38]. Their algorithm use ABC optimisation to cluster different intensity values in abdominal CT images, followed by mathematical morphological operations to manipulate and separate the clusters. This process eliminates small and thin regions, such as flesh regions or organ edges. The extracted regions form an initial estimate of the liver area, which is further enhanced using a region-growing technique. The proposed approach demonstrates a segmentation accuracy of 93.73% on a test dataset of 38 CT images, taken in the pre-contrast phase.

A 3D deeply supervised network (DSN) is introduced by Dou et al. (2016) [39]. The proposed architecture consists of 11 layers, including 6 convolutional layers, 2 max-pooling layers, 2 deconvolution layers, and 1 softmax layer. The network is designed in a 3D format to effectively capture spatial information. The 3D DSN employs deep supervision via additional deconvolutional layers to counteract vanishing gradients, thus improving the training process. The learning objective is to minimise per-voxel-wise binary classification errors, with deep supervision injected at specific layers. The MICCAI-SLiver07 dataset is used for evaluation, demonstrating that the 3D DSN has a faster convergence and lower errors when compared to traditional 3D convolutional neuronal networks (CNNs).

Christ et al. (2017) [14] propose a cascaded fully CNN (CFCN) on CT slices that sequentially segments the liver and lesions. First, various preprocessing steps, including Hounsfield unit windowing and contrast enhancement, are applied. Then, the cascaded approach involving two U-Net architectures is used for liver and lesion segmentation. Finally, 3D conditional random fields (CRFs) are used to refine the segmentation results.

Generalisation and scalability to different modalities and real-life datasets, including a diffusion-weighted magnetic resonance imaging (MRI) dataset and a large multi-centre CT dataset, are shown.

Hiraman (2018) [40] presents a slice alignment method that addresses the challenges through optimal threshold selection, skeletonization, and enhanced correlation coefficient (ECC) alignment. Next, a CNN-based liver region of interest detection method is proposed to classify 2D slices for focused processing.

The study presented by Wang et al. (2019) [41] investigates the application of a generalised CNN for automated liver segmentation and biometry using cross-sectional data from abdominal CT and MRI scans. Their retrospective study included a sample of 563 abdominal scans from 530 adults, covering different imaging modalities. The CNN was initially trained on 300 unenhanced multiecho 2D SPGR MRI sets and then subjected to transfer learning for generalisation across different imaging methods. The accuracy of the CNN was evaluated using internal and external validation datasets. This study also investigates the impact of training data size on the segmentation accuracy and explores the feasibility of using automated liver segmentation for volumetry and hepatic PDF quantification.

Almotairi et al. (2020) [42] explore the application of the SegNet architecture. The proposed modified SegNet model uses the VGG-16 network as an encoder. Tests were performed on a standard dataset for liver CT scans (3D-IRCADb01 [43]), achieving a tumour accuracy of up to 99.9% in the training phase and 86% for tumour identification.

Ayalew et al. (2021) [44] present a modified U-Net architecture and introduce a new class balancing method. To address the class imbalance between the liver and tumours, a weighting factor is applied and slices without a tumour are removed during data preparation. The U-Net-based network architecture includes batch normalisation, dropout layers, and filter size reduction. Training involves tuning hyperparameters, such as the learning rate and batch size. The datasets used are derived from the 3D-IRCADb01 [43] and LiTS [45] databases and the results achieve a Dice Similarity Coefficient (DSC) of 0.96 and 0.74, respectively. The algorithm also introduces a novel approach for direct tumour segmentation from abdominal CT scan images, with a comparable performance to existing two-step methods.

The study of Scicluna (2022) [46] is motivated by challenges such as the Combined Healthy Abdominal Organ Segmentation (CHAOS) Challenge [47], which focuses on healthy abdominal organs. The study focuses on replicating the v16pUNet1.1C model, which demonstrated a superior performance in Task 2 of the CHAOS Challenge. Results from the v16pUNet1.1C model are presented and compared with variations in the loss function and scaling transformation. The application of a 3D largest-connected-component filter is discussed, showing improvements in mean scores.

A deep semantic segmentation CNN is used by Ezzat et al. (2023) [48]. A three-stage architecture is proposed, including pre-processing with data augmentation, deep CNN training, and testing. The CNN-based semantic segmentation model is shown to be robust, achieving a test accuracy of 98.8%. The approach does not require user input, making it accessible to non-experts.

Shao et al. (2024) [49] present the Attention Connect Network (AC-Net) for liver tumour segmentation in CT and MRI images. The AC-Net consists of two main modules: an axial attention module (AAM) and a Vision Transformer module (VTM). The AAM uses an axial attention mechanism to merge features of matching dimensions, maximising the use of spatial features extracted by a CNN. The VTM processes high-level semantic features extracted by the CNN using a methodology similar to Vision Transformers (ViTs) [50]. The network achieves a DSC of 0.90, a Jaccard coefficient (JC) of 0.82, a recall of 0.92, a precision of 0.89, a Hausdorff distance (HD) of 11.96, and an average symmetric surface distance (ASSD) of 4.59.

3.2. Other Review Papers

The search described in Section 2 retrieved six literature review documents. For one of the works, however, the full document was not available. The remaining five are briefly presented in the following.

A comparative analysis of various available techniques, focusing on their advantages and disadvantages, is given in [51]. Recognising the challenges posed by the variable shape of the liver and the weak edges in adjacent organ regions, this survey covers approaches such as threshold, model, level set, region, active contour, and clustering. This paper also divides its investigation into sections, covering both image pre-processing and segmentation techniques, providing an overview of the current landscape in liver segmentation from CT images.

Study [52] provides a survey of 3D image segmentation methods, focusing on selected binarization and segmentation techniques suitable for processing volume images. For thresholding methods, both global and local techniques are considered, and challenges such as hysteresis in dealing with voxel value distributions are addressed. The region growing section explores voxel-based procedures, including growing by grey value and adaptive region growing. In addition, deformable surfaces and level set methods are discussed, before other segmentation concepts such as fuzzy connectedness and watershed algorithms are introduced. The concluding remarks underline the complexity of image segmentation, emphasising the absence of a universal solution and the need to carefully evaluate and select methods based on specific tasks and dataset characteristics. The challenges posed by 3D data, including the data volume and issues of interactivity and visualisation, are also acknowledged.

Study [53] reviews the literature on methods for segmenting liver images, distinguishing between semi-automatic and fully automated techniques. The challenges of liver image segmentation, such as low contrast, blurred edges, and the complexity of the liver morphology, are discussed. Different approaches are reviewed, including neuronal-network-based methods, support-vector-machine-based methods, clustering-based methods, and hybrid methods. It is concluded that, despite progress, liver image segmentation remains a challenging task, and the authors encourage further development of hybrid approaches for more accurate segmentation.

Various segmentation methods, including statistical shape models, probabilistic atlas-based approaches, geometric deformable models, and machine-learning-based methods, are reviewed in [54]. This review includes information on available databases and challenges in liver tumour segmentation, highlighting the scarcity of public datasets and the need for improved segmentation methods. Liver blood vessel segmentation and computer-assisted diagnosis (CAD) systems are also reviewed. The conclusion highlights the importance of segmentation, particularly in pathological cases, and the need for improved CAD systems with accurate segmentation for comprehensive analysis of liver treatment.

The survey paper [55] provides a comparative analysis of various available techniques, focusing on their advantages and disadvantages. Grey-level-based techniques, such as region growing and active contour methods, are highlighted as effective for liver segmentation. This survey acknowledges the challenges of detecting early-phase liver lesions and emphasises the need for a combination of methods to achieve seamless segmentation, with region growing and active contour methods considered more efficient than other segmentation techniques.

This survey differs from the other documents in this section in a number of ways. Firstly, the most recent of the review papers found dates to 2022. One of the contributions of this work is to present a more up-to-date view of the works published since then. In addition, none of the other works present a historical perspective on the subject, starting from 1990, as is the case in the present review.

4. Findings

A summary of the papers reviewed in Section 3.1 is given in Table 1. The columns of the table contain the following information: identification of each study (column Authors); publication year (column Year); general category or approach used in the segmentation method as defined in [18] (column Segmentation Category); the specific segmentation technique or algorithm (column Method); whether the method is fully-automatic or semi-automatic (column Autom. Level for Automation Level); whether the segmentation is performed in 2D or 3D (column Dim. for Dimensionality); the dataset or database used for evaluation (column Database); and the key results of the best segmentation method in each paper (column Results). Graphical depictions of the metrics used in one or more studies are shown in Figure 4.

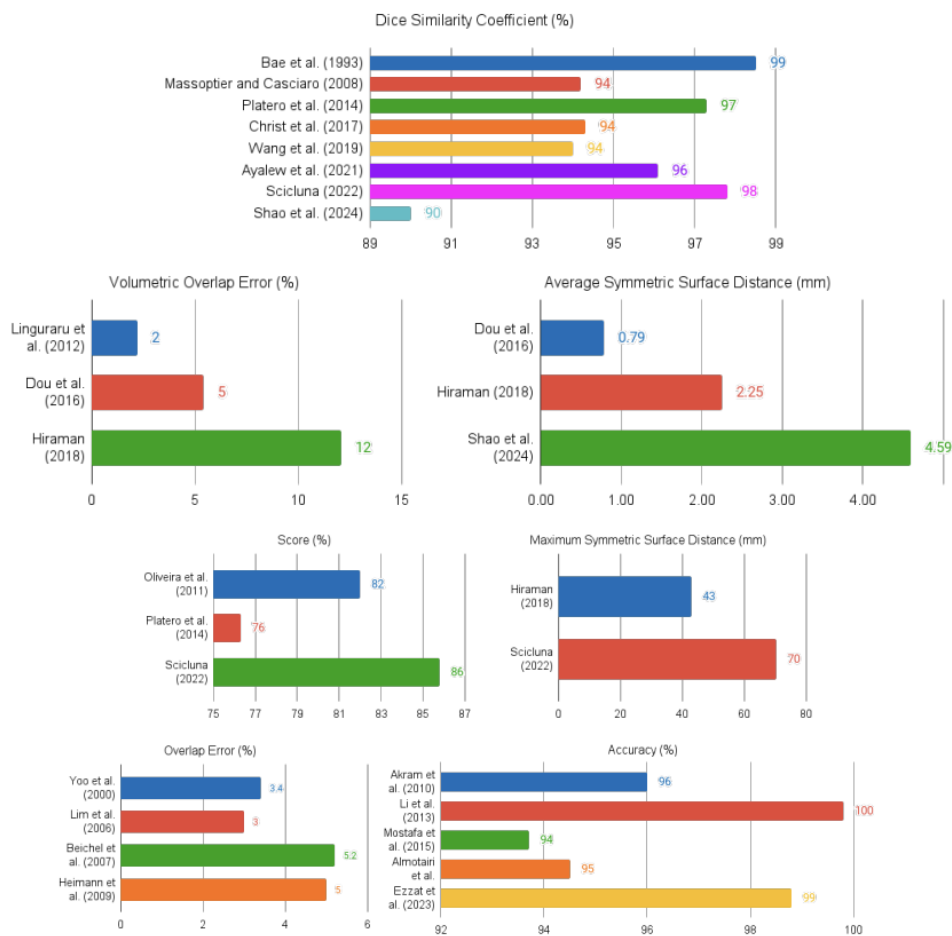


Figure 4. Summary of the quantitative results presented in the reviewed documents.

Table 1. Summary of reviewed documents.

Authors	Year	Segmentation Category	Method	Autom. Level	Dim.	Database	Results
Bae et al. [19]	1993	Threshold	Grey-level thresholding	Semi	2D	Private	0.985 DSC with mean percent error within 10%.
Gao et al. [20]	1996	Edge	Parametrically deformable contour model	Fully	3D	Private	13.2% of the results required user modifications.
Soler et al. [21]	1997	Edge	Deformable models	Fully	3D	Private	Claimed to be comparable to manual segmentations.
Yoo et al. [22]	2000	Threshold	Threshold	Fully	2D	Private	3.41% error.
Pan and Dawant [23]	2001	Edge	Level sets	Fully	Both	Private	[0.874, 0.963] average similarities.
Saitoh et al. [24]	2002	Threshold	Threshold	Fully	3D	Private	~20 min computation time.
Masumoto et al. [25]	2003	Region	Differences between time-phase images	Fully	3D	Private	67% volume ratio average; 32% in the worst cases.
Lim et al. [26]	2004	Region	Watershed	Fully	2D	Private	Only qualitative.
Liu et al. [27]	2005	Edge	GVF snake	Semi	2D	Private	5.3% median value of the difference ratios.
Lim et al. [28]	2006	Semantic	Labeling-based search	Fully	2D	Private	96% average correctness; 3% average error rate.
Beichel et al. [29]	2007	Region	Graph cuts	Semi	3D	Private	5.2% average overlap error.
Massopier and Casciaro [30]	2008	Edge	Active contour	Fully	3D	Private	94.2% mean DSC.
Heimann et al. [31]	2009	Several	Majority Voting	Both	Both	Private	5% overlap error; −0.7 volume difference; 0.8 average distance; 1.7 RMS distance; 19.1 max distance.
Akram et al. [32]	2010	Threshold	Global Threshold	Fully	3D	Private	0.9% average accuracy; 0.0017 std; 96% accurately segmented; 4% poorly segmented.
Oliveira et al. [33]	2011	Edge	Level sets	Semi	2D	SLiver07	82.05 overall score.
Linguraru et al. [34]	2012	Region	Graph cuts	Fully	3D	Private; SLiver07	2.2 VOE.
Li et al. [35]	2013	Edge	Fuzzy clustering and level set	Fully	2D	Private	0.9986 average accuracy; 0.9989 average specificity.
Platero et al. [37]	2014	Region	Graph cuts	Semi	3D	SLiver07	76.3 maximum score; 0.973 DSC.
Mostafa et al. [38]	2015	Cluster	ABC optimization	Fully	2D	Private	93.73% accuracy; 84.82% average SI.
Dou et al. [39]	2016	NN	3D DSN	Fully	3D	SLiver07	5.42% VOE; 0.79 mm ASSD.
Christ et al. [14]	2017	NN	CFCN	Fully	2D	3D-IRCADb01	94.3% mean DSC.
Hiraman [40]	2018	NN	CNN	Fully	2D	SLiver07	12.07% average VOE; −1.96% RVD; 2.25 mm ASSD; 2.60 mm RMSD; 43.01 mm MSSD.
Wang et al. [41]	2019	NN	CNN	Fully	3D	Private	0.94 ± 0.06 DSC.
Almotairi et al. [42]	2020	NN	SegNet	Fully	3D	3D-IRCADb01	94.57% overall accuracy.
Ayalew et al. [44]	2021	NN	U-Net	Fully	2D	3D-IRCADb01; LITS	0.9612 DSC.
Scicluna [46]	2022	NN	UNet; VGG16UNetC	Fully	2D	CHAOS	85.84 mean score; 97.85 DSC; 80.33 RAVD; 94.80 ASSD score; 70.38 MSSD.
Ezzat et al. [48]	2023	NN	CNN	Fully	2D	Private	96.80% accuracy.
Shao et al. [49]	2024	NN	AC-Net	Fully	3D	Private; LITS	0.90 DSC; 0.82 JC; 0.92 recall; 0.89 precision; 11.96 HD; 4.59 ASSD.

ABC: Artificial Bee Colony; AC-Net: Attention Connect Network; ASSD: average symmetric surface distance; CFCN: cascaded fully CNN; CNN: convolutional neural network; DSC: Dice Similarity Coefficient; DSN: deeply supervised network; GVF: gradient vector flow; HD: Hausdorff distance; JC: Jaccard coefficient; MSSD: maximum symmetric surface distance; NN: neuronal network; RAVD: relative absolute volume difference; RMS: root mean square; RMSD: root mean square Symmetric Surface Distance; RVD: relative volume difference; SI: similarity index; std: standard deviation; VOE: volumetric overlap error.

Table 2 presents a brief description of each segmentation category, highlights its main advantages and disadvantages, and presents some application details. It is clear from Table 1 that prior to 2016 there was no predominant category. Region-based segmentation, edge-based segmentation, threshold segmentation, semantic segmentation, cluster-based segmentation, and even combinations of several methods were tried. However, since 2016, neuronal-network-based techniques have dominated the field. Remembering the huge impact that AlexNet [56] had in winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [57], it is clear that liver segmentation in CT scans took a few years to catch up with the state-of-the-art research. This is probably due to the arrival of U-Net [58], which was proposed in 2015, that is specifically designed for biomedical image segmentation.

Table 2. Comparison of segmentation categories.

Category	Description	Main Advantages	Main Limitations	Applicability
Threshold	Segments based on intensity thresholds	Simple, fast, easy to implement	Sensitivity to threshold selection, suffers from noise and artefacts	Commonly used in cases where clear intensity differences between the ROI and other regions exist
Edge	Segments based on intensity transitions between regions	Accurate delineation of organ boundaries and structures	Sensitive to noise, difficulties with capturing complex structures	Suitable for images with clear organ boundaries and well-defined edges, but may struggle with low-contrast areas
Region	Segments based on homogeneous regions within an image	Simple implementation, intuitive methodology	Sensitive to initialisation	Often used in cases where interpretability is a concern, but may struggle with fine details
Semantic	Segments based on semantic meaning of pixels	Pixel-level segmentation, fine-grained structural detail	Complex to implement, resource-intensive, and computationally expensive	Suitable for segmenting anatomical structures with distinct features
Cluster	Segments based on similar data patterns or clusters	Efficient grouping and identification of similar data patterns	Sensitivity to initialisation and noise, limited to specific data distributions	Useful for identifying patterns and groups within the data, but can struggle with irregular shapes
NN	Learns models to segment images based on learned features	High accuracy, efficient learning from data	Requires large training datasets, computationally intensive	Suitable for different types of data due to its flexibility and adaptability

In terms of automation, both fully automated and semi-automated techniques have been explored (Figure 5). While it is nice to have fully-automatic, accurate and fast techniques, the final decision should always belong to the specialist. Thus, we advocate fully automatic methods for contour initialisation, together with the development of intuitive tools that allow specialists to modify the fully automatically generated contour if they feel the need to do so. This point of view is in line with current clinical practice, where the specialist follows contouring guidelines from respected entities such as the European Society for Radiotherapy and Oncology, the American Society for Radiation Oncology, or the Global Harmonization Group [59], while being allowed to use built-in auto-delineation and interpolation tools [60]. We therefore agree with Sarria et al. [60] in that, while AI can improve the accuracy and consistency of contouring, it cannot replace the knowledge and clinical judgement of radiation oncologists, physicists, or radiation therapists. AI should be used as a tool to support and optimise clinical decision making and not as a substitute for human expertise.

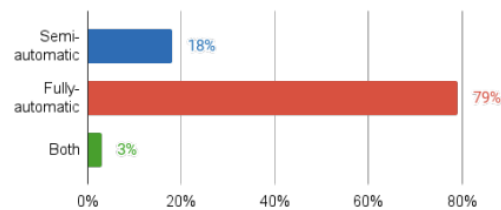


Figure 5. Percentage of studies according to their level of automation (semi-automatic, fully automatic, or both).

Both slice-based (2D) and volume-based (3D) methods have been developed (Figure 6). As an anatomical structure, we believe that liver segmentation methods should be inherently three-dimensional. This would have two main advantages. On the one hand, fully 3D methods would make use of more contextual information and thus potentially provide better segmentations. On the other hand, the development of fully three-dimensional techniques would avoid the need for slice-based methods to aggregate all of the segmentations into a coherent volume. This aggregation could not only lead to errors and anatomically incorrect structures, but would also increase the computational time. Although we do advocate for true three-dimensional methods, we recognise their drawbacks, notably the increased complexity of implementation, the need for more computational resources, the possibly lower efficiency, the increased data requirements, and difficulties in visualisation. The choice must be an informed one, and made on a case-by-case basis.

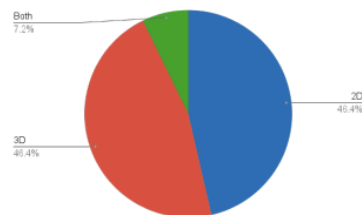


Figure 6. Percentage of studies according to their image dimensionality (2D vs. 3D).

Data availability is a major concern, as most methods use private datasets (Figure 7). This inhibits reproducibility of the results. In addition, researchers who do not have access to hospitals or other facilities with CT scanners cannot develop new techniques for this particular problem. We advocate making the data available, while respecting all the ethical considerations that are important when dealing with medical data and properly anonymising any sensitive information. Some notable exceptions to publicly available datasets are listed in Section 5.1.

5. Discussion

As seen in the previous section, significant advances in liver segmentation techniques have been presented, with particular impact due to the adoption of AI methods, more specifically neuronal network techniques. In this section, we provide some considerations of publicly available databases (Section 5.1), the impact of the widespread adoption of neuronal networks since 2017 (Section 5.2), a comparison between 2D and 3D implementations (Section 5.3), and Section 5.4 present answers to the research questions posed in the introductory section.

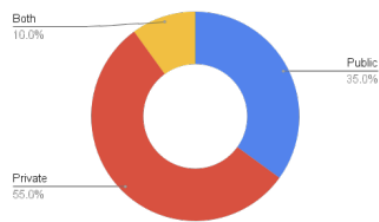


Figure 7. Percentage of database types.

5.1. Public Dataset Analysis

The most common public datasets used in studies of liver segmentation on CT scans include 3D-IRCADb01, LiTS17, and MICCAI-SLiver07 (Figure 8). According to a dataset comparison provided by Al-Saeed et al. [61] (shown in Table 3), it is possible to identify several key differences that may have implications for data processing and analysis, such as different formats and differences in resolution between datasets which may require different approaches to processing and interpretation.

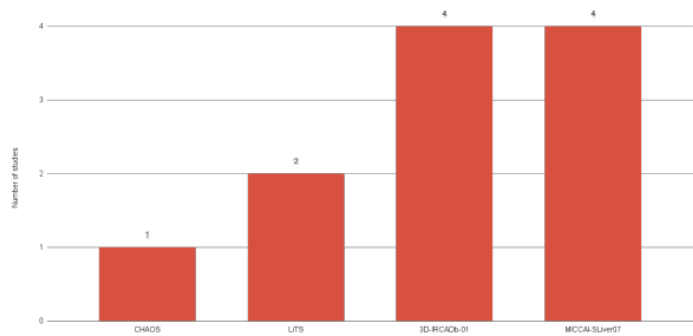


Figure 8. Number of studies using each public dataset.

Table 3. Characteristics of each dataset used by the main analysed studies (based on [61]).

Dataset	Date	Format	Number of Subjects	Slices per Subject	Resolution
MICCAI-SLiver07 [62]	2007	RAW	30	74 to 260	512 × 512
3D-IRCADb01 [43]	2010	DICOM	20	74 to 260	512 × 512
LiTS17 [45]	2017	RAW	200	42 to 1024	Variable

5.2. Impact of the Adoption of Neuronal-Network-Based Methods

The growth of neuronal-network-based approaches has led to remarkable progress in liver segmentation, particularly with respect to CT scans. These models have led to a new era of accuracy and efficiency, significantly outperforming traditional methods [56,58]. This improved accuracy has become critical in the field of medical imaging, where the correct interpretation of CT scans can directly affect the diagnosis and treatment plans of patients. Furthermore, the efficiency of these neuronal network models translates into faster processing times, allowing for more agile decision making in clinical settings.

Another positive aspect of neuronal networks in liver segmentation is their ability to cope with the complexity of the liver anatomy. Neuronal networks, with their systematic

and complex pattern recognition capabilities, are able to navigate these anatomical variations. As a result, they are better able to deal with the variety of appearances that liver tissue can have on CT scans. This ability to handle complex datasets ensures that neuronal networks can provide consistent and accurate segmentation in a wide range of cases.

5.3. Comparison between 2D and 3D Methods for Liver Segmentation

- Importance of Choosing between 2D and 3D Methods
 - In medical imaging, and in particular liver segmentation, the choice between slice-based 2D and volume-based 3D segmentation methods is crucial. This decision is highly dependent on the anatomical structure of the liver. Given the complex, three-dimensional nature of the liver, 3D segmentation techniques often prove to be the most appropriate choice [21,40]. These methods are inherently designed to understand and process the volumetric characteristics of the liver, which is a critical consideration for accurate segmentation results.
- Two-Dimensional Segmentation Limitations
 - Although 2D slice-based segmentation is widely used, it has limitations, particularly when it comes to dealing with complex organs such as the liver. The main challenge with 2D methods is their inability to fully capture all the regions of the liver. They involve working with individual slices, which can provide a fragmented understanding of the organ structure, but this fragmentation can lead to inconsistencies and errors when these individual slices are aggregated to form a complete image [41].
- Three-Dimensional Segmentation Advantages
 - In order to overcome the limitations of 2D segmentation, 3D segmentation has the ability to use more contextual information. Unlike 2D methods, which visualise the liver in individual slices, 3D techniques consider the organ as a whole, as they have the ability to ensure anatomical correctness by processing the liver as a single, continuous volume, avoiding errors that can arise from the aggregation of 2D slices [14,39]. In 2D segmentation, inconsistencies can occur when individual slices are combined, leading to inaccuracies in the representation of the liver anatomy. The holistic view provided by 3D segmentation results in more accurate segmentation, as it takes into account the spatial relationships and continuity between the different sections of the liver. The inclusion of this additional contextual information can potentially lead to segmentation results, especially in complex cases where the shape and size of the liver can vary considerably.

5.4. Exploring Research Questions

Following the specific analysis of the studies presented and their main findings, the questions raised in Section 1 are answered as follows.

- RQ1—What are the challenges and limitations associated with accurate liver segmentation in CT scans?
 - The challenges and limitations associated with accurate liver segmentation in CT images include under-segmentation, over-segmentation, low contrast, poor boundary detection, and background segmentation due to noise. In addition, liver segmentation in CT scans is further challenged by the presence of artefacts, such as partial volumes, noise, and low sharpness and contrast between organs, making it difficult to identify the boundaries between different tissues.
- RQ2—How does the choice of the method impact the accuracy and efficiency of liver segmentation in CT scans?
 - The choice of the method has an important impact on the accuracy and efficiency of liver segmentation in CT scans. Traditional techniques such as image processing and region growing approaches have shown varying degrees of sensitivity

and specificity, with some challenges in dealing with large injuries. In contrast, newer methods such as FCN and DBN-DNN and techniques like ResU-Net and SegNet showed a higher accuracy, with some reaching the highest accuracy levels. Notably, the use of GPUs has reduced processing times, thus contributing towards more efficient and accurate liver segmentation methods.

- RQ3—What are the evaluation metrics commonly used to assess the performance of AI models and traditional methods for liver segmentation in CT scans?
 - Some of the key metrics used to measure the outcome of segmentation techniques include the Dice Similarity Coefficient (DSC), accuracy, precision, sensitivity, specificity, and the segmentation speed. There is not much consistency in the metrics presented by the various studies except for the DSC.

6. Conclusions and Future Work

The evolution of liver segmentation techniques throughout history reflects the broad impact of AI technologies across a wide range of disciplines. The transition to fully automated segmentation methods has been an important breakthrough in the process, although the indispensable involvement of medical experts continues to play a key role in ensuring the accuracy and clinical relevance of these techniques. The emerging prevalence of 3D segmentation methods, which follow the structure of the liver, promises more accurate and anatomically consistent results.

However, there are a number of challenges that need to be addressed in order to advance the field. The lack of public datasets is one of the main barriers to the advancement of liver segmentation technologies. Research has mainly been conducted on private datasets, often restricted to specific medical centres, which limits wider participation in research and makes it difficult to replicate.

In terms of future developments, the outlook for the evolution of liver segmentation is unfolding in several core areas that promise to impact research and application in this field. The increasing availability of public datasets is key to fostering innovation, enabling the contribution of researchers from diverse backgrounds and promoting a dynamic research environment. In addition, the definition of specific standardised evaluation metrics is crucial to allow meaningful comparisons between segmentation methods and to guide the development towards more efficient, accurate, and user-friendly solutions, such as contour- and region-based metrics, performance metrics, and user intervention metrics.

The further application of 3D segmentation techniques may be a good investment, as they can provide more anatomically accurate and consistent anatomical results, overcoming the limitations of 2D segmentation. Furthermore, the effective integration of medical expertise in segmentation automation is indeed essential, with the aim of developing interfaces that allow specialists to interact with automated segmentation results, ensuring that liver segmentation tools are both technically advanced and clinically relevant and feasible. Collaboration between cutting-edge technology and human expertise is a good approach, combining the efficiency of automation with the refined understanding of healthcare professionals, whose oversight remains critical to maintain the accuracy and reliability of the final results.

Author Contributions: Conceptualisation, J.B.; Methodology, S.B.N. and J.B.; Software, S.B.N.; Validation, S.B.N., I.D. and J.B.; Formal analysis, S.B.N., I.D. and J.B.; Investigation, S.B.N.; Resources, S.B.N.; Data curation, S.B.N.; Writing—original draft preparation, S.B.N.; Writing—review and editing, S.B.N., I.D. and J.B.; Supervision, J.B. and I.D.; Project administration, J.B. and I.D.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Francque, S. The Liver and the Cardiovascular System: Two of a Kind? *J. Am. Heart Assoc.* **2021**, *10*, e020286. [\[CrossRef\]](#)
2. Trefts, E.; Gannon, M.; Wasserman, D.H. The liver. *Curr. Biol.* **2017**, *27*, R1147–R1151. [\[CrossRef\]](#)
3. Mendes, B.; Domingues, I.; Silva, A.; Santos, J. Prostate Cancer Aggressiveness Prediction Using CT Images. *Life* **2021**, *11*, 1164. [\[CrossRef\]](#)
4. Pereira, G.; Domingues, I.; Martins, P.; Abreu, P.H.; Duarte, H.; Santos, J. Registration of CT with PET: A Comparison of Intensity-Based Approaches. In *International Workshop on Combinatorial Image Analysis (IWCIA)*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 134–149.
5. Zhang, H.; Luo, K.; Deng, R.; Li, S.; Duan, S. Deep Learning-Based CT Imaging for the Diagnosis of Liver Tumor. *Comput. Intell. Neurosci.* **2022**, *2022*, 1–7. [\[CrossRef\]](#)
6. Vernuccio, F.; Cannella, R.; Bartolotta, T.V.; Galia, M.; Tang, A.; Brancatelli, G. Advances in liver US, CT, and MRI: moving toward the future. *Eur. Radiol. Exp.* **2021**, *5*, 52. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Domingues, I.; Cardoso, J.S. Using Bayesian surprise to detect calcifications in mammogram images. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 1091–1094.
8. Mendes, B.; Domingues, I.; Santos, J. Multi-class Semantic Segmentation for Prostate Cancer Radiotherapy Treatment Optimization. In Proceedings of the International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMA2SC), Porto, Portugal, 27–29 June 2022.
9. Bechar, M.E.A.; Settouti, N.; Domingues, I. Deep Learning vs. Super Pixel Classification for Breast Masses Segmentation. In *Deep Learning for Biomedical Applications*; CRC Press: Boca Raton, FL, USA, 2021; pp. 121–156.
10. Oliveira, A.C.; Domingues, I.; Duarte, H.; Santos, J.; Abreu, P.H. Going Back to Basics on Volumetric Segmentation of the Lungs in CT: A Fully Image Processing Based Technique. In *Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11868 LNCS, pp. 322–334.
11. Carbone, I.; Martins, P.; Teixeira, A.; Silva, A. A Vocal Tract Segmentation and Analysis over a European Portuguese MRI Database. *Electronica E Telecomunicações* **2008**, *4*, 1050–1053.
12. Zhou, L.Q.; Wang, J.Y.; Yu, S.Y.; Wu, G.G.; Wei, Q.; Deng, Y.B.; Wu, X.L.; Cui, X.W.; Dietrich, C.F. Artificial intelligence in medical imaging of the liver. *World J. Gastroenterol.* **2019**, *25*, 672–682. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Chlebus, G.; Schenk, A.; Moltz, J.H.; van Ginneken, B.; Hahn, H.K.; Meine, H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci. Rep.* **2018**, *8*, 15497. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Christ, P.F.; Ettlinger, F.; Grün, F.; Elshaera, M.E.A.; Lipkova, J.; Schlecht, S.; Ahmaddy, F.; Tatavirt, S.; Bickel, M.; Bilic, P.; et al. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. *arXiv* **2017**, arXiv: 1702.05970. <http://arxiv.org/abs/1702.05970>
15. Ansari, M.Y.; Abdalla, A.; Ansari, M.Y.; Ansari, M.I.; Malluhi, B.; Mohanty, S.; Mishra, S.; Singh, S.S.; Abinayed, J.; Al-Ansari, A.; et al. Practical utility of liver segmentation methods in clinical surgeries and interventions. *BMC Med. Imaging* **2022**, *22*, 97.
16. Le, D.C.; Chinnasarn, K.; Chansangrat, J.; Keeratibharat, N.; Horkaew, P. Semi-automatic liver segmentation based on probabilistic models and anatomical constraints. *Sci. Rep.* **2021**, *11*, 6106. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Halevi, G.; Moed, H.; Bar-Ilan, J. Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature. *J. Inf.* **2017**, *11*, 823–834. [\[CrossRef\]](#)
18. Sakshi; Kukreja, V. Image Segmentation Techniques: Statistical, Comprehensive, Semi-Automated Analysis and an Application Perspective Analysis of Mathematical Expressions. *Arch. Comput. Methods Eng.* **2023**, *30*, 457–495. [\[CrossRef\]](#)
19. Bae, K.T.; Giger, M.L.; Chen, C.T.; Kahn, C.E., Jr. Automatic segmentation of liver structure in CT images. *Med. Phys.* **1993**, *20*, 71–78. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Gao, L.; Heath, D.G.; Kuszyk, B.S.; Fishman, E.K. Automatic liver segmentation technique for three-dimensional visualization of CT data. *Radiology* **1996**, *201*, 359–364. [\[CrossRef\]](#)
21. Soler, L.; Malandain, G.; Montagnat, J.; Delingette, H.; Ayache, N.; Clément, J.M.; Roy, C.; Russier, Y.; Tassetti, V.; Marescaux, J. Automatic Segmentation of Portal Vein in CT-Scans of the Liver. In Proceedings of the World Congress on Medical Physics and Biomedical Engineering (MPBE), Nice, France, 14–19 September 1997; p. 788.
22. Yoo, S.W.; Cho, J.S.; Noh, S.M.; Shin, K.S.; Park, J.W. Advanced Liver Segmentation by Using Pixel Ratio in Abdominal CT Image. In Proceedings of the IEEE Conference, Pusan, Korea, 2000; The Institute of Electronics and Information Engineers: Seoul, Korea 2000; pp. 39–42.
23. Pan, S.; Dawant, B.M. Automatic 3D segmentation of the liver from abdominal CT images: a level-set approach. In *Medical Imaging: Image Processing*; SPIE: Bellingham, WA, USA, 2001; Volume 4322, pp. 128–138.
24. Saitoh, T.; Tamura, Y.; Kaneko, T. Automatic segmentation of liver region through blood vessels on multi-phase CT. In Proceedings of the International Conference on Pattern Recognition, Quebec, ON, Canada, 11–15 August 2002; IEEE: Piscataway, NJ, USA, Volume 1, 2002; pp. 735–738.
25. Masumoto, J.; Hori, M.; Sato, Y.; Murakami, T.; Johkoh, T.; Nakamura, H.; Tamura, S. Automated liver segmentation using multislice CT images. *Syst. Comput. Jpn.* **2003**, *34*, 71–82. [\[CrossRef\]](#)

26. Lim, S.J.; Jeong, Y.Y.; Lee, C.W.; Ho, Y.S. Automatic segmentation of the liver in CT images using the watershed algorithm based on morphological filtering. In Proceedings of the Medical Imaging: Image Processing, San Diego, CA, USA, 6–19 February 2004; SPIE: Bellingham, WA, USA, 2004; Volume 5370, pp. 1658–1666.
27. Liu, F.; Zhao, B.; Kijewski, P.K.; Wang, L.; Schwartz, L.H. Liver segmentation for CT images using GVF snake. *Med. Phys.* **2005**, *32*, 3699–3706. [\[CrossRef\]](#)
28. Lim, S.J.; Jeong, Y.Y.; Ho, Y.S. Automatic liver segmentation for volume measurement in CT Images. *J. Vis. Commun. Image Represent.* **2006**, *17*, 860–875. [\[CrossRef\]](#)
29. Beichel, R.; Bauer, C.; Bornik, A.; Sorantin, E.; Bischof, H. Liver Segmentation in CT Data: A Segmentation Refinement Approach. In *Segmentation in The Clinic: A Grand Challenge*, 2007; pp. 235–245. Available online: <https://graz.elsevierpure.com/en/publications/liver-segmentation-in-ct-data-a-segmentation-refinement-approach> (accessed on 4 March 2024).
30. Massoptier, L.; Casciaro, S. A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans. *Eur. Radiol.* **2008**, *18*, 1658–1665. [\[CrossRef\]](#)
31. Heimann, T.; Van Ginneken, B.; Styner, M.A.; Arzhaeva, Y.; Aurich, V.; Bauer, C.; Beck, A.; Becker, C.; Beichel, R.; Bekes, G.; et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Med. Imaging* **2009**, *28*, 1251–1265. [\[CrossRef\]](#)
32. Akram, M.U.; Khanum, A.; Iqbal, K. An automated system for liver CT enhancement and segmentation. *Icgst J. Graph. Vis. Image Process. (ICGST-GVIP)* **2010**, *10*, 17–22.
33. Oliveira, D.A.; Feitosa, R.Q.; Correia, M.M. Segmentation of liver, its vessels and lesions from CT images for surgical planning. *Biomed. Eng. Online* **2011**, *10*, 1–23. [\[CrossRef\]](#)
34. Linguraru, M.G.; Richbourg, W.J.; Liu, J.; Watt, J.M.; Pamulapati, V.; Shijun Wang.; Summers, R.M. Tumor Burden Analysis on Computed Tomography by Automated Liver and Tumor Segmentation. *IEEE Trans. Med. Imaging* **2012**, *31*, 1965–1976. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Li, X.; Luo, S.; Li, J. Liver Segmentation from CT Image Using Fuzzy Clustering and Level Set. *J. Signal Inf. Process.* **2013**, *4*, 36–42. [\[CrossRef\]](#)
36. Jena, B.; Krishna Nayak, G.; Saxena, S. Comprehensive Review of Abdominal Image Segmentation using Soft and Hard Computing Approaches. In Proceedings of the International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 13–14 March 2020; pp. 1–5.
37. Platero, C.; Tobar, M.C.; et al. A multias segmentation using graph cuts with applications to liver segmentation in CT scans. *Comput. Math. Methods Med.* **2014**, *2014*, 182909. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Mostafa, A.; Fouad, A.; Abd Elfattah, M.; Hassanien, A.E.; Hefny, H.; Zhu, S.Y.; Schaefer, G. CT liver segmentation using artificial bee colony optimisation. *Procedia Comput. Sci.* **2015**, *60*, 1622–1630. [\[CrossRef\]](#)
39. Dou, Q.; Chen, H.; Jin, Y.; Yu, L.; Qin, J.; Heng, P.A. 3D deeply supervised network for automatic liver segmentation from CT volumes. In Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Athens, Greece, 17–21 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 149–157.
40. Hiranman, A. Liver Segmentation Using 3D CT Scans. Ph.D. Thesis, University of Kwazulu-Natal, Durban, South Africa, 2018.
41. Wang, K.; Mamidipalli, A.; Retson, T.; Bahrami, N.; Hasenstab, K.; Blansit, K.; Bass, E.; Delgado, T.; Cunha, G.; Middleton, M.S.; et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. *Radiol. Artif. Intell.* **2019**, *1*, 180022. [\[CrossRef\]](#)
42. Almotairi, S.; Kareem, G.; Aouf, M.; Almutairi, B.; Salem, M.A.M. Liver tumor segmentation in CT scans using modified SegNet. *Sensors* **2020**, *20*, 1516. [\[CrossRef\]](#)
43. Soler, L.; Hostettler, A.; Agnus, V.; Charnoz, A.; Fasquel, J.; Moreau, J.; Osswald, A.; Bouhadjar, M.; Marescaux, J. *3D Image Reconstruction for Comparison of Algorithm Database: A Patient Specific Anatomical and Medical Image Database*; Technical Report; IRCAD: Strasbourg, France, 2010.
44. Ayalew, Y.A.; Fante, K.A.; Mohammed, M.A. Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method. *BMC Biomed. Eng.* **2021**, *3*, 4. [\[CrossRef\]](#)
45. LiTS Challenge Dataset. Available online: <https://competitions.codalab.org/competitions/17094> (accessed on 27 November 2023).
46. Scicluna, D. Automatic Segmentation of Healthy Liver in Abdominal Computed Tomography Scans. Master's Thesis, University of Malta, Msida, Malta, 2022.
47. Kavur, A.E.; Gezer, N.S.; Barış, M.; Aslan, S.; Conze, P.H.; Groza, V.; Pham, D.D.; Chatterjee, S.; Ernst, P.; Özkan, S.; et al. CHAOS Challenge-combined (CT-MR) healthy abdominal organ segmentation. *Med. Image Anal.* **2021**, *69*, 101950. [\[CrossRef\]](#)
48. Ezzat, K.A.; Omran, L.N.; El Seddawy, A.I.B. Automatic liver segmentation in computed tomography scans using deep semantic segmentation. *Bull. Electr. Eng. Inform.* **2023**, *12*, 250–256. [\[CrossRef\]](#)
49. Shao, J.; Luan, S.; Ding, Y.; Xue, X.; Zhu, B.; Wei, W. Attention Connect Network for Liver Tumor Segmentation from CT and MRI Images. *Technol. Cancer Res. Treat.* **2024**, *23*, 15330338231219366. [\[PubMed\]](#)
50. Mauricio, J.; Domingues, I.; Bernardino, J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Appl. Sci.* **2023**, *13*, 5521. [\[CrossRef\]](#)

51. Priyadarsini, S.; Selvathi, D. Survey on segmentation of liver from CT images. In Proceedings of the IEEE international conference on advanced communication control and computing technologies (ICACCCT), Ramanathapuram, India, 23–25 August 2012; pp. 234–238.
52. Mharib, A.M.; Ramli, A.R.; Mashohor, S.; Mahmood, R.B. Survey on liver CT image segmentation methods. *Artif. Intell. Rev.* **2012**, *37*, 83–95. [\[CrossRef\]](#)
53. Khan, F. Automated segmentation of CT liver images: A review. *J. Commun. Technol. Electron. Comput. Sci.* **2018**, *19*, 5–9.
54. Moghbel, M.; Mashohor, S.; Mahmud, R.; Saripan, M.I.B. Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artif. Intell. Rev.* **2018**, *50*, 497–537. [\[CrossRef\]](#)
55. Vanmore, S.V.; Chougule, S.R. Survey on automatic liver segmentation techniques from abdominal CT images. In Proceedings of the International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 22–24 May 2024; IEEE: Piscataway, NJ, USA, 2019; pp. 1030–1035.
56. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. [\[CrossRef\]](#)
57. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis. (IJCV)* **2015**, *115*, 211–252. [\[CrossRef\]](#)
58. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597. <http://arxiv.org/abs/1505.04597>
59. Mir, R.; Kelly, S.M.; Xiao, Y.; Moore, A.; Clark, C.H.; Clementel, E.; Coming, C.; Ebert, M.; Hoskin, P.; Hurkmans, C.W.; et al. Organ at risk delineation for radiation therapy clinical trials: Global Harmonization Group consensus guidelines. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* **2020**, *150*, 30–39.
60. Sarria, G.R.; Kugel, F.; Roehner, F.; Layer, J.; Dejonckheere, C.; Scafa, D.; Koeksal, M.; Leitzen, C.; Schmeel, L.C. Artificial Intelligence-Based Autosegmentation: Advantages in Delineation, Absorbed Dose-Distribution, and Logistics. *Adv. Radiat. Oncol.* **2024**, *9*, 101394. [\[PubMed\]](#)
61. Al-Saeed, Y.; Gab-Allah, W.; Elmogy, M. Hepatic tumors diagnosis system based on fuzzy c-means using computed tomography images. *Res. Sq. Prepr.* **2022**. [\[CrossRef\]](#)
62. MICCAI-Sliver07. Available online: <https://sliver07.grand-challenge.org/> (accessed on 27 November 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Appendix F – MDPI Encyclopedia Journal Entry

Publication of an entry (**Algorithms for Liver Segmentation in Computed Tomography Scans**) in *MDPI Encyclopedia* journal:

Algorithms for Liver Segmentation in Computed Tomography Scans

Subjects: Engineering, Biomedical

Contributor: Stephanie Batista Niño, Jorge Bernardino, Inês Domingues

Oncology has emerged as a crucial field of study in the domain of medicine. Computed tomography has gained widespread adoption as a radiological modality for the identification and characterisation of pathologies, particularly in oncology, enabling precise identification of affected organs and tissues. However, achieving accurate liver segmentation in computed tomography scans remains a challenge due to the presence of artefacts and the varying densities of soft tissues and adjacent organs.

Keywords: artificial intelligence ; computed tomography ; hepatic pathologies ; liver segmentation

1. Introduction

As one of the most important organs in the digestive system, the liver performs critical functions such as breaking down nutrients, producing bile and eliminating toxic substances. However, liver-related diseases, particularly oncological diseases, pose health risks, and liver cancer is a leading cause of cancer-related mortality worldwide [1,2]. Computed tomography (CT) has become an integral part of diagnosis, treatment planning and monitoring the progress of oncological diseases [3,4], providing detailed cross-sectional images for accurate visualisation of internal structures, including liver tumours [5,6].

With the advancement of technology and artificial intelligence (AI) in medicine, there is a growing need to optimise the identification of oncological diseases [7]. Medical image segmentation is emerging as a fundamental step in the pipeline [8,9,10,11]. Liver segmentation in CT scans has emerged as a critical area, requiring accurate identification and delimitation of the liver region for treatment planning and progress monitoring, as well as for early detection of liver lesions and metastases to other organs [12,13,14]. However, accurate liver segmentation on CT scans is challenging due to factors such as artefacts, varying soft tissue densities and the complexity caused by adjacent organ proximity [15,16].

2. Historical Overview

The oldest work found to tackle liver segmentation is the one by Bae et al. (1993) [17], presenting a similar sequential image-by-image segmentation technique using a reference image, where the liver occupies a significant portion of the abdomen cross-section. Image processing techniques, including grey-level thresholding, Gaussian smoothing and connectivity tracking, are employed to extract the liver boundaries. The resulting boundaries are then smoothed using mathematical morphology techniques and B-splines. This study focuses on a living-donor liver transplant program, and the computer-determined boundaries are compared with those drawn by a radiologist, showing agreement within 10% of the calculated areas.

Gao et al. (1996) [18] focus on facilitating 3D visualisations for surgical planning. Their method employs a global histogram analysis, morphologic operations and a parametrically deformable contour model to delineate the liver boundary. Ten cases were used to validate the approach and promising results were found with minimal operator intervention required.

Soler et al. (1997) [19] propose an automatic method for segmenting the portal vein, with the primary objective of achieving accurate segmentation with detailed branching and topological information, facilitating the localisation of liver tumours concerning Couinaud's anatomical segmentation. This approach involves the initial detection of liver contours using 3D deformable models, followed by limiting the CT images to a liver mask containing hepatic tissue, vascular trees and potential tumours. Classification of anatomical structures is performed using Gaussian curves fitted to an intensity histogram. The vascular trees and tumours are segmented through a hysteresis thresholding technique based on a distance map, considering the Gaussian parameters. An isotropic image is obtained through shape-based interpolation and the portal vein is reconstructed using skeletonization, eliminating short branches and correcting errors. Results

demonstrate that the algorithm automatically extracts the first three main bifurcations of the portal vein, comparable to manual segmentation.

Yoo et al. (2000) ^[20] focus on the use of pixel ratios. By analysing the grey value range of a normal liver in CT images, a binary image is generated and then processed into four mesh images based on hole ratios to eliminate noise. A template representing the general outline of the liver is generated from the union image of these mesh images and subtracted from the binary image to accurately represent the organ boundary. The pixel ratio, which takes into account the distribution of organ pixels, was used to discriminate between the organ and noise, especially in cases where organs have similar grey value ranges. The proposed method reduced the processing time compared to existing methods and was validated against manual segmentation by medical experts.

Pan and Dawant (2001) ^[21] introduce a level-set approach, which addresses the challenge of defining appropriate speed functions for contour propagation. A speed function is proposed to stop the propagation of the contour at organ boundaries with weak edges by incorporating the accumulative speed based on the path of the contour, enhancing the robustness of segmentation in noisy images. The method also leverages a priori anatomical information to improve the accuracy. Tested on five CT datasets, including cases with abnormal livers, the method demonstrates good agreement with manual delineations.

Saitoh et al. (2002) ^[22] present an automated method for segmenting the liver region from the third phase of abdominal CT scans. Their approach involves the extraction of blood vessels using a threshold, followed by morphological dilation to define an approximate liver region useful for the removal of adjacent organs. The final liver region is then extracted using a threshold. The method is thus based on mathematical morphology and thresholding techniques, using the unique characteristics of blood vessels to functionally identify the liver region. The experiments performed on eight CT datasets show a good agreement between the automatically and manually detected liver regions.

Masumoto et al. (2003) ^[23] use multislice CT images. Their method uses two time-varying images acquired during the contrast medium circulation phase, highlighting the liver region through CT value changes. The proposed scheme involves generating a liver likelihood image by analysing CT value changes and subsequently extracting the liver region while considering the geometric characteristics of blood vessels and tumours. The evaluation, based on receiver operating characteristic (ROC) analyses, demonstrates the superiority of the proposed method over other approaches, especially when using information from both phases.

The scheme proposed by Lim et al. (2004) ^[24] uses an ROI approach to optimise computational efficiency. Morphological filters, incorporating a priori knowledge of liver location and intensity, detect the initial boundary. The algorithm then generates a gradient image using the weighted initial boundary and employs an immersion-based watershed algorithm for segmentation. Post-processing includes region merging based on statistical information to refine the segmentation.

Liu et al. (2005) ^[25] present a gradient vector flow (GVF) snake-based method for the semi-automatic segmentation of liver volumes in contrast-enhanced CT images. The algorithm follows a stepwise approach, starting with the computation of an initial edge map using the Canny edge detector and the estimation of a liver template. The edge map is then modified to suppress edges within the liver using the liver template, and a concavity removal algorithm is applied to refine the liver boundary. The GVF field is computed based on the modified edge map, and the initial liver contour is determined by considering the candidate initial contour and the computed GVF field. The final liver contour is obtained by deforming the initial contour using the snake. The method was evaluated on 20 contrast-enhanced volumetric liver images, and the results were compared with a radiologist's manual delineation. The median difference ratio between the computer-generated results and manual results is 5.3%, with a range of 2.9% to 7.6%.

A three-stage approach is used by Lim et al. (2006) ^[26]. The first stage involves image simplification as preprocessing, where an ROI is identified and thresholds are determined using multilevel thresholding. The second stage detects a search range using multiscale morphological filtering, region labelling, and partition clustering. The third stage uses a contour-based segmentation approach with a labelling-based search algorithm to refine the initial liver boundary. The effectiveness of the algorithm is demonstrated through experimental results on contrast-enhanced abdominal CT images, with an average segmentation accuracy of 96%. Volume measurement is performed based on the segmented liver regions, with an average error rate of 3%.

Beichel et al. (2007) ^[27] introduce a two-step process. First, initial segmentation is performed using graph cuts, overcoming challenges such as the high variability in liver shape and grey-value appearance. Second, an interactive refinement step is introduced, allowing users to correct segmentation errors in a 3D environment. The refinement is facilitated by a hybrid desktop/virtual reality (VR) user interface. This approach is demonstrated on ten contrast-enhanced

2D Liver Segmentation in Abdominal Computed Tomography Scans

liver CT scans, demonstrating robustness to variations in patient data. The results also indicate an improved segmentation quality with low interaction times.

Massoptier and Casciaro (2008) ^[28] present a fully automated method that uses a statistical-model-based approach to distinguish liver tissue from other abdominal organs. An active contour technique using gradient vector flow is used for smoother segmentation of the liver surface. Automatic classification is performed to isolate hepatic lesions from liver parenchyma. The method was evaluated on 21 datasets and demonstrated robust and efficient liver and lesion segmentations close to the ground truth, with an average processing time of 11.4 s per 512×512 pixel slice. The volume overlap for liver surface segmentation is 94.2%, and the accuracy is 3.7 mm. Tumour detection achieved a sensitivity and specificity of 82.6% and 87.5%, respectively.

Heimann et al. (2009) ^[29] focus on the comparison and evaluation of different methods. The image data, acquired from different CT scanners, consisted of contrast-dye-enhanced scans showing pathological conditions like tumours and cysts. Radiology experts manually delineated the liver contours in transversal slices to create reference segmentations. A total of 40 images were divided into training and test sets for algorithm evaluation. Evaluation measures included volumetric overlap, relative volume difference, and surface distances. Fully automated and interactive segmentation methods were employed, with the former showing discernible performance differences. The best-performing automated approaches used statistical shape models. Interactive methods achieved higher scores with more user interaction. A combined approach using majority voting from the best-performing methods outperformed individual automated and interactive results.

A three-step procedure is outlined by Akram et al. (2010) ^[30]. Firstly, a pre-processing step involves converting the image to greyscale and applying a 3×3 median filter to reduce noise. The second step focuses on liver segmentation, with a global threshold and morphological operations to obtain the final segmented liver region. Finally, post-processing steps include adaptive histogram equalisation, Gaussian smoothing, and grey-level transformations to enhance the segmented liver region. Experimental tests on 100 CT images demonstrate the accuracy of the proposed method by comparing automated segmentation results with images manually segmented by hepatologists and oncologists.

The approach of Oliveira et al. (2011) ^[31] involves a sequence of four steps. First, the liver is segmented using level sets with parameters optimised by a genetic algorithm (GA). A Gaussian fit is employed to define the speed image for level set propagation. Secondly, vessels and nodules are segmented using a Gaussian mixture model, focusing on adipose nodules. A region-growing method with information from the Gaussian model is applied. Thirdly, vessels are classified into portal veins or hepatic veins using a vein tracking method. Finally, a geometric approach based on the identified veins is used to segment the liver into different Couinaud regions. Liver segmentation is based on the assumption that the liver parenchyma homogeneity and veins are mainly inside the liver. The parameters are estimated using a GA, and the fitness evaluation involves comparing the segmentation with a reference using five disparity metrics. The proposed method shows good performance, ranking among the top methods in the MICCAI-SLiver07 conference evaluation.

The method developed by Linguraru et al. (2012) ^[32] uses a robust parametrisation of 3D surfaces for point-to-point correspondence, overcoming challenges such as inconsistent contrast enhancements and imaging artefacts. A shape descriptor that is invariant under rotation and scale is used to compare the local shape features of organs. Initial liver segmentation is refined using a shape-driven geodesic active contour, and hepatic tumours are detected and segmented using graph cuts and support vector machines (SVMs). This technique is evaluated on a dataset of 101 CT scans and shows improvements in the liver segmentation accuracy, particularly in cases with large tumours and segmentation errors. Furthermore, the method identifies liver tumours with a low rate of false positives.

Li et al. (2013) ^[33] discuss a method that makes use of fuzzy clustering and level set techniques. The Fuzzy C-Means (FCM) clustering algorithm is employed, which assigns pixels to different categories based on fuzzy memberships, considering both the grey level intensity and spatial information. The FCM algorithm is iteratively optimised by minimising a cost function, allowing the fuzziness of the resulting partition. To overcome the limitations of standard FCM, a spatial FCM algorithm is introduced that incorporates spatial information into fuzzy membership functions. This paper also introduces the level set method, a continuous deformable model for segmentation. Distance-Regularized Level Set Evolution (DRLSE) is proposed to address reinitialisation issues and improve the efficiency. The proposed method is evaluated using accuracy, sensitivity, and specificity metrics and demonstrates a high performance in liver segmentation, especially in cases with unclear boundaries. A comprehensive review of abdominal image segmentation using soft and hard computing approaches is provided in ^[34].

Platero et al. (2014) ^[35] integrate a multi-atlas segmentation approach with graph cuts. Their method includes several steps: (1) obtaining an initial solution using low-level operations to define the ROI around the liver; (2) constructing a fast probabilistic atlas for the ROI and computing a coarse binary segmentation using segmentation-affine registration; (3) ranking the atlases based on segmentation similarity and propagating selected atlases to the target image; (4) improving the segmentation accuracy through label fusion, minimising the discrete energy function; and (5) evaluating the approach using a public liver segmentation database. The experimental results show a high accuracy, competitive with human expert segmentation.

Artificial Bee Colony (ABC) optimisation is used by Mostafa et al. (2015) ^[36]. Their algorithm use ABC optimisation to cluster different intensity values in abdominal CT images, followed by mathematical morphological operations to manipulate and separate the clusters. This process eliminates small and thin regions, such as flesh regions or organ edges. The extracted regions form an initial estimate of the liver area, which is further enhanced using a region-growing technique. The proposed approach demonstrates a segmentation accuracy of 93.73% on a test dataset of 38 CT images, taken in the pre-contrast phase.

A 3D deeply supervised network (DSN) is introduced by Dou et al. (2016) ^[37]. The proposed architecture consists of 11 layers, including 6 convolutional layers, 2 max-pooling layers, 2 deconvolution layers, and 1 softmax layer. The network is designed in a 3D format to effectively capture spatial information. The 3D DSN employs deep supervision via additional deconvolutional layers to counteract vanishing gradients, thus improving the training process. The learning objective is to minimise per-voxel-wise binary classification errors, with deep supervision injected at specific layers. The MICCAI-SLiver07 dataset is used for evaluation, demonstrating that the 3D DSN has a faster convergence and lower errors when compared to traditional 3D convolutional neuronal networks (CNNs).

Christ et al. (2017) ^[34] propose a cascaded fully CNN (CFCN) on CT slices that sequentially segments the liver and lesions. First, various preprocessing steps, including Hounsfield unit windowing and contrast enhancement, are applied. Then, the cascaded approach involving two U-Net architectures is used for liver and lesion segmentation. Finally, 3D conditional random fields (CRFs) are used to refine the segmentation results. Generalisation and scalability to different modalities and real-life datasets, including a diffusion-weighted magnetic resonance imaging (MRI) dataset and a large multi-centre CT dataset, are shown.

Hiraman (2018) ^[38] presents a slice alignment method that addresses the challenges through optimal threshold selection, skeletonization, and enhanced correlation coefficient (ECC) alignment. Next, a CNN-based liver region of interest detection method is proposed to classify 2D slices for focused processing.

The study presented by Wang et al. (2019) ^[39] investigates the application of a generalised CNN for automated liver segmentation and biometry using cross-sectional data from abdominal CT and MRI scans. Their retrospective study included a sample of 563 abdominal scans from 530 adults, covering different imaging modalities. The CNN was initially trained on 300 unenhanced multiecho 2D SPGR MRI sets and then subjected to transfer learning for generalisation across different imaging methods. The accuracy of the CNN was evaluated using internal and external validation datasets. This study also investigates the impact of training data size on the segmentation accuracy and explores the feasibility of using automated liver segmentation for volumetry and hepatic PDFF quantification.

Almotairi et al. (2020) ^[40] explore the application of the SegNet architecture. The proposed modified SegNet model uses the VGG-16 network as an encoder. Tests were performed on a standard dataset for liver CT scans (3D-IRCADb01 ^[41]), achieving a tumour accuracy of up to 99.9% in the training phase and 86% for tumour identification.

Ayalew et al. (2021) ^[42] present a modified U-Net architecture and introduce a new class balancing method. To address the class imbalance between the liver and tumours, a weighting factor is applied and slices without a tumour are removed during data preparation. The U-Net-based network architecture includes batch normalisation, dropout layers, and filter size reduction. Training involves tuning hyperparameters, such as the learning rate and batch size. The datasets used are derived from the 3D-IRCADb01 ^[41] and LITS ^[43] databases and the results achieve a Dice Similarity Coefficient (DSC) of 0.96 and 0.74, respectively. The algorithm also introduces a novel approach for direct tumour segmentation from abdominal CT scan images, with a comparable performance to existing two-step methods.

The study of Scicluna (2022) ^[44] is motivated by challenges such as the Combined Healthy Abdominal Organ Segmentation (CHAOS) Challenge ^[45], which focuses on healthy abdominal organs. The study focuses on replicating the v16pUNet1.1C model, which demonstrated a superior performance in Task 2 of the CHAOS Challenge. Results from the v16pUNet1.1C model are presented and compared with variations in the loss function and scaling transformation. The application of a 3D largest-connected-component filter is discussed, showing improvements in mean scores.

2D Liver Segmentation in Abdominal Computed Tomography Scans

A deep semantic segmentation CNN is used by Ezzat et al. (2023) [46]. A three-stage architecture is proposed, including pre-processing with data augmentation, deep CNN training, and testing. The CNN-based semantic segmentation model is shown to be robust, achieving a test accuracy of 98.8%. The approach does not require user input, making it accessible to non-experts.

Shao et al. (2024) [47] present the Attention Connect Network (AC-Net) for liver tumour segmentation in CT and MRI images. The AC-Net consists of two main modules: an axial attention module (AAM) and a Vision Transformer module (VTM). The AAM uses an axial attention mechanism to merge features of matching dimensions, maximising the use of spatial features extracted by a CNN. The VTM processes high-level semantic features extracted by the CNN using a methodology similar to Vision Transformers (ViTs) [48]. The network achieves a DSC of 0.90, a Jaccard coefficient (JC) of 0.82, a recall of 0.92, a precision of 0.89, a Hausdorff distance (HD) of 11.96, and an average symmetric surface distance (ASSD) of 4.59.

3. Public Dataset Analysis

The most common public datasets used in studies of liver segmentation on CT scans include 3D-IRCADb01, LITS17, and MICCAI-SLiver07 (Figure 1). According to a dataset comparison provided by Al-Saeed et al. [49] (shown in Table 1), it is possible to identify several key differences that may have implications for data processing and analysis, such as different formats and differences in resolution between datasets which may require different approaches to processing and interpretation.

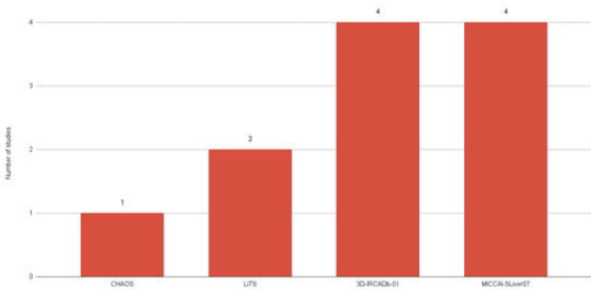


Figure 1. Number of studies using each public dataset.

Table 1. Characteristics of each dataset used by the main analysed studies (based on [49]).

Dataset	Date	Format	Number of Subjects	Slices per Subject	Resolution
MICCAI-SLiver07 [50]	2007	RAW	30	74 to 260	512 × 512
3D-IRCADb01 [41]	2010	DICOM	20	74 to 260	512 × 512
LITS17 [43]	2017	RAW	200	42 to 1024	Variable

4. Impact of the Adoption of Neuronal-Network-Based Methods

The growth of neuronal-network-based approaches has led to remarkable progress in liver segmentation, particularly with respect to CT scans. These models have led to a new era of accuracy and efficiency, significantly outperforming traditional methods [51][52]. This improved accuracy has become critical in the field of medical imaging, where the correct interpretation of CT scans can directly affect the diagnosis and treatment plans of patients. Furthermore, the efficiency of these neuronal network models translates into faster processing times, allowing for more agile decision making in clinical settings.

Another positive aspect of neuronal networks in liver segmentation is their ability to cope with the complexity of the liver anatomy. Neuronal networks, with their systematic and complex pattern recognition capabilities, are able to navigate these anatomical variations. As a result, they are better able to deal with the variety of appearances that liver tissue can have on CT scans. This ability to handle complex datasets ensures that neuronal networks can provide consistent and accurate segmentation in a wide range of cases.

5. Comparison between 2D and 3D Methods for Liver Segmentation

- Importance of Choosing between 2D and 3D Methods
 - In medical imaging, and in particular liver segmentation, the choice between slice-based 2D and volume-based 3D segmentation methods is crucial. This decision is highly dependent on the anatomical structure of the liver. Given the complex, three-dimensional nature of the liver, 3D segmentation techniques often prove to be the most appropriate choice ^{[19][38]}. These methods are inherently designed to understand and process the volumetric characteristics of the liver, which is a critical consideration for accurate segmentation results.
- Two-Dimensional Segmentation Limitations
 - Although 2D slice-based segmentation is widely used, it has limitations, particularly when it comes to dealing with complex organs such as the liver. The main challenge with 2D methods is their inability to fully capture all the regions of the liver. They involve working with individual slices, which can provide a fragmented understanding of the organ structure, but this fragmentation can lead to inconsistencies and errors when these individual slices are aggregated to form a complete image ^[39].
- Three-Dimensional Segmentation Advantages
 - In order to overcome the limitations of 2D segmentation, 3D segmentation has the ability to use more contextual information. Unlike 2D methods, which visualise the liver in individual slices, 3D techniques consider the organ as a whole, as they have the ability to ensure anatomical correctness by processing the liver as a single, continuous volume, avoiding errors that can arise from the aggregation of 2D slices ^{[44][22]}. In 2D segmentation, inconsistencies can occur when individual slices are combined, leading to inaccuracies in the representation of the liver anatomy. The holistic view provided by 3D segmentation results in more accurate segmentation, as it takes into account the spatial relationships and continuity between the different sections of the liver. The inclusion of this additional contextual information can potentially lead to segmentation results, especially in complex cases where the shape and size of the liver can vary considerably.

References

1. Francque, S. The Liver and the Cardiovascular System: Two of a Kind? J. Am. Heart Assoc. 2021, 10, e020286.
2. Trefts, E.; Gannon, M.; Wasserman, D.H. The liver. Curr. Biol. 2017, 27, R1147–R1151.
3. Mendes, B.; Domingues, I.; Silva, A.; Santos, J. Prostate Cancer Aggressiveness Prediction Using CT Images. Life 2021, 11, 1164.
4. Pereira, G.; Domingues, I.; Martins, P.; Abreu, P.H.; Duarte, H.; Santos, J. Registration of CT with PET: A Comparison of Intensity-Based Approaches. In International Workshop on Combinatorial Image Analysis (IWCI); Springer: Berlin/Heidelberg, Germany, 2018; pp. 134–149.
5. Zhang, H.; Luo, K.; Deng, R.; Li, S.; Duan, S. Deep Learning-Based CT Imaging for the Diagnosis of Liver Tumor. Comput. Intell. Neurosci. 2022, 2022, 1–7.
6. Vernuccio, F.; Cannella, R.; Bartolotta, T.V.; Galia, M.; Tang, A.; Brancatelli, G. Advances in liver US, CT, and MRI: Moving toward the future. Eur. Radiol. Exp. 2021, 5, 52.
7. Domingues, I.; Cardoso, J.S. Using Bayesian surprise to detect calcifications in mammogram images. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 1091–1094.
8. Mendes, B.; Domingues, I.; Santos, J. Multi-class Semantic Segmentation for Prostate Cancer Radiotherapy Treatment Optimization. In Proceedings of the International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMA2SC), Porto, Portugal, 27–29 June 2022.
9. Bechar, M.E.A.; Settouti, N.; Domingues, I. Deep Learning vs. Super Pixel Classification for Breast Masses Segmentation. In Deep Learning for Biomedical Applications; CRC Press: Boca Raton, FL, USA, 2021; pp. 121–156.
10. Oliveira, A.C.; Domingues, I.; Duarte, H.; Santos, J.; Abreu, P.H. Going Back to Basics on Volumetric Segmentation of the Lungs in CT: A Fully Image Processing Based Technique. In Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA); Springer: Berlin/Heidelberg, Germany, 2019; Volume 11868 LNCS, pp. 322–334.

2D Liver Segmentation in Abdominal Computed Tomography Scans

11. Carbone, I.; Martins, P.; Teixeira, A.; Silva, A. A Vocal Tract Segmentation and Analysis over a European Portuguese MRI Database. *Electrónica E Telecomunicações* 2008, 4, 1050–1053.
12. Zhou, L.Q.; Wang, J.Y.; Yu, S.Y.; Wu, G.G.; Wei, Q.; Deng, Y.B.; Wu, X.L.; Cui, X.W.; Dietrich, C.F. Artificial intelligence in medical imaging of the liver. *World J. Gastroenterol.* 2019, 25, 672–682.
13. Chlebus, G.; Schenk, A.; Moltz, J.H.; van Ginneken, B.; Hahn, H.K.; Meine, H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci. Rep.* 2018, 8, 15497.
14. Christ, P.F.; Ettlinger, F.; Grün, F.; Elshaera, M.E.A.; Lipkova, J.; Schlecht, S.; Ahmaddy, F.; Tatavarty, S.; Bickel, M.; Bilic, P.; et al. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. *arXiv* 2017, arXiv:1702.05970.
15. Ansari, M.Y.; Abdalla, A.; Ansari, M.Y.; Ansari, M.I.; Malluhi, B.; Mohanty, S.; Mishra, S.; Singh, S.S.; Abinayed, J.; Al-Ansari, A.; et al. Practical utility of liver segmentation methods in clinical surgeries and interventions. *BMC Med. Imaging* 2022, 22, 97.
16. Le, D.C.; Chinnasam, K.; Chansangrat, J.; Keeratibharat, N.; Horkaew, P. Semi-automatic liver segmentation based on probabilistic models and anatomical constraints. *Sci. Rep.* 2021, 11, 6106.
17. Bae, K.T.; Giger, M.L.; Chen, C.T.; Kahn, C.E., Jr. Automatic segmentation of liver structure in CT images. *Med. Phys.* 1993, 20, 71–78.
18. Gao, L.; Heath, D.G.; Kuszyk, B.S.; Fishman, E.K. Automatic liver segmentation technique for three-dimensional visualization of CT data. *Radiology* 1996, 201, 359–364.
19. Soler, L.; Malandain, G.; Montagnat, J.; Delingette, H.; Ayache, N.; Clément, J.M.; Roy, C.; Russier, Y.; Tasseti, V.; Marescaux, J. Automatic Segmentation of Portal Vein in CT-Scans of the Liver. In *Proceedings of the World Congress on Medical Physics and Biomedical Engineering (MPBE)*, Nice, France, 14–19 September 1997; p. 788.
20. Yoo, S.W.; Cho, J.S.; Noh, S.M.; Shin, K.S.; Park, J.W. Advanced Liver Segmentation by Using Pixel Ratio in Abdominal CT Image. In *Proceedings of the IEEK Conference*, Pusan, Korea, 2000; The Institute of Electronics and Information Engineers: Seoul, Korea, 2000; pp. 39–42.
21. Pan, S.; Dawant, B.M. Automatic 3D segmentation of the liver from abdominal CT images: A level-set approach. In *Medical Imaging: Image Processing*; SPIE: Bellingham, WA, USA, 2001; Volume 4322, pp. 128–138.
22. Saitoh, T.; Tamura, Y.; Kaneko, T. Automatic segmentation of liver region through blood vessels on multi-phase CT. In *Proceedings of the International Conference on Pattern Recognition*, Quebec, ON, Canada, 11–15 August 2002; IEEE: Piscataway, NJ, USA, 2002; Volume 1, pp. 735–738.
23. Masumoto, J.; Hori, M.; Sato, Y.; Murakami, T.; Johkoh, T.; Nakamura, H.; Tamura, S. Automated liver segmentation using multislice CT images. *Syst. Comput. Jpn.* 2003, 34, 71–82.
24. Lim, S.J.; Jeong, Y.Y.; Lee, C.W.; Ho, Y.S. Automatic segmentation of the liver in CT images using the watershed algorithm based on morphological filtering. In *Proceedings of the Medical Imaging: Image Processing*, San Diego, CA, USA, 6–19 February 2004; SPIE: Bellingham, WA, USA, 2004; Volume 5370, pp. 1658–1666.
25. Liu, F.; Zhao, B.; Kijewski, P.K.; Wang, L.; Schwartz, L.H. Liver segmentation for CT images using GVF snake. *Med. Phys.* 2005, 32, 3699–3706.
26. Lim, S.J.; Jeong, Y.Y.; Ho, Y.S. Automatic liver segmentation for volume measurement in CT Images. *J. Vis. Commun. Image Represent.* 2006, 17, 860–875.
27. Beichel, R.; Bauer, C.; Bornik, A.; Sorantin, E.; Bischof, H. Liver Segmentation in CT Data: A Segmentation Refinement Approach. In *Segmentation in The Clinic: A Grand Challenge*; 2007; pp. 235–245. Available online: <https://graz.elsevierpure.com/en/publications/liver-segmentation-in-ct-data-a-segmentation-refinement-approach> (accessed on 4 March 2024).
28. Massotier, L.; Casciaro, S. A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans. *Eur. Radiol.* 2008, 18, 1658–1665.
29. Heimann, T.; Van Ginneken, B.; Styner, M.A.; Arzhaeva, Y.; Aurich, V.; Bauer, C.; Beck, A.; Becker, C.; Beichel, R.; Bekes, G.; et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Med. Imaging* 2009, 28, 1251–1265.
30. Akram, M.U.; Khanum, A.; Iqbal, K. An automated system for liver CT enhancement and segmentation. *Icgst J. Graph. Vis. Image Process. (ICGST-GVIP)* 2010, 10, 17–22.
31. Oliveira, D.A.; Feitosa, R.Q.; Correia, M.M. Segmentation of liver, its vessels and lesions from CT images for surgical planning. *Biomed. Eng. Online* 2011, 10, 1–23.

32. Linguraru, M.G.; Richbourg, W.J.; Liu, J.; Watt, J.M.; Pamulapati, V.; Wang, S.; Summers, R.M. Tumor Burden Analysis on Computed Tomography by Automated Liver and Tumor Segmentation. *IEEE Trans. Med. Imaging* 2012, 31, 1965–1976.
33. Li, X.; Luo, S.; Li, J. Liver Segmentation from CT Image Using Fuzzy Clustering and Level Set. *J. Signal Inf. Process.* 2013, 4, 36–42.
34. Jena, B.; Krishna Nayak, G.; Saxena, S. Comprehensive Review of Abdominal Image Segmentation using Soft and Hard Computing Approaches. In *Proceedings of the International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Gunupur, India, 13–14 March 2020; pp. 1–5.
35. Platero, C.; Tobar, M.C. A multiatlas segmentation using graph cuts with applications to liver segmentation in CT scans. *Comput. Math. Methods Med.* 2014, 2014, 182909.
36. Mostafa, A.; Fouad, A.; Abd Elfattah, M.; Hassanien, A.E.; Hefny, H.; Zhu, S.Y.; Schaefer, G. CT liver segmentation using artificial bee colony optimisation. *Procedia Comput. Sci.* 2015, 60, 1622–1630.
37. Dou, Q.; Chen, H.; Jin, Y.; Yu, L.; Qin, J.; Heng, P.A. 3D deeply supervised network for automatic liver segmentation from CT volumes. In *Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Athens, Greece, 17–21 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 149–157.
38. Hiranman, A. Liver Segmentation Using 3D CT Scans. Ph.D. Thesis, University of Kwazulu-Natal, Durban, South Africa, 2018.
39. Wang, K.; Mamidipalli, A.; Retson, T.; Bahrami, N.; Hasenstab, K.; Blansit, K.; Bass, E.; Delgado, T.; Cunha, G.; Middleton, M.S.; et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. *Radiol. Artif. Intell.* 2019, 1, 180022.
40. Almotairi, S.; Kareem, G.; Aouf, M.; Almutairi, B.; Salem, M.A.M. Liver tumor segmentation in CT scans using modified SegNet. *Sensors* 2020, 20, 1516.
41. Soler, L.; Hostettler, A.; Agnus, V.; Charnoz, A.; Fasquel, J.; Moreau, J.; Osswald, A.; Bouhadjar, M.; Marescaux, J. 3D Image Reconstruction for Comparison of Algorithm Database: A Patient Specific Anatomical and Medical Image Database; Technical Report; IRCAD: Strasbourg, France, 2010.
42. Ayalew, Y.A.; Fante, K.A.; Mohammed, M.A. Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method. *BMC Biomed. Eng.* 2021, 3, 4.
43. LiTS Challenge Dataset. Available online: <https://competitions.codalab.org/competitions/17094> (accessed on 27 November 2023).
44. Scicluna, D. Automatic Segmentation of Healthy Liver in Abdominal Computed Tomography Scans. Master's Thesis, University of Malta, Msida, Malta, 2022.
45. Kavur, A.E.; Gezer, N.S.; Barış, M.; Aslan, S.; Conze, P.H.; Groza, V.; Pham, D.D.; Chatterjee, S.; Ernst, P.; Özkan, S.; et al. CHAOS Challenge-combined (CT-MR) healthy abdominal organ segmentation. *Med. Image Anal.* 2021, 69, 101950.
46. Ezzat, K.A.; Omran, L.N.; El Seddawy, A.I.B. Automatic liver segmentation in computed tomography scans using deep semantic segmentation. *Bull. Electr. Eng. Inform.* 2023, 12, 250–256.
47. Shao, J.; Luan, S.; Ding, Y.; Xue, X.; Zhu, B.; Wei, W. Attention Connect Network for Liver Tumor Segmentation from CT and MRI Images. *Technol. Cancer Res. Treat.* 2024, 23, 15330338231219366.
48. Maurício, J.; Domingues, I.; Bernardino, J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Appl. Sci.* 2023, 13, 5521.
49. Al-Saeed, Y.; Gab-Allah, W.; Elmogy, M. Hepatic tumors diagnosis system based on fuzzy c-means using computed tomography images. *Res. Sq. Prepr.* 2022.
50. MICCAI-Sliver07. Available online: <https://sliver07.grand-challenge.org/> (accessed on 27 November 2023).
51. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 2012, 25.
52. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* 2015, arXiv:1505.04597.

Retrieved from <https://encyclopedia.pub/entry/history/show/126786>

Appendix G– Biomedical Engineering Conference

Poster presentation at the *Biomedical Engineering Conference 2024 (JEB'24)* held at ISEC (**Application of Artificial Intelligence in Medical Imaging**):



Application of Artificial Intelligence in Medical Imaging

Stephanie Batista ¹, Miguel Couceiro ^{1,2,3}, Ricardo Filipe ⁴, Paulo Rachinhas ⁵, Jorge Isidoro ⁵ and Inês Domingues ^{1,6}

¹ Polytechnic Institute of Coimbra, Coimbra Institute of Engineering, Rua Pedro Nunes - Quinta da Nora, 3030-199 Coimbra, Portugal; ² Institute of Applied Research (I2A), Portugal; ³ Laboratory for High Performance Computing (LoCED), Portugal; ⁴ Altice Labs, S.A., Portugal; ⁵ Coimbra Hospital and University Center, Portugal; ⁶ IPO Porto Research Center (CHPOP): Medical Physics, Radiobiology & Radiation Protection Research Group, Portugal

KEYWORDS: Artificial Intelligence, Medical Imaging, Diagnostics and Treatment.

ABSTRACT: Artificial intelligence is transforming healthcare by improving diagnostic and treatment approaches. The integration of artificial intelligence into medical devices has enhanced the performance of diagnostic tools and therapeutic procedures, making care more patient-centric and minimally invasive. Particularly in medical imaging, artificial intelligence analyzes large volumes of medical data to detect patterns and anomalies more accurately than traditional methods. This capability not only speeds up diagnostics but also boosts the predictive power of imaging systems, enabling early detection and tailored interventions. This study investigates the transformative impact of artificial intelligence in medical imaging, focusing on technological advances, integration challenges, and the importance of standardization, data privacy, and regulatory compliance.

1 INTRODUCTION

Artificial Intelligence (AI) consists of a domain of science intended to provide computers with the ability to perform functions such as logic, reasoning, organization, learning and conceptualization [1]. It comprises methodologies such as machine learning (ML), deep learning (DL), and computer vision (CV), whose relationship is presented in Fig. 1. These methodologies allow machines to perform tasks which usually require human intelligence, involving the ability to reason, discover meaning, generalize, or learn from previous experiences. In the context of healthcare, AI concerns the use of these advanced algorithms and software to emulate human cognition in the analysis, interpretation and understanding of complex medical data [2].

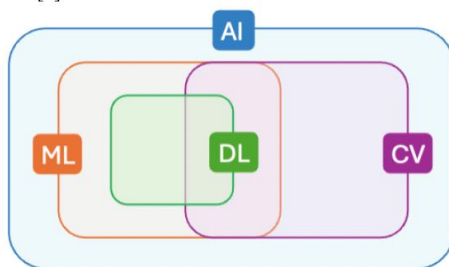


Fig. 1 Relationship between AI, ML, DL and CV

Applying AI in healthcare involves many fronts, redefining the paradigms of patient care through more accurate diagnosis, optimized treatment guidelines and better patient management in various domains. On a specific front, AI technologies are increasingly integrated into medical imaging, where they interpret large sets of health data, identifying patterns and predicting health outcomes with a level of precision and speed (Fig. 2). This application not only aids in the detection and characterization of diseases, but also in monitoring progression and response to treatment, thus optimizing patient outcomes.

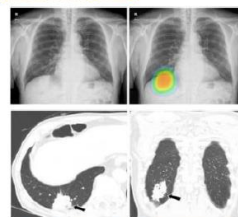


Fig. 2 Example of application of AI algorithms in medical imaging [3]

This study briefly explores how AI is transforming medical imaging with advanced diagnostic algorithms, some applications and its main challenges in Section 2 and Section 3 presents the main conclusions.

2 ARTIFICIAL INTELLIGENCE APPLICATION IN MEDICAL DIAGNOSTIC AND TREATMENT

Medical imaging modalities, such as radiography, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound, provide crucial information about the internal structures of the body, but interpreting these images is complex and time-consuming. AI attends as a valuable assistant for healthcare professionals, helping to analyze medical images to identify subtle patterns and anomalies. Using AI algorithms can accelerate the diagnosis process, leading to faster treatments and enhanced patient outcomes. While AI supports the diagnostic and treatment processes, healthcare professionals ultimately make the final decisions based on the insights provided by AI [4].

The number of publications using AI techniques to analyze static medical images has risen from hundreds to thousands in recent years. This growth is driven by the pattern-recognition demands of diagnostic tasks and the availability of organized images. Key applications in medical imaging include denoising, dose reduction, automatic segmentation, case triage, and image restoration.

The development of AI models in this field typically involves five main phases:

- **Data collection:**

Creation and characterization of the dataset to be used, in this case, consisting of medical images.

- **Data annotation/labelling:**

Creating a ground truth dataset involves a specialized healthcare professional manually identifying and recognizing study objects. This serves as a reference to train and validate the performance of models.

- **Pre-processing, dataset division, and data augmentation:**

For model evaluation, the dataset is commonly divided into two parts: one for training and one for testing. The training set may undergo data augmentation, a process of generating new data from existing images (through modifications like rotation, reflection, zoom, and translation) to provide more inputs for the algorithm. This assists the algorithm learn to recognize patterns and improve performance.

- **Algorithm training:**

Adapting the model training it on a dataset to adjust parameters, such as the weights of Neural Network connections, for intended tasks.

- **Algorithm testing:**

Using the testing dataset to provide an unbiased evaluation of a final model fit on the training dataset.

Some applications of AI in medical imaging within the hospital sector are present below in Tab. 1:

Tab. 1 Applications of AI in medical imaging.

Application	Description
Stroke Identification	CT and MRI scans to identify strokes for immediate treatment, detecting critical signs of acute strokes.
Heart Disease Prediction	AI-enhanced cardiac imaging predicts heart diseases early by analyzing cardiac images, helping preempt clinical manifestations.
Bone Fracture Detection and Classification	AI accurately detects and classifies bone fractures from a radiography, identifying subtle fractures that might be overlooked, enhancing orthopedic diagnostics.

Regulatory oversight and AI management systems are essential components in ensuring the safe and effective deployment of AI [5]. An interesting point to note is that since 1995, nearly 700 AI and ML-enabled medical devices have been approved by the FDA, reflecting a growing trust and reliance on AI technologies in healthcare industry [6].

3 CONCLUSIONS

The integration of AI with medical imaging promises to revolutionize healthcare by enhancing diagnostic accuracy and improving service accessibility. To fully leverage the potential of AI, it is crucial to address challenges like data bias, transparency, and ongoing regulatory changes. Collaborative efforts between technologists, healthcare providers, and regulators are essential to ensure ethical practices and patient safety. Addressing data privacy and maintaining strict regulatory oversight are also vital. By overcoming these challenges, AI can significantly advance medical imaging and patient care.

REFERENCES

- [1] J. Andreu-Perez, F. Deligianni, D. Ravi, and G.-Z. Yang, "Artificial intelligence and robotics", arXiv:1803.10813, 2018
- [2] K. Borys, Y. Schmitt, et al., "Explainable AI in medical imaging: An overview for clinical practitioners – Saliency-based XAI approaches", *European journal of radiology*, 2023, vol. 162.
- [3] Radiological Society of North America. Chest X-rays obtained as part of a health checkup in a 71-year-old male patient. Retrieved March 26, 2024 from <https://www.eurekalert.org/multimedia/989179>
- [4] A. Esteve A, Chou K, Yeung S, et al. "Deep learning-enabled medical computer vision". *NPJ Digit Med*. 2021, vol 8;4(1):5
- [5] E. Jillson, "Aiming for truth, fairness, and equity in your company's use of AI", FTC, 2021. Retrieved April 7, 2024 from <https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai>
- [6] U.S Food and Drug Administration (FDA). Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices, 2023. Retrieved April 7, 2024 from <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>

Application of Artificial Intelligence in Medical Imaging

Stephanie Batista¹, Miguel Couceiro^{1,2,3,4}, Ricardo Filipe⁵, Paulo Rachinhas⁶,
Jorge Isidoro⁶ and Inês Domingues^{1,7}

¹ Polytechnic Institute of Coimbra, Coimbra Institute of Engineering, Rua Pedro Nunes - Quinta da Nora, 3030-199 Coimbra, Portugal;

² Institute of Applied Research (IZA), Portugal; ³ Laboratory for High Performance Computing (LoCED), Portugal;

⁴ Laboratory of Instrumentation and Experimental Particle Physics (IIP-Coimbra), Rua Larga da Universidade de Coimbra, 3004-516 Coimbra, Portugal;

⁵ AITice Labs, S.A., Portugal; ⁶ Coimbra Hospital and University Center, Portugal; ⁷ IPO Porto Research Center (CHPOP): Medical Physics, Radiobiology & Radiation Protection Research Group, Portugal

Abstract

Artificial intelligence is transforming healthcare, optimizing diagnostic assistance systems and treatment strategies to be more precise and patient-centered. Especially in medical imaging, artificial intelligence analyzes large volumes of medical data to identify patterns and anomalies, enhancing diagnosis and predictive capabilities for individualized interventions in an increasingly less invasive approach wherever possible. This study briefly explores the impact of artificial intelligence on medical imaging and how it can be applied.

Keywords: Artificial Intelligence, Medical Imaging, Diagnostics and Treatment.

Introduction

Artificial Intelligence (AI) is a domain of science intended to provide computers with the ability to perform functions such as:



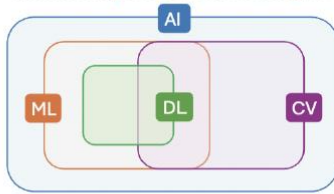
AI Methodologies Applied in Medical Imaging

Machine Learning (ML): teaches computers to learn and make decisions from data.

Deep Learning (DL): uses neural networks with many layers to analyze various levels of data features.

Computer Vision (CV): trains computers to interpret and understand visual data (digital images and videos).

Relationship between Methodologies



AI in **healthcare** enhances patient care by improving diagnostics, optimizing treatments, and managing health outcomes through advanced medical imaging and data analysis.

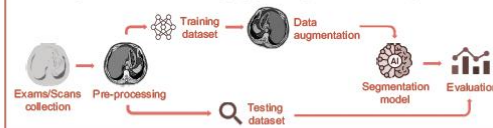
Application in Medical Diagnostic and Treatment

The integration of **AI in medical imaging** has grown extensively, propelled by the need for effective pattern recognition and the access to structured datasets.

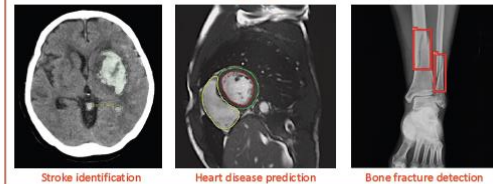
This technology enhances diagnostics by supporting functions like denoising, segmentation, and image restoration, improving accuracy and efficiency.

Development of AI Algorithms

AI algorithms for medical imaging go through several critical phases:



Medical Imaging Use Cases



Meeting Healthcare's High Safety Regulations and Standards

AI deployment in healthcare is strictly regulated to ensure the **safety and efficacy** for patients, providing trust and reliance on AI healthcare technologies.

Since 1995, the **FDA** has approved **≈ 700** AI and ML medical devices

Conclusions

AI improves diagnostic accuracy

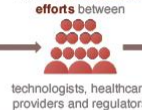
Key challenges

- Databases
- Ensure transparency
- Regulatory compliance

Prioritize the protection



Promote collaborative efforts between



The integration of **AI and medical imaging** promises to revolutionize healthcare

References

1. J. Andreu-Perez, F. Deligianni, D. Ravi, and G.-Z. Yang, "Artificial intelligence and robotics", arXiv:1803.10813, 2018
2. K. Borys, Y. Schmitt, et al., "Explainable AI in medical imaging: An overview for clinical practitioners – Saliency-based XAI approaches", European journal of radiology, 2023, vol 162.
3. Radiological Society of North America. Chest X-rays obtained as part of a health checkup in a 71-year-old male patient. Retrieved March 26, 2024 from <https://www.eurekalent.org/multimedia/589179>
4. A. Esteve A, Chou K, Yeung S, et al. "Deep learning-enabled medical computer vision". NPJ Digit Med. 2021, vol 8;4(1):5
5. E. Jillson, "Aiming for truth, fairness, and equity in your company's use of AI", FTC, 2021. Retrieved April 7, 2024 from <https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai>
6. U.S Food and Drug Administration (FDA). Artificial Intelligence and Machine Learning (AIML)-Enabled Medical Devices, 2023. Retrieved April 7, 2024 from <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>

Appendix H – International Conference on Mathematical Analysis and Applications in Science and Engineering

Poster presentation at the *International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMA²SC'24)* [53] (**Performance Evaluation Metrics for Automatic Liver Segmentation on Medical Imaging Scans**):

International Conference on Mathematical Analysis and Applications in Science and Engineering
ICMA²SC'24

ISEP Porto-Portugal, June 20 - 22, 2024

Performance Evaluation Metrics for Automatic Liver Segmentation on Medical Imaging Scans

Stephanie Batista^{1*} Miguel Couceiro^{1,2,3,4} Ricardo Filipe⁵
Paulo Rachinhas⁶ Jorge Isidoro⁶ Inês Domingues^{1,7}

¹ Polytechnic Institute of Coimbra, Coimbra Institute of Engineering,
Rua Pedro Nunes - Quinta da Nora, 3030-199 Coimbra, Portugal

² Institute of Applied Research (i2A), Portugal

³ Laboratory for High Performance Computing (LaCED), Portugal

⁴ Laboratory of Instrumentation and Experimental Particle Physics (LIP-Coimbra),
Rua Larga da Universidade de Coimbra, 3004-516 Coimbra, Portugal

⁵ Altice Labs, S.A, Portugal

⁶ Coimbra Hospital and University Center, Portugal

⁷ IPO Porto Research Center (CI-IPOP): Medical Physics,
Radiobiology & Radiation Protection Research Group, Portugal

^{1*}stephanieabn@gmail.com

Abstract: Liver segmentation in medical imaging scans is fundamental for the diagnosis and treatment planning of liver-related pathologies. It aids in identifying and characterizing diseases, providing accurate recognition of affected organs and tissues. Despite its significance, achieving precise segmentation is a major challenge due to the presence of artifacts and the different densities of soft tissues and neighboring organs. Given the inherent difficulty associated with the subjectivity regarding the validation and acceptance of manual segmentation by clinicians, this work emphasizes the importance of quantitative metrics to complement the subjective assessment for automatic liver segmentation outcome.

keywords: Medical Imaging; Liver; Automatic Liver Segmentation; Evaluation Metrics.

1 Introduction

The evolution of medical imaging represents a hallmark of technological advancement in healthcare, revolutionizing diagnostic practices from invasive to non-invasive methods, thus enhancing patient comfort and healthcare standards [1]. Beyond diagnosis, medical imaging plays a crucial role in treatment planning and follow-up care, signifying a dedication to improving patient well-being across the continuum of medical intervention. Stemming from the pioneering efforts of figures like Marie Curie [2], medical imaging modalities such as CT, PET, SPECT, and MRI continue to expand treatment options and elevate the standard of care.

One significant advancement in this field is the progress made in organ segmentation, particularly with liver segmentation [3]. This is crucial for monitoring the health of the digestive system and treat pathologies related to liver. Despite obstacles such as artifacts and the close proximity of tissues, automatic techniques have seen substantial improvements [4, 5]. However, consistent evaluation metrics for these techniques remain lacking, hindering comparative analysis of their evolution and effectiveness [6].

2 Segmentation Evaluation Metrics

There are two main types of metrics, region-based and contour-based metrics. **Region-based metrics** assess the correctness of the area covered by segmentation, regardless of border precision. These include Accuracy, Dice Similarity Coefficient (DSC), Precision, and Volume Overlap Error (VOE). **Contour-based metrics** measure the correctness of boundary delineations. These include Average Symmetric Surface Distance (ASSD) and Mean Symmetric Surface Distance (MSSD). Figures 1 and 2 illustrate the calculation of these metrics, with True Positives (TP) in green, True negatives (TN) in blue, False negatives (FN) in red, False positives (FP) in orange.



Figure 1: Illustration of Image Segmentation results: (a) Original image; (b) Ground truth segmentation; (c) Automatic segmentation; (d) Classification outcomes.

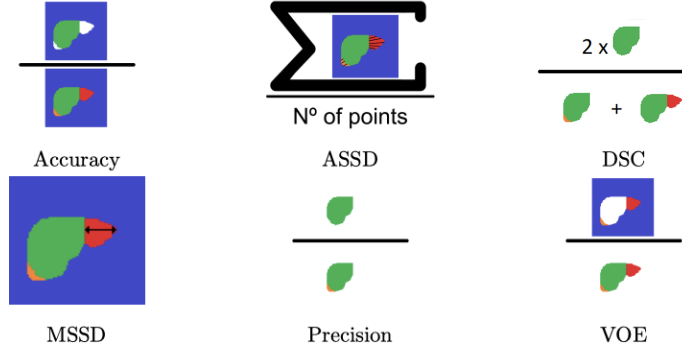


Figure 2: Illustration of metrics calculation.

Research on liver segmentation in medical imaging demonstrates promise, yet the lack of consistent evaluation metrics is often overlooked [7, 8]. A literature review conducted previously [6] selected studies dated between 2014 and 2024, focusing on fully automatic liver segmentation techniques. As a result, the most commonly used metrics in these studies were the ones shown in Figure 3.

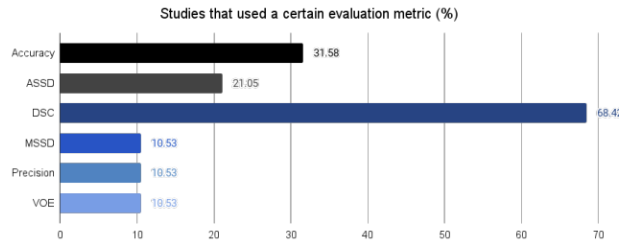


Figure 3: Distribution of the metrics used in the selected studies.

2D Liver Segmentation in Abdominal Computed Tomography Scans

Table 1 presents an overview of the metrics, along with their main advantages and shortcomings.

Table 1: Summary of the main evaluation metrics used by automatic liver segmentation studies.

Metric	Definition	Main Advantages	Main Disadvantages
Accuracy	Measures the proportion of correctly identified voxels (both liver and non-liver), including TPs and true negatives, to the total number of voxels.	-Provides a comprehensive overview of the segmentation performance. -Easy to calculate and interpret.	-May be misleading in datasets with a large imbalance between liver and non-liver voxels. -Does not specifically focus on liver tissue accuracy.
ASSD	Calculates the average distance between corresponding points on the surfaces of the segmented liver region and the ground truth, in both directions.	-Offers a detailed assessment of boundary accuracy. -Highlights discrepancies in the segmentation's surface alignment.	-More complex to compute and interpret compared to overlap-based metrics. -Can be influenced by outliers or extreme values.
DSC	Measures the overlap between the segmented liver region and the ground truth, calculated as the size of the intersection divided by the summed size of the two regions.	-Provides a clear, straightforward measure. -Easy to interpret and widely used in medical imaging.	-Might not fully capture small discrepancies in boundary delineation. -Can be overly optimistic in cases of large overlap but poor boundary alignment.
MSSD	Determines the maximum distance between any point on the segmented liver surface and the nearest point on the ground truth surface, and vice versa.	-Identifies the worst-case scenario in segmentation discrepancies. -Useful for assessing the reliability of segmentation in critical areas.	-Focused on the maximum error, potentially overlooking overall segmentation quality. -Can be overly punitive for generally accurate segmentations with minor outliers.
Precision	Assesses the proportion of voxels correctly identified as liver tissue out of all voxels classified as liver by the algorithm.	-Highlights the algorithm's ability to avoid over-segmentation.	-Does not account for missed liver tissue (FNs). -It may only identify very clear cases of liver boundaries and ignore the more ambiguous cases.
VOE	Quantifies the volumetric discrepancy between the segmented liver region and the ground truth as the percentage of non-overlapping volume to union volume.	-Directly related to clinical relevance, reflecting the actual volume error. -Intuitive and easy to understand.	-Can be sensitive to the size of the liver, with larger livers potentially showing more significant absolute errors.

The results presented indicate that DSC, Accuracy, and ASSD stand out as the most frequently employed metrics in liver segmentation evaluations. Despite some studies validating the effectiveness of these metrics in tackling common challenges such as under-segmentation and over-segmentation, there remains an evident need for the adoption of standardized metrics in the field. We advocate for the use of at least one metric within each type of evaluation metric (region and contour). The introduction of such standards would greatly advance segmentation studies by underscoring key aspects, enabling straightforward comparisons across various methods, and significantly improving the reliability and effectiveness of automated liver segmentation in medical imaging.

3 Conclusions and Future Work

Non-invasive imaging techniques have vast potential to transform patient care, emphasizing the need for consistent metrics in automatic liver segmentation. By developing unified, quantitative standards, the accuracy, reliability, and comparability of segmentation methods can be significantly enhanced. This will not only aid clinicians but also steer advancements toward more effective and user-friendly solutions.

Standardizing evaluation metrics is essential for progressing liver segmentation technology. Clear metrics will enable better comparisons, foster innovation, and improve the development and validation of new methods. This will directly benefit clinical practices by improving diagnostics and treatment planning.

It is crucial to align technological developments with clinical needs to incorporate automatic liver segmentation into routine medical care. This alignment will improve diagnoses and help customize treatments for liver diseases, enhancing patient outcomes.

References

- [1] Friebe, M; Illanes, A, Advancements in Medical Imaging and Image-Guided Procedures: A Potential-Or Rather Likely-Paradigm Shift in Diagnosis and Therapy: Understand Disruption and Take Advantage of It!, *Appl. Sci.*, **13** (2023), 9218.
- [2] American Institute of Physics, Marie Curie - War Duty (1914-1919), n.d. Retrieved March 24, 2024 from <https://history.aip.org/exhibits/curie/war1.htm>
- [3] Torres, A, Segmentação de imagens médicas visando a construção de modelos médicos, Instituto Politécnico de Bragança, Escola Superior de Tecnologia e Gestão, 2012.
- [4] Trefts, E; Gannon, M; Wasserman, D, Primer - The liver, *Current Biology Magazine*, 2017.
- [5] Francque, S, The liver and the cardiovascular system: Two of a kind?, *J Am Heart Assoc*, **10**(8) (2021).
- [6] Niño, SB; Bernardino, J; Domingues, I, Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective, *Sensors*, **24** (2024), 1752.
- [7] Crum, WR; Camara, O; Hill, DL, Generalized overlap measures for evaluation and validation in medical image analysis, *IEEE Trans Med Imaging*, **25**(11) (2006), 1451-1461. PMID: 17117774.
- [8] Bilic, P; Christ, P; Li, HB; Vorontsov, E; Ben-Cohen, A; Kaissis, G; Szeskin, A; Jacobs, The Liver Tumor Segmentation Benchmark (LiTS), *Medical Image Analysis*, **84** (2023), 102680, ISSN 1361-8415.

2D Liver Segmentation in Abdominal Computed Tomography Scans

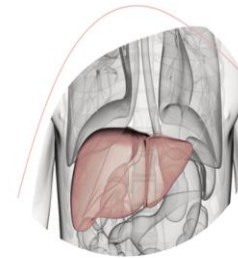


PERFORMANCE EVALUATION METRICS FOR AUTOMATIC LIVER SEGMENTATION ON MEDICAL IMAGING SCANS

STEPHANIE BATISTA

POLYTECHNIC INSTITUTE OF COIMBRA, COIMBRA INSTITUTE OF ENGINEERING
stephanieabn@gmail.com

Table of Contents



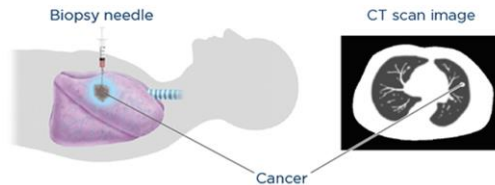
01 Introduction

02 Segmentation Evaluation Metrics

03 Conclusions

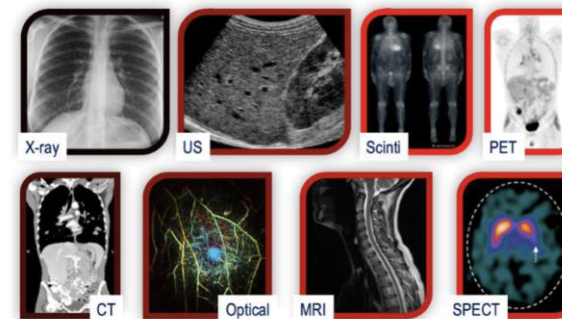
01 Introduction

Medical imaging keeps evolving for healthcare, advancing from invasive to non-invasive diagnostics and improving patient comfort and care.



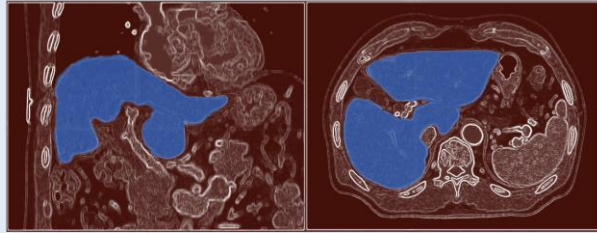
Beyond diagnosis, medical imaging is crucial for treatment planning and follow-up care, improving patient well-being throughout their medical intervention.

Medical Imaging Modalities



One significant advancement in this field is the progress made in organs segmentation.

Liver segmentation is crucial for monitoring digestive health and treating liver-related diseases. Despite challenges like artifacts and nearby tissues, automatic methods have significantly improved.



Despite this, there has been a lack of **consistent evaluation metrics** for these methods, which makes comparative assessment of their evolution and effectiveness more difficult.

5

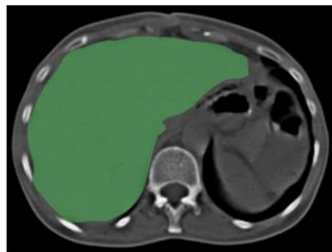
02 Segmentation Evaluation Metrics

Segmentation evaluation metrics are crucial for developing effective liver segmentation tools. They assess the reliability and consistency of segmentation algorithms, enabling standardized comparisons between different methods.

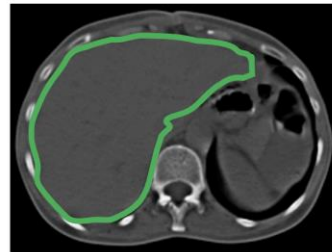


6

Classification of Segmentation Evaluation Metrics



Region-based metrics



Contour-based metrics

7

Region-based Metrics

Assess the correctness of the area covered by segmentation, regardless of contour precision.

These include:

- Accuracy
- Precision
- Dice Similarity Coefficient
- Area/Volume Overlap Error

8

2D Liver Segmentation in Abdominal Computed Tomography Scans

Precision

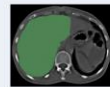
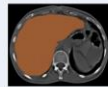
Assesses the proportion of pixels or voxels correctly identified as liver tissue out of all the pixels/voxels classified as liver by the algorithm.

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



Main disadvantage:

- Ignores false negatives (FN)



GT
Ground Truth
AS
Algorithm segmentation

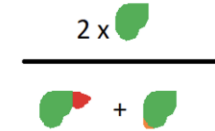
10

Dice Similarity Coefficient

(Sørensen-Dice coefficient/index or Dice coefficient)

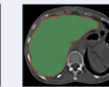
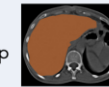
Measures the similarity between the segmented liver region obtained from an algorithm and the ground truth.

$$\text{DSC} = \frac{2 * |S \cap G|}{|S| + |G|}$$



Main disadvantage:

- Overly optimistic for large overlap but poor contour delineation



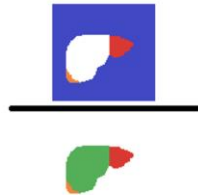
GT
Ground Truth
AS
Algorithm segmentation

11

Area/Volume Overlap Error

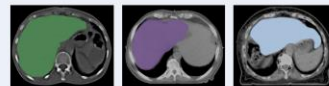
Quantifies the discrepancy in the area/volumetric overlap between the segmented liver region produced by an algorithm and the ground truth.

$$\text{AOE/VOE} = \left(1 - \frac{|S \cap G|}{|S \cup G|}\right) * 100\%$$

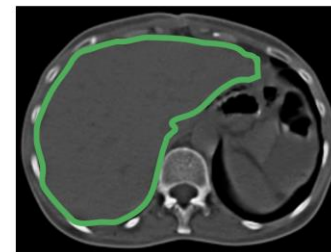


Main disadvantage:

- Sensitive to liver size:
Larger livers show more errors



12



Contour-based Metrics

Measure the correctness of boundary delineations.

These include:

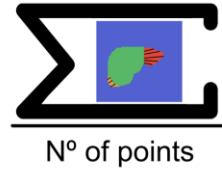
- Average Symmetric Surface Distance
- Maximum Symmetric Surface Distance

13

Average Symmetric Surface Distance

Assesses how well a segmentation algorithm can delineate the boundaries of the liver from the surrounding tissues.

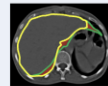
$$ASSD = \frac{1}{2} \left(\frac{\sum_{i=1}^N d(S_i, G)}{N} + \frac{\sum_{j=1}^M d(G_j, S)}{M} \right)$$



Nº of points

Main disadvantage:

- Influenced by outliers



GT
Ground Truth
AS
Algorithm segmentation
Difference between GT and AS

14

Maximum Symmetric Surface Distance

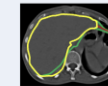
Measures the maximum distance between the surfaces of the segmented liver region and the ground truth.

$$MSSD = \max \left(\max_{i \in S} \min_{j \in G} d(S_i, G_j), \max_{j \in G} \min_{i \in S} d(G_j, S_i) \right)$$



Main disadvantages:

- Focuses on maximum error
- Overly punitive

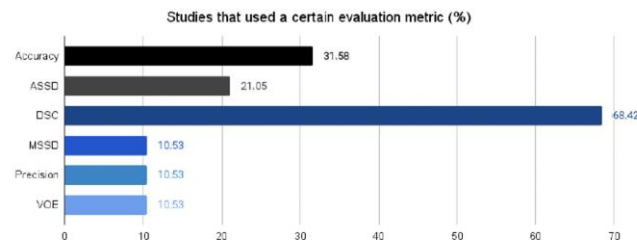


GT
Ground Truth
AS
Algorithm segmentation
Maximum distance between GT and AS

15

Distribution of the Metrics Used

A literature review with 19 studies dated between 2014 and 2024 about **automatic liver segmentation methods** showed the following distribution of use of the evaluation metrics under analysis:



16

03 Conclusions

The most used evaluation metrics in the studies:

DSC

Accuracy

ASSD

These are effective metrics to address

Undersegmentation and Oversegmentation



17

2D Liver Segmentation in Abdominal Computed Tomography Scans

Need for standardized metrics

At least one metric from each type should be used

DSC

ASSD

18



By developing unified, quantitative standards, the efficiency, reliability, and comparability of segmentation methods can be significantly enhanced.



19

References

- ▶ Friebe, M; Illanes, A, Advancements in Medical Imaging and Image-Guided Procedures: A Potential-Or Rather Likely-Paradigm Shift in Diagnosis and Therapy: Understand Disruption and Take Advantage of It!, Appl. Sci., 13 (2023), 9218
- ▶ American Institute of Physics, Marie Curie - War Duty (1914-1919), n.d. Retrieved March 24, 2024 from <https://history.aip.org/exhibits/curie/war1.htm>
- ▶ Torres, A, Segmentação de imagens médicas visando a construção de modelos médicos, Instituto Politécnico de Bragança, Escola Superior de Tecnologia e Gestão, 2012
- ▶ Trefts, E; Gannon, M; Wasserman, D, Primer - The liver, Current Biology Magazine, 2017
- ▶ Francque, S, The liver and the cardiovascular system: Two of a kind?, J Am Heart Assoc, 10(8) (2021)
- ▶ Niño, SB; Bernardino, J; Domingues, I, Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective, Sensors, 24 (2024), 1752
- ▶ Crum, WR; Camara, O; Hill, DL, Generalized overlap measures for evaluation and validation in medical image analysis, IEEE Trans Med Imaging, 25(11) (2006), 1451-1461. PMID: 17117774
- ▶ Bilic, P; Christ, P; Li, HB; Vorontsov, E; Ben-Cohen, A; Kaissis, G; Szeskin, A; Jacobs, The Liver Tumor Segmentation Benchmark (LiTS), Medical Image Analysis, 84 (2023), 102680, ISSN 1361-8415

20

THANKS

Do you have any questions?

Contact me!

 stephanieabn@gmail.com

21

Appendix I – Springer Proceedings Journal

Publication of a paper (**Comparative Analysis of Quantitative Assessment Metrics for Liver Segmentation**) in *Springer Proceedings* journal:

Chapter 1

Comparative Analysis of Quantitative Assessment Metrics for Liver Segmentation

Abstract Liver segmentation in medical imaging scans is fundamental for the diagnosis and treatment planning of liver-related pathologies. It aids in identifying and characterizing diseases, providing accurate recognition of affected organs and tissues. Despite its significance, achieving precise segmentation is a major challenge due to the presence of artifacts and the different densities of soft tissues and neighboring organs. Given the inherent difficulty associated with the subjectivity regarding the validation and acceptance of manual segmentation by clinicians, this work emphasizes the importance of quantitative metrics to complement the subjective assessment for automatic liver segmentation outcome.

keywords: Medical Imaging; Liver; Automatic Liver Segmentation; Evaluation Metrics.

1.1 Introduction

The evolution of medical imaging represents a hallmark of technological advancement in healthcare, revolutionizing diagnostic practices from invasive to non-invasive methods, thus enhancing patient comfort and healthcare standards [18]. Beyond diagnosis, medical imaging plays a crucial role in treatment planning and follow-up care, signifying a dedication to improving patient well-being across the continuum of medical intervention. Stemming from the pioneering efforts of figures like Marie Curie [3], medical imaging modalities such as CT, PET, SPECT, and MRI continue to elevate the standard of care [27]. These techniques have become an integral part of the diagnosis, treatment planning and monitoring the progress of oncological diseases [30, 39], providing detailed cross-sectional images for accurate visualization of internal structures [60, 54].

Physicians require tools that allow them to localize organs and tissues with higher precision and rapidity, while identifying and characterizing pathologies for accurate diagnosis. As one of the most important organs in the digestive system, the liver performs critical functions such as breaking down nutrients, producing bile and

eliminating toxic substances. However, liver-related diseases, particularly oncological diseases, pose health risks, and liver cancer is a leading cause of cancer-related mortality worldwide [17, 52, 49].

With the advancement of technology and artificial intelligence (AI) in medicine, there is a growing need to optimize the identification of oncological diseases [14]. One advancement in this field is the progress made in organ segmentation, particularly with liver segmentation [51, 29, 7, 38, 9]. This is crucial for monitoring the health of the digestive system and treat pathologies related to liver, requiring accurate identification and delimitation of the liver region for treatment planning and progress monitoring, as well as for early detection of liver lesions and metastases to other organs [62, 11, 12]. However, accurate liver segmentation on CT scans is challenging due to factors such as artefacts, varying soft tissue densities, the complexity of adjacent organ proximity and dynamic anatomical modifications, such as tumor contraction during treatment, there is a compelling need for reliable and fast segmentation to facilitate adaptive radiotherapy [4, 24, 21, 8, 44]. Despite these obstacles, automatic techniques have shown substantial improvements [52, 17]. However, the evolution of these diverse approaches has underscored the critical need for a rigorous and standardized evaluation framework. Accurate performance metrics are not just a tool for assessing the effectiveness of a given method, they are essential for ensuring the reproducibility and comparability of results across different studies and applications. In an era where Machine Learning (ML) models and AI-driven techniques are rapidly advancing, the establishment of robust evaluation protocols is necessary [35]. These metrics serve as the foundation for objectively validating the improvements claimed by new methods, facilitating a transparent and consistent comparison against existing benchmarks.

As the field of liver segmentation and, more broadly, medical image analysis, continues to grow, the ability to critically assess the performance of various models becomes increasingly crucial. Standardized metrics not only drive the adoption of best practices but also contribute to the broader scientific discourse by enabling researchers and practitioners to communicate their findings effectively. This ensures that innovations are not only theoretically sound but also practically viable, leading to meaningful advancements in clinical applications. Therefore, the careful selection and application of evaluation metrics should be viewed as a cornerstone of method development, guiding the trajectory of research and ultimately influencing clinical outcomes.

This chapter expands upon a presentation given at the International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMASC) 2024, titled "Performance Evaluation Metrics for Automatic Liver Segmentation on Medical Imaging Scans," [6] which provided a brief overview of performance evaluation metric types for segmentation models, focusing primarily on introducing the metrics and comparing their usage across selected studies. In contrast, this chapter offers a more comprehensive and in-depth examination of the subject, analyzing the scope and relevance of rigorous and efficient validation of segmentation models, providing a detailed assessment of the advantages and disadvantages of using each metric.

This thorough research aims to assist researchers in understanding the benefits of choosing specific metrics and supports the prioritization of metrics based on their definitions. While all the metrics presented in this document are meaningful, certain metrics are more impactful in assessing performance than others, with some metrics tending towards optimism and others towards pessimism. Moreover, understanding the implications of the value of each metric is essential to correctly interpreting the results of segmentation models. This chapter emphasizes the importance of using standardized evaluation metrics, allowing researchers to effectively compare their results with previous studies, while also underlining the limitations of comparing different metrics that evaluate different aspects of performance, emphasizing the need for a standardized approach in this field.

Besides introduction, this paper is structured in four sections. Section 1.2 introduces the fundamental concepts related to the liver, namely its importance and of liver segmentation applications. Section 1.3 details the most commonly used performance evaluation metrics from the literature review. Section 1.4 describes the analysis of liver segmentation evaluation metrics in recent medical imaging studies. Section 1.5 discusses the findings from this research. Section 1.6 presents the conclusions and main suggestions for the performance evaluation of automatic liver segmentation on medical imaging scans.

1.2 Liver

Given that the liver is the organ studied through this paper, this section provides relevant information to better understand its main functions and relevance for study.

1.2.1 Definition

The liver is a vital organ located in the upper right quadrant of the abdomen, beneath the diaphragm. It is the largest internal organ and gland in the human body, weighing approximately 1.5 kilograms in an average adult. Structurally, the liver is divided into two main lobes: the right lobe, which is larger, and the left lobe. This organ is composed of specialized cells called hepatocytes, which perform a myriad of essential functions that are crucial for maintaining homeostasis and overall health [23].

1.2.2 Importance

The liver plays a critical role in several processes, making it essential for human survival. One of its primary functions is detoxification. The liver processes toxins

and drugs, rendering them harmless or facilitating their excretion from the body. Additionally, it is responsible for the production of bile, which is crucial for the digestion and absorption of fats and fat-soluble vitamins in the small intestine [33, 34].

The liver also performs an important function in metabolism. It regulates blood sugar levels by storing and releasing glucose as needed and is involved in the metabolism of proteins and lipids. Furthermore, the liver synthesizes essential proteins such as albumin and clotting factors, which are vital for maintaining blood volume and proper clotting function. The organ also stores vitamins and minerals, including iron and vitamin A, and converts ammonia, a toxic byproduct of protein metabolism, into urea for excretion.

1.2.3 Significance of Liver Studies

Studying the liver in detail is extremely important due to its critical functions and the impact of liver diseases. Liver diseases, including cirrhosis, hepatitis, and liver cancer, are among the leading causes of death worldwide, being estimated that about 2500 people die each year due to liver diseases [26, 53].

As Portugal is considered to be the 10th country in the world with the highest alcohol consumption and alcohol is one of the main causes of liver diseases, the Portuguese Society of Gastroenterology estimates a five-year mortality rate of 95% for people with liver diseases in Portugal, and liver cancer is the 5th leading cause of death in Portugal according to the Directorate General of Health (DGS) [13], mostly associated with men, approximately 75% [57].

These diseases not only affect millions of individuals, but also represent a substantial burden on health systems.

Understanding the complex physiology and pathology of the liver is essential for developing effective treatments and preventive measures. Advances in liver research have the potential to improve diagnostic techniques, therapeutic strategies, and patient outcomes. Additionally, studying the liver can provide insights into related conditions, such as metabolic disorders, cardiovascular diseases, and obesity, given the organ's integral role in metabolism and detoxification.

1.2.4 Relevance of Liver Segmentation

Liver segmentation in medical imaging is paramount due to its critical role in oncological treatment planning and progress monitoring [33, 46, 28]. The complex vascular structure and vital functions of the liver needs precise imaging to accurately localize and characterize tumors. This precision enables tailored radiotherapy and surgical planning, ensuring effective treatment while preserving healthy liver tissue. Furthermore, follow-up segmentation allows for meticulous monitoring of treatment

response and early detection of disease progression, facilitating timely adjustments to therapeutic strategies. Despite its importance, liver segmentation in CT scans remains a challenging task [36, 50, 19, 31, 42].

1.3 Performance Evaluation Metrics

There are two main types of evaluation metrics of segmentation performance, region-based (Section 1.3.1) and contour-based metrics (Section 1.3.2). To better understand the performance evaluation metrics, it is relevant to first define some key concepts. These concepts form the foundation for comparing the algorithm's segmentation results with the ground truth (expert-annotated) data:

- **True Positives (TP)** are the pixels/voxels correctly identified as liver tissue by both the algorithm and the ground truth.
- **True Negatives (TN)** are the pixels/voxels correctly identified as non-liver tissue by both the algorithm and the ground truth.
- **False Positives (FP)** are the pixels/voxels incorrectly identified as liver tissue by the algorithm but not by the ground truth.
- **False Negatives (FN)** are the pixels/voxels incorrectly identified as non-liver tissue by the algorithm but identified as liver tissue by the ground truth.

To visualize these concepts, Figure 1.1 illustrates an original image alongside its corresponding ground truth and automatic segmentation. In this figure, TP are shown in green, TN in blue, FN in red, and FP in orange.



Fig. 1.1 Illustration of image segmentation results. (a) Original image. (b) Ground truth segmentation. (c) Automatic segmentation. (d) Classification outcomes

For 2D scans, these metrics are calculated based on the value of the pixels (picture elements). In the case of 3D scans, the metrics consider voxels (volume pixels). Voxels are analogous to pixels in 2D scans. Each voxel contains a value that represents some property (such as color, density, or intensity) in the 3D space.

1.3.1 Region-based Metrics

Region-based metrics assess the correctness of the area covered by segmentation, regardless of contour precision (Figure 1.2). These include Accuracy, Precision, Dice Similarity Coefficient (DSC), and Volume Overlap Error (VOE).

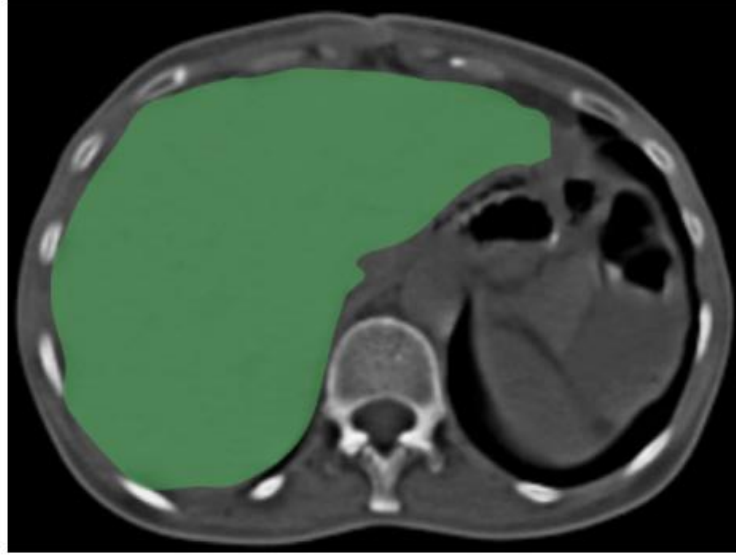


Fig. 1.2 Region-based illustration.

1.3.1.1 Accuracy

Accuracy is calculated as the ratio of the number of correctly identified pixels/voxels (both liver and non-liver) to the total number of pixels/voxels in the image or volume of interest. In mathematical terms, accuracy can be defined as presented in Equation 1.1 [59, 22].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1.1)$$

Figure 1.3 illustrates the calculation of the accuracy metric in order to translate the mathematical equation into a visual explanation.

Accuracy provides a general overview of the segmentation performance, combining both the TP rate (sensitivity) and the TN rate (specificity) into a single measure. A high accuracy value, close to 1 (or 100% when expressed as a percentage), indicates a high level of agreement between the segmentation of the algorithm and the ground truth, implying an effective segmentation performance. Conversely, a lower accu-

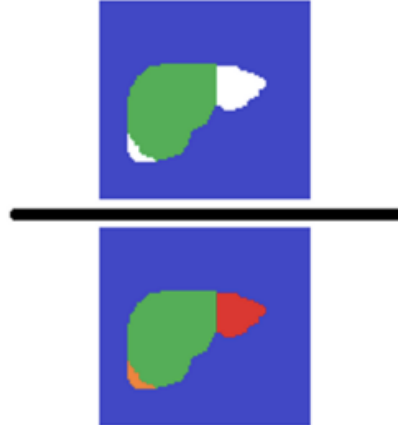


Fig. 1.3 Illustration of the calculation of accuracy, with TP in green, TN in blue, FN in red, and FP in orange

racy value suggests discrepancies between the output of the algorithm and the ground truth, highlighting potential areas for improvement in the segmentation process.

The accuracy value can be misleading due to potential imbalances in the data. This is particularly true in segmentation problems, where there are often many more pixels belonging to the background than to the region of interest. In such cases, FP and FN may be "diluted" compared to the much larger number of TN.

1.3.1.2 Precision

Precision is focused on the quality of the positive predictions made by the segmentation algorithm, meaning it assesses the proportion of pixels/voxels correctly identified as liver tissue out of all the pixels/voxels classified as liver by the algorithm [22]. Mathematically, the precision metric is defined as presented in Equation 1.2.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1.2)$$

Figure 1.4 illustrates the calculation of the precision metric in order to translate the mathematical equation into a visual explanation.

A high precision indicates that a large majority of the pixels/voxels labeled by the segmentation algorithm as part of the liver are indeed liver tissue according to the ground truth. This implies that the algorithm is effective in avoiding under-segmentation, where non-liver tissues are mistakenly included within the segmented region of the liver.



Fig. 1.4 Illustration of the calculation of precision, with TP in green, and FP in orange

1.3.1.3 Dice Similarity Coefficient

The DSC metric for assessing the performance of liver segmentation algorithms in CT scans, also known as the Sørensen-Dice coefficient/index or Dice coefficient, is a widely utilized evaluation metric in the field of medical imaging [22, 55]. The DSC equation is presented in (1.3).

$$\text{DSC} = \frac{2 \times |S \cap G|}{|S| + |G|} \quad (1.3)$$

where:

- $|S \cap G|$ represents the intersection, or the common pixels/voxels, between the segmented liver region S and the ground truth G . This intersection quantifies the number of pixels/voxels accurately identified as liver tissue by the segmentation algorithm.
- $|S|$ and $|G|$ denote the total number of pixels/voxels within the segmented liver region and the ground truth, respectively.

Figure 1.5 illustrates the calculation of the DSC metric in order to translate the mathematical equation into a visual explanation.



Fig. 1.5 Illustration of the calculation of DSC, with TP in green, FN in red, and FP in orange

The DSC ranges from 0 to 1, where 0 indicates no overlap between the segmented liver and the ground truth (indicating poor segmentation performance), and 1 represents perfect alignment or overlap (indicating ideal segmentation accuracy).

1.3.1.4 Volumetric Overlap Error

VOE is expressed as a percentage, representing the proportion of the total volume that does not overlap between the segmentation of the algorithm and the ground truth [58]. In 2D images, this is often referred to as the Area Overlap Error (AOE) or simply the overlap error, and it works on the same principle of comparing areas instead of volumes. It is calculated using Equation 1.4.

$$\text{VOE} = \left(1 - \frac{|S \cap G|}{|S \cup G|} \right) \times 100\% \quad (1.4)$$

where:

- $|S \cap G|$ denotes the intersection between the segmented liver region S and the ground truth G , as above.
- $|S \cup G|$ signifies the union area/volume of S and G , encompassing all the area/volume identified as liver tissue by either the segmentation algorithm or the ground truth.

Figure 1.6 illustrates the calculation of the VOE metric in order to translate the mathematical equation into a visual explanation.

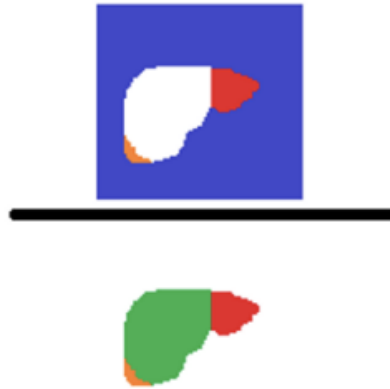


Fig. 1.6 Illustration of the calculation of VOE, with TP in green, TN in blue, FN in red, and FP in orange

A lower VOE indicates a higher degree of overlap between the segmented liver and the ground truth, reflecting a more accurate segmentation. Conversely, a higher

VOE points to a greater discrepancy between the segmented region and the ground truth, signaling potential inaccuracies in the segmentation algorithm.

1.3.2 Contour-based Metrics

Contour-based metrics measure the correctness of boundary delineations (Figure 1.7). These include Average Symmetric Surface Distance (ASSD) and Mean Symmetric Surface Distance (MSSD).

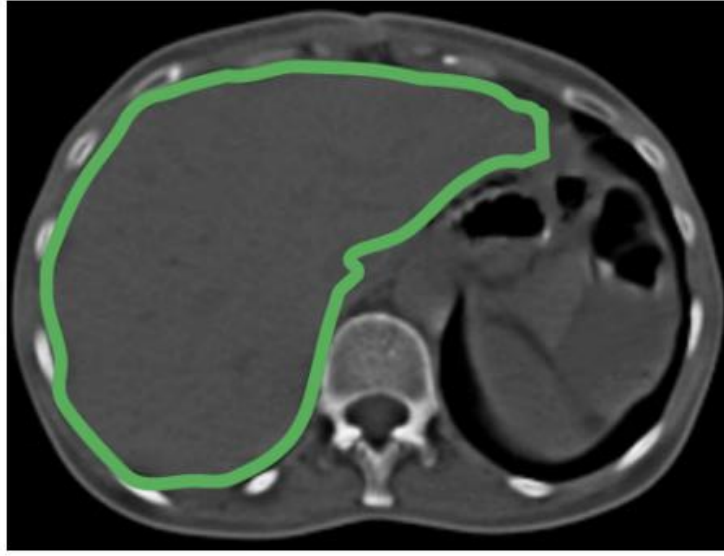


Fig. 1.7 Contour-based illustration

1.3.2.1 Average Symmetric Surface Distance

ASSD is calculated by measuring the average distance from the surface points of the segmented liver to ground truth [55, 58]. To ensure symmetry and a comprehensive evaluation, this process is often performed in both directions: from the segmented liver to the ground truth and vice versa. This bidirectional approach ensures that the metric accounts for all discrepancies, not just the most favorable ones. The ASSD equation is as presented in (1.5).

$$ASSD = \frac{1}{2} \left(\frac{\sum_{i=1}^N d(S_i, G)}{N} + \frac{\sum_{j=1}^M d(G_j, S)}{M} \right) \quad (1.5)$$

where:

- S_i denotes a point on the segmented liver surface.
- G_j denotes a point on the ground truth liver surface.
- $d(S_i, G)$ is the shortest distance from the point S_i on the segmented surface to any point on the ground truth surface G .
- $d(G_j, S)$ is the shortest distance from the point G_j on the ground truth surface to any point on the segmented surface S .
- N is the total number of points on the segmented liver surface.
- M is the total number of points on the ground truth liver surface.

Figure 1.8 illustrates the calculation of the ASSD metric in order to translate the mathematical equation into a visual explanation.

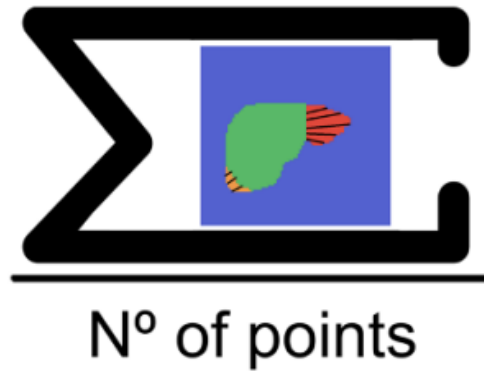


Fig. 1.8 Illustration of the calculation of ASSD, with TP in green, TN in blue, FN in red, and FP in orange

The ASSD value provides a numerical representation of the segmentation performance, with lower values indicating a higher accuracy and a better alignment between the segmented liver and the ground truth.

1.3.2.2 Maximum Symmetric Surface Distance

Maximum Symmetric Surface Distance (MSSD), also known as the Symmetric Hausdorff Distance, measures the maximum distance between the surfaces of the segmented liver region and the ground truth. MSSD focuses on the worst-case scenario by identifying the largest single discrepancy between the two surfaces [55, 58].

The computation of MSSD involves calculating the maximum distance from any point on the segmented liver surface to the closest point on the ground truth surface and vice versa. This ensures that the metric captures the most important errors in segmentation, regardless of their direction. The MSSD is mathematically defined as in Equation 1.6.

$$\text{MSSD} = \max \left(\max_{i \in S} \min_{j \in G} d(S_i, G_j), \max_{j \in G} \min_{i \in S} d(G_j, S_i) \right) \quad (1.6)$$

where:

- S_i represents a point on the surface of the segmented liver region.
- G_j represents a point on the surface of the ground truth liver region.
- $d(S_i, G_j)$ denotes the Euclidean distance between a point S_i on the segmented surface and a point G_j on the ground truth surface.
- $d(G_j, S_i)$ denotes the Euclidean distance between a point G_j on the ground truth surface to a point S_i on the segmented surface.

Figure 1.9 illustrates the calculation of the MSSD metric in order to translate the mathematical equation into a visual explanation.

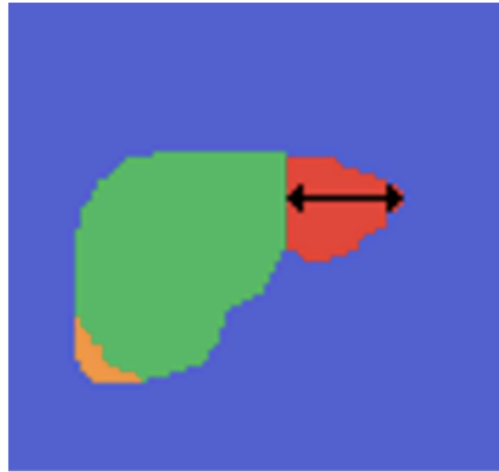


Fig. 1.9 Illustration of the calculation of MSSD, with TP in green, TN in blue, FN in red, and FP in orange

MSSD is expressed in physical units, such as millimeters, reflecting the spatial resolution of the CT scans. A lower MSSD value indicates a closer match between the segmented liver and the ground truth, signifying higher segmentation accuracy. Conversely, a higher MSSD value points to discrepancies between the segmented and ground truth surfaces, highlighting areas where the segmentation algorithm may need improvement.

1.4 Research

Research on liver segmentation in medical imaging demonstrates promise, yet the lack of consistent evaluation metrics is often overlooked. The upcoming analysis will prioritize examining these metrics, focusing on those commonly used in automatic segmentation studies. The selected studies for the analysis of the evaluation metrics

most commonly used, dated between 2014 and 2024, employ fully automatic liver segmentation techniques and metrics aligned with those in related works (Table 1.1). The research was sourced from **Google Scholar**, **Scopus**, and **Association for Computing Machinery Digital Library (ACM Digital Library)**, using keywords like “*Liver Segmentation*” and “*Segmentation Techniques*” to identify relevant studies.

Table 1.1 Summary of segmentation methods, main performance metrics used in the selected studies, and abbreviations.

Author	Year	Main Metrics Used
Platero et al. [41]	2014	DSC
Mostafa et al. [32]	2015	Accuracy
Dou et al. [15]	2016	ASSD; VOE
Chartrand et al. [10]	2017	DSC
Christ et al. [12]	2017	DSC
Zhang et al. [61]	2017	DSC
Rafiei et al. [43]	2018	DSC
Hiraman [20]	2018	ASSD; MSSD; VOE
Ahmad et al. [1]	2019	DSC
Wang et al. [56]	2019	DSC
Lebre et al. [25]	2019	DSC
Almotairi et al. [2]	2020	Accuracy
Ayalew et al. [5]	2021	DSC
Pettit et al. [40]	2022	Accuracy
Sabir et al. [45]	2022	Accuracy; DSC; Precision
Scicluna [47]	2022	ASSD; DSC; MSSD
Olivares et al. [37]	2023	Accuracy; DSC
Ezzat et al. [16]	2023	Accuracy
Shao et al. [48]	2024	ASSD; DSC; Precision

ASSD: Average Symmetric Surface Distance

MSSD: Maximum Symmetric Surface Distance

DSC: Dice Similarity Coefficient

VOE: Volumetric Overlap Error

As a result, the most commonly used metrics in these studies were the ones shown in Figure 1.10, which have been selected for analysis in this sub-chapter.

The results presented indicate that DSC, Accuracy, and ASSD stand out as the most frequently employed metrics in liver segmentation evaluations. Despite some studies validating the effectiveness of these metrics in tackling common challenges such as under-segmentation and over-segmentation, there remains an evident need for the adoption of standardized metrics in the field. We advocate for the use of at least one metric within each type of evaluation metric (region and contour). The introduction of such standards would greatly advance segmentation studies by underscoring key aspects, enabling straightforward comparisons across various methods,

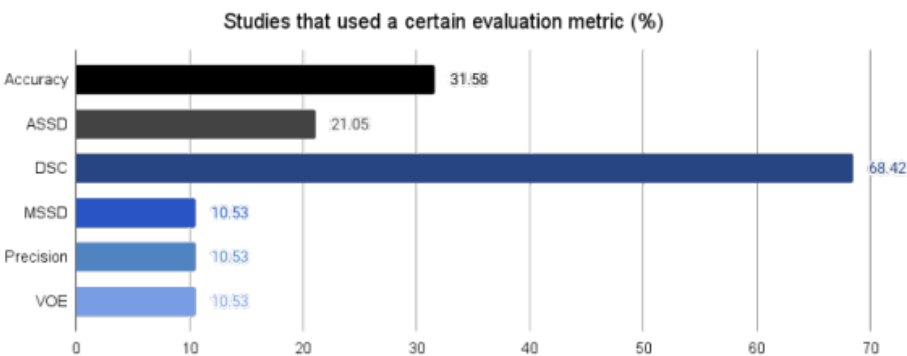


Fig. 1.10 Distribution of the metrics used in the selected studies

and improving the reliability and effectiveness of automated liver segmentation in medical imaging.

1.5 Discussion

Table 1.2 presents an overview of the metrics, along with their main advantages and shortcomings.

Despite some studies validating the effectiveness of these metrics in tackling common challenges such as under-segmentation (Figure 1.11 (a)) and over-segmentation (Figure 1.11 (b)), which often appears when the Hounsfield unit (HU) of liver is close to the HU of other tissues or background, there remains an evident need for the adoption of standardized metrics in the field.

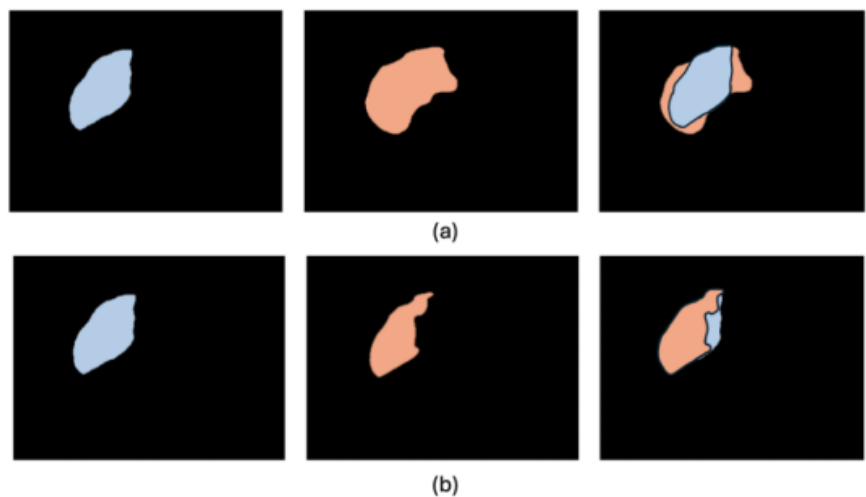


Fig. 1.11 Common segmentation challenges. (a) Under-segmentation. (b) Over-segmentation.

Table 1.2 Summary of the main evaluation metrics used by automatic liver segmentation studies.

Category	Main Advantages	Main Disadvantages
Accuracy	- Provides a comprehensive overview of the segmentation performance. - Easy to calculate and interpret.	- May be misleading in datasets with a large imbalance between liver and non-liver voxels. - Does not specifically focus on liver tissue accuracy.
Precision	- Highlights the ability of the algorithm to avoid over-segmentation. - Crucial for applications where precision is more important than recall.	- Does not account for missed liver tissue (FN). - Can be misleading if the algorithm is overly conservative.
DSC	- Provides a clear, straightforward measure of segmentation accuracy. - Easy to interpret and widely used in medical imaging.	- Might not fully capture small discrepancies in boundary delineation. - Can be overly optimistic in cases of large overlap but poor boundary alignment.
VOE	- Directly related to clinical relevance, reflecting the actual volume error. - Intuitive and easy to understand.	Can be sensitive to the size of the liver, with larger livers potentially showing more absolute errors.
ASSD	- Offers a detailed assessment of boundary accuracy. - Highlights discrepancies in the surface alignment of the segmentation.	- More complex to compute and interpret compared to overlap-based metrics. - Can be influenced by outliers or extreme values.
MSSD	- Identifies the worst-case scenario in segmentation discrepancies. - Useful for assessing the reliability of segmentation in critical areas.	- Focused on the maximum error, potentially overlooking overall segmentation quality. - Can be overly punitive for generally accurate segmentations with minor outliers.

- **Under-segmentation**

Represents when the regions within an image have not been sufficiently segmented, leading to the merging of regions that should be separate.

- **Over-segmentation**

Occurs when the regions within an image are divided into too many segments, creating unnecessary and small regions that should be combined.

It is considered ideal to use at least one metric within each type of evaluation metric (region and contour). The introduction of such standards would greatly advance segmentation studies by highlighting key aspects, allowing for direct comparisons across various methods, and improving the reliability and effectiveness of automated liver segmentation in medical imaging.

1.6 Conclusions

Non-invasive imaging techniques have vast potential to transform patient care, emphasizing the need for consistent metrics in automatic liver segmentation. By developing unified, quantitative standards, the accuracy, reliability, and comparability of segmentation methods can be enhanced. This will not only aid clinicians but also steer advancements toward more effective and user-friendly solutions.

Standardizing evaluation metrics is essential for progressing liver segmentation technology. Clear metrics will enable better comparisons, foster innovation, and improve the development and validation of new methods. This will directly benefit clinical practices by improving diagnostics and treatment planning.

It is crucial to align technological developments with clinical needs to incorporate automatic liver segmentation into routine medical care. This alignment will improve diagnoses and help customize treatments for liver diseases, enhancing patient outcomes.

References

1. M. Ahmad et al. Deep belief network modeling for automatic liver segmentation. *IEEE Access*, 7:20585–20595, 2019.
2. S. Almotairi, G. Kareem, et al. Liver tumor segmentation in CT scans using modified SegNet. *Sensors*, 20(5):1516, 2020.
3. American Institute of Physics. Marie Curie - war duty (1914-1919). <https://history.aip.org/exhibits/curie/war1.htm>. Accessed March 24, 2024.
4. M.Y. Ansari, A. Abdalla, M.Y. Ansari, M.I. Ansari, B. Malluhi, S. Mohanty, S. Mishra, S.S. Singh, J. Abinshed, and A. Al-Ansari. Practical utility of liver segmentation methods in clinical surgeries and interventions. *BMC Medical Imaging*, 22:97, 2022.
5. Y.A. Ayalew, K.A. Fante, and M.A. Mohammed. Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method. *BMC Biomedical Engineering*, 3(1):4, 2021.
6. S. Batista, I. Domingues, M. Couceiro, and R. Filipe. Performance evaluation metrics for automatic liver segmentation on medical imaging scans. In *International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMA2SC)*, Porto, Portugal, 2024.
7. M. Bechar, N. Settouti, and I. Domingues. *Deep Learning vs. Super Pixel Classification for Breast Masses Segmentation*, pages 121–156. CRC Press, Boca Raton, FL, USA, 2021.
8. J.L. Boxerman, T.J. Mosher, E.R. McVeigh, E. Atalar, J. Lima, and D.A. Bluemke. Advanced MR imaging techniques for evaluation of the heart and great vessels. *Radiographics*, 18(3):543–564, 1998.
9. I. Carbone, P. Martins, A. Teixeira, and A. Silva. A vocal tract segmentation and analysis over a european portuguese MRI database. *Electrónica e Telecomunicações*, 4:1050–1053, 2008.
10. G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, et al. Liver segmentation on CT and MR using laplacian mesh optimization. *IEEE Transactions on Biomedical Engineering*, 64, 2017.
11. G. Chlebus, A. Schenk, J.H. Moltz, B. van Ginneken, H.K. Hahn, and H. Meine. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Scientific Reports*, 8:15497, 2018.

12. P.F. Christ, F. Ettlinger, F. Grün, M.E.A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, et al. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. *arXiv*, 2017.
13. Directorate General of Health. National health plan 2012-2016, 2012.
14. I. Domingues and J.S. Cardoso. Using bayesian surprise to detect calcifications in mammogram images. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1091–1094, Chicago, IL, USA, 2014.
15. Q. Dou, H. Chen, et al. 3D deeply supervised network for automatic liver segmentation from CT volumes. In *19th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 149–157, 2016.
16. K.A. Ezzat, L.N. Omran, and A. El Seddawy. Automatic liver segmentation in computed tomography scans using deep semantic segmentation. *Bulletin of Electrical Engineering and Informatics*, 12(1):250–256, 2023.
17. S. Francque. The liver and the cardiovascular system: Two of a kind? *Journal of the American Heart Association*, 10:e020286, 2021.
18. M. Friebe and A. Illanes. Advancements in medical imaging and image-guided procedures: A potential-or rather likely-paradigm shift in diagnosis and therapy: Understand disruption and take advantage of it! *Applied Sciences*, 13:9218, 2023.
19. T. Heimann, B. Van, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, 2009.
20. A. Hiranman. Liver segmentation using 3D CT scans. Master's thesis, University of Kwazulu-Natal, 2018.
21. J.M. Hoffman. New advances in brain tumor imaging. *Current Opinion in Oncology*, 13(3):148–153, 2001.
22. N. Huynh. Understanding evaluation metrics in medical image segmentation. https://medium.com/@nghihuynh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a3732023. Accessed: October 30, 2023.
23. A. Kalra, E. Yetiskul, C.J. Wehrle, et al. Physiology, liver. <https://www.ncbi.nlm.nih.gov/books/NBK535438>, 2023. Accessed October 23, 2023.
24. D.C. Le, K. Chinnasarn, J. Chansangrat, N. Keeratibharat, and P. Horkaew. Semi-automatic liver segmentation based on probabilistic models and anatomical constraints. *Scientific Reports*, 11:6106, 2021.
25. M. Lebre et al. A robust multi-variability model-based liver segmentation algorithm for CT-scan and MRI modalities. *Computerized Medical Imaging and Graphics*, 2019.
26. R. Marinho. Conselhos para o bem-estar do fígado. <https://www.future-healthcare.pt/news2/sp/doenca.html>. Accessed March 20, 2024.
27. P. Martins, I. Domingues, A. Silva, and A. Teixeira. An MRI study of european portuguese nasals. In *INTERSPEECH*, pages 58–61. ISCA, 2007.
28. Mayo Clinic. Liver disease. <https://www.mayoclinic.org/diseases-conditions/liver-problems/diagnosis-treatment/drc-20374507>, 2024. Accessed October 29, 2023.
29. B. Mendes, I. Domingues, and J. Santos. Multi-class semantic segmentation for prostate cancer radiotherapy treatment optimization. In *International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMA2SC)*, Porto, Portugal, 2022.
30. B. Mendes, I. Domingues, A. Silva, and J. Santos. Prostate cancer aggressiveness prediction using CT images. *Life*, 11:1164, 2021.
31. A. Mharib, A. Ramli, S. Mashohor, and R. Mahmood. Survey on liver CT image segmentation methods. *Artificial Intelligence Review*, 37(2):83–95, 2012.
32. A. Mostafa, A. Fouad, et al. CT liver segmentation using artificial bee colony optimisation. In *Procedia Computer Science*, pages 1622–1630, 2015.
33. National Cancer Institute. What is liver cancer. <https://www.cancer.gov/types/liver/what-is-liver-cancer>, 2024. Accessed October 23, 2023.

34. H. Nishikawa and Y. Osaki. Liver cirrhosis: Evaluation, nutritional status, and prognosis. *Mediators of Inflammation*, 2015:872152, 2015.
35. S.B. Niño, J. Bernardino, and I. Domingues. Algorithms for liver segmentation in computed tomography scans: A historical perspective. *Sensors*, 24:1752, 2024.
36. N. Ohri, L.A. Dawson, S. Krishnan, et al. Radiotherapy for hepatocellular carcinoma: New indications and directions for future study. *Journal of the National Cancer Institute*, page 108, 2016.
37. J. Olivares, N. Satpute, and J. Gómez-Luna. Cross-modality-guided contrast enhancement on liver segmentation. In *18th Iberian Conference on Information Systems and Technologies (CISTI)*, 2023.
38. A. Oliveira, I. Domingues, H. Duarte, J. Santos, and P. Abreu. Going back to basics on volumetric segmentation of the lungs in CT: A fully image processing based technique. In *Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, volume 11868 LNCS, pages 322–334, Berlin/Heidelberg, Germany, 2019. Springer.
39. G. Pereira, I. Domingues, P. Martins, P.H. Abreu, H. Duarte, and J. Santos. Registration of CT with PET: A comparison of intensity-based approaches. In *International Workshop on Combinatorial Image Analysis (IWCIA)*, pages 134–149. Springer, 2018.
40. R. Pettit, B.B. Marlatt, S.J. Corr, J. Havelka, and A. Rana. nnU-Net deep learning method for segmenting parenchyma and determining liver volume from computed tomography images. *Annals of Surgery Open: perspectives of surgical history, education, and clinical approaches*, 3(2), 2022.
41. C. Platero and M.C. Tobar. A multiatlas segmentation using graph cuts with applications to liver segmentation in CT scans. *Computational and Mathematical Methods in Medicine*, 2014.
42. R. Punia and S. Singh. Review on machine learning techniques for automatic segmentation of liver images. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(4):666–670, 2013.
43. S. Rafiei et al. Liver segmentation in CT images using three dimensional to two dimensional fully convolutional network. In *25th IEEE International Conference on Image Processing (ICIP)*, pages 2067–2071, 2018.
44. S.J. Riederer. Recent technical advances in MR imaging of the abdomen. *Journal of Magnetic Resonance Imaging*, 6(5):822–832, 1996.
45. M. Sabir et al. Segmentation of liver tumor in CT scan using ResU-Net. *Applied Sciences*, 12(17):8650, 2022.
46. P. Sachdev. What is liver failure? <https://www.webmd.com/digestive-disorders/digestive-diseases-liver-failure>, 2023. Accessed October 28, 2023.
47. D. Scicluna. Automatic segmentation of healthy liver in abdominal computed tomography scans. Master's thesis, University of Malta, 2022.
48. J. Shao, S. Luan, et al. Attention connect network for liver tumor segmentation from CT and MRI images. *Technology in Cancer Research and Treatment*, 23, 2024.
49. A. Sorensen. Magnetic resonance as a cancer imaging biomarker. *Journal of Clinical Oncology*, 24(20):3274–3281, 2006.
50. S.K. Tanguturi, J.Y. Wo, A.X. Zhu, L.A. Dawson, and T.S. Hong. Radiation therapy for liver tumors: ready for inclusion in guidelines? *Oncologist*, 19:868–879, 2014.
51. A. Torres. *Segmentação de Imagens Médicas Visando a Construção de Modelos Médicos*. PhD thesis, Instituto Politécnico de Bragança, Escola Superior de Tecnologia e Gestão, 2012.
52. E. Trefts, M. Gannon, and D.H. Wasserman. The liver. *Current Biology*, 27:R1147–R1151, 2017.
53. S. Verissimo. Encefalopatia hepática: o que é, sintomas e tratamento. <https://www.maisquecuidar.com/encefalopatia-hepatica-o-que-e-sintomas-e-tratamento>. Accessed March 20, 2024.
54. F. Vernuccio, R. Cannella, T.V. Bartolotta, M. Galia, A. Tang, and G. Brancatelli. Advances in liver US, CT, and MRI: moving toward the future. *European Radiology Experimental*, 5:52, 2021.

55. E. Vorontsov, M. Cerny, P. Régner, et al. Deep learning for automated segmentation of liver lesions at CT in patients with colorectal cancer liver metastases. PowerPoint Presentation, 2019. Journal Reading.
56. K. Wang, A. Mamidipalli, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. *Radiology: AI*, 1(2):180022, 2019.
57. E.M. Wu, L.L. Wong, B.Y. Hernandez, J.F. Ji, W. Jia, S.A. Kwee, and S. Kalathil. Gender differences in hepatocellular cancer: disparities in nonalcoholic fatty liver disease/steatohepatitis and liver transplantation. *Hepatoma Research*, 4:66, 2018.
58. V. Yeghiazaryan and V. Irina. An overview of current evaluation methods used in medical image segmentation. Technical report, University of Oxford, Department of Computer Science, 2015.
59. N. Ying-Hwey, W. Bernice, L. Nadya, S. O'Doherty, et al. Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset. *Computers in Biology and Medicine*, 134:104497, 2021.
60. H. Zhang, K. Luo, R. Deng, S. Li, and S. Duan. Deep learning-based CT imaging for the diagnosis of liver tumor. *Computational Intelligence and Neuroscience*, pages 1–7, 2022.
61. Y. Zhang, Z. He, and C. Zhong. Fully convolutional neural network with post-processing methods for automatic liver segmentation from CT. In *Chinese Automation Congress (CAC)*, 2017.
62. L.Q. Zhou, J.Y. Wang, S.Y. Yu, G.G. Wu, Q. Wei, Y.B. Deng, X.L. Wu, X.W. Cui, and C.F. Dietrich. Artificial intelligence in medical imaging of the liver. *World Journal of Gastroenterology*, 25:672–682, 2019.

Appendix J – RECPAD 2024 Publication

Paper submitted to the *Portuguese Conference on Pattern Recognition (RECPAD 2024)* (Adoption of Artificial Neural Network-based Methods for Medical Image Segmentation):

Adoption of Artificial Neural Network-based Methods for Medical Image Segmentation

Stephanie Batista¹
stephanieabn@gmail.com

Miguel Couceiro^{1,2,3}
couceiro@isec.pt

Ricardo Filipe⁴
ricardo.angelo.filipe@gmail.com

Paulo Rachinhas⁵
rac@ulsc Coimbra.min-saude.pt

Jorge Isidoro⁵
jisidoro@ulsc Coimbra.min-saude.pt

Inês Domingues^{1,6}
inesdomingues@gmail.com

¹ Polytechnic University of Coimbra, Rua da Misericórdia, Lagar dos Cortiços, S. Martinho do Bispo, 3045-093 Coimbra, Portugal

² Institute of Applied Research (i2A), Portugal

³ Laboratory for High Performance Computing (LaCED), Portugal

⁴ Altice Labs, S.A, Portugal

⁵ Coimbra Hospital and University Center, Portugal

⁶ IPO Porto Research Center (CI-IPOP): Medical Physics, Radiobiology & Radiation Protection Research Group, Portugal

Abstract

Medical imaging segmentation is a crucial process for diagnostics and treatment planning, particularly for complex organs. Traditional image segmentation techniques, such as semantic segmentation, edge detection and region-growing methods, face significant challenges when applied to medical data, which often contains noise, artifacts, and irregular organ boundaries. With the advancements of artificial intelligence, particularly artificial neural networks, the potential for more accurate and efficient segmentation has become evident. This paper explores the choice of adopting artificial neural networks-based techniques instead of traditional methods, focusing on the benefits and challenges of adopting these modern approaches in medical image segmentation.

1 Introduction

Medical imaging segmentation is the process of dividing digital medical images into segments to simplify and modify the representation of the image into something that is easier to analyse. This technique is especially useful for diagnostics, treatment planning and monitoring of various diseases (Figure 1). Accurate segmentation of medical images allows for better visualization and analysis of anatomical structures, improving decision-making in clinical settings [1]. Traditional image segmentation methods, such as thresholding, clustering, and region-based approaches, have been widely used, but they often struggle with the complexity and variability of medical data, particularly in images affected by noise, artifacts, or irregular organ boundaries. The introduction of Artificial Neural Network (ANN)-based techniques has revolutionized this field, offering more precise and efficient segmentation capabilities.

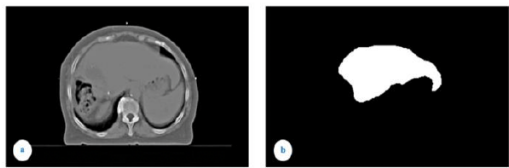


Figure 1: Medical image: (a) Original DICOM image. (b) Liver segmentation

ANNs, especially Convolutional Neural Networks (CNNs), have proven to be particularly effective in medical image segmentation due to their ability to learn complex patterns and features from large datasets. These models, combined with advances in computational power and availability of annotated medical images, have enabled significant improvements in segmentation accuracy and processing speed. This shift from traditional methods to ANN-based approaches is driving innovation in fields such as radiology, oncology, and histology, where precise segmentation is essential for disease detection, treatment planning, and monitoring patient outcomes [2]. This paper compares the adoption of ANN-based methods over traditional techniques, highlighting the growing role of artificial intelligence (AI) in advancing medical image segmentation.

2 Image Segmentation Techniques

The evolution of image processing technologies has given rise to a spectrum of segmentation categories. Inspired by existing segmentation techniques, authors in [3] noted that these methods can be systematically classified. While common types like instance and semantic segmentation

stand alone, other techniques often involve AI, thresholding, clustering, or other region/edge-based approaches.

Table 1 provides an overview of the image segmentation techniques previously mentioned, focusing on their definition, strengths, and weaknesses, taking into account that each approach has its own niche depending on the image type, complexity, and computational requirements.

Table 1. Image segmentation techniques description

Technique	Description
Threshold	Uses intensity thresholds for fast, simple segmentation. Effectiveness depends on threshold selection, sensitive to noise.
Cluster	Identifies similar data patterns or clusters, ideal for images with clear patterns. Sensitive to noise and irregular shapes.
Edge	Focuses on intensity transitions to delineate boundaries, effective for well-contrasted images but sensitive to noise.
Region	Divides an image into homogeneous regions, simple and intuitive but may miss fine details and complex boundaries.
Semantic	Categorizes each pixel based on its meaning, offering fine-grained segmentation but requires high computational power.
Neural Network	Uses ANN models to segment images based on learned features. Requires large training datasets, computationally intensive.

A comprehensive study carried out in [4] through the analysis of 28 studies, one study per year between the years 1990 and 2024, presented the distribution of the use of segmentation methods shown in Figure 2. The adoption of ANNs for the segmentation process over the years stands out, and it is interesting to evaluate the fact that the other most used methods are based on edges and regions.

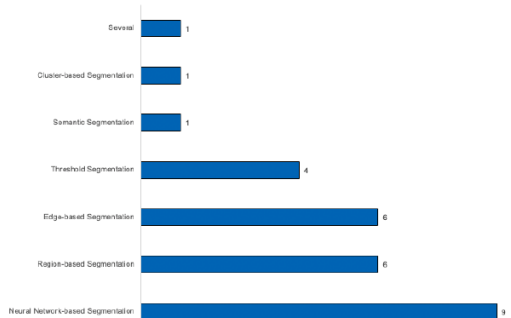


Figure 2: Distribution of image segmentation techniques

3 Artificial Neural Network-based Segmentation

Through the integration of ANNs into medical imaging, ANNs can analyse large volumes of medical data to detect patterns and anomalies with greater accuracy than traditional methods. The number of publications using ANN techniques to analyse static medical images has risen from hundreds to thousands in recent years. This growth is driven by the pattern-recognition demands of diagnostic tasks and the availability of organized images. Key applications in medical imaging include denoising, automatic segmentation, case triage, and image restoration [2].

The application of medical image segmentation extends various medical fields, significantly impacting areas such as radiology, gastroenterology, histology, ultrasound imaging, and cancer detection. For instance, in radiology, AI-based segmentation helps in processing vast amounts of images from X-rays, computed tomography (CT) scans, and magnetic resonance imaging (MRI), enabling more accurate and scalable diagnostic capabilities. Similarly, in gastroenterology, segmentation techniques can improve the detection of conditions like cancerous polyps and ulcers, often identifying issues that might be overlooked by human specialists. Histology benefits from segmentation by accurately applying staining protocols, thus aiding in the detailed analysis of tissue samples [5].

Automation and computational costs are crucial concepts that expand the understanding of AI applications in healthcare. Automation refers to the use of technology to perform tasks with minimal or no human intervention, which enhances the efficiency of clinical workflows, reduces manual errors, and allows healthcare professionals to focus more on patient care. On the other hand, computational costs involve the resources, such as time, memory, and power, required to process and analyse large datasets and complex algorithms. These costs are particularly significant in the context of AI in healthcare, where efficient management is essential to ensuring the sustainability and effectiveness of AI-driven applications.

Based on the studies described in [4], Table 2 shows a more specific comparison of the main results of studies between 2016 and 2024 that employed ANNs as their segmentation method, in which it is interesting to observe that a significant percentage of these studies opted for using CNNs due to their ability to automatically learn features such as edges, textures, and shapes, making them highly effective in identifying structures in modalities such as MRI and CT scans. Also widely used within ANNs is the U-Net segmentation method, a well-known encoder-decoder model, which is particularly suitable for small medical datasets, utilizing a U-shaped structure to capture context and provide precise localization, which is especially useful in tasks such as medical organs segmentation.

Table 2. ANN-based segmentation methods used by the selected studies

Author	Year	Method	Main results
Dou et al. [6]	2016	3D DSN	5.42% VOE; 0.79mm ASSD
Christ et al. [7]	2017	CFCN	94.3% mean DSC
Hiraman [8]	2018	CNN	2.25mm ASSD; 43.01mm MSSD
Wang et al. [9]	2019	CNN	0.94 ± 0.06 DSC
Almotairi et al. [10]	2020	SegNet	94.57% overall accuracy
Ayalew et al. [11]	2021	U-Net	0.9612 DSC
Scicluna [12]	2022	U-Net and VGG16UNetC	97.85 DSC; 94.80 ASSD; 70.38 MSSD
Ezzat et al. [13]	2023	CNN	98.80% accuracy
Shao et al. [14]	2024	AC-Net	0.90 DSC; 0.89 precision; 4.59 ASSD

AC-Net: Attention Connect Network
ASSD: Average Symmetric Surface Distance
CFCN: Cascaded fully CNN
CNN: Convolutional Neural Network

DSC: Dice Similarity Coefficient
DSN: Deeply Supervised Network
MSSD: Maximum Symmetric Surface Distance
VOE: Volumetric Overlap Error

4 Conclusions

The growth of ANNs-based approaches has led to remarkable progress in medical image segmentation. These models have led to a new era of accuracy and efficiency, significantly outperforming traditional methods. This improved accuracy has become critical in the field of medical imaging, where the correct interpretation of medical scans can directly affect the diagnosis and treatment plans of patients. Furthermore, the efficiency of these ANN models translates into faster processing times, allowing for more agile decision making in clinical settings.

Another positive aspect of ANNs in medical segmentation is their ability to cope with the complexity of the anatomy of the organs. ANNs, with their systematic and complex pattern recognition capabilities, can navigate these anatomical variations. As a result, they are better able to deal with the variety of appearances that tissues can have on medical scans. This ability to handle complex datasets ensures that ANNs can provide consistent and accurate segmentation in a wide range of cases.

Applying ANNs in healthcare involves many fronts, redefining the paradigms of patient care through more accurate diagnosis, optimized treatment guidelines and better patient management in various domains. On a specific front, the adoption of ANNs is increasingly applied into medical imaging, where they interpret large sets of medical data, identifying patterns and predicting health outcomes with a level of precision and speed.

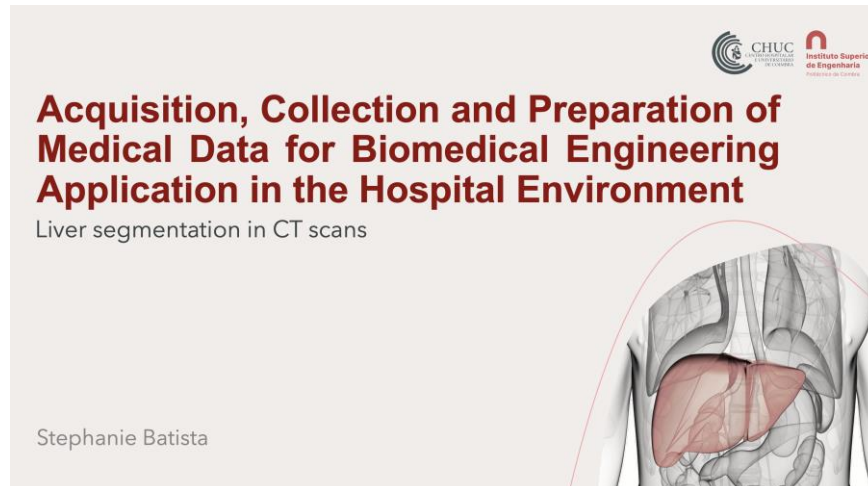
References

- [1] Conze, P.-H.; Kavur, A. E.; Gall, E. C.-L.; Gezer, N. S.; Meur, Y. L.; Selver, M. A.; Rousseau, F., "Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks", *Artificial Intelligence in Medicine*, 117:102109, July 2021
- [2] Sakib, S.; Ahmed, N.; Kabir, A. J.; Ahmed, H. "An Overview of Convolutional Neural Network: Its Architecture and Applications", *Preprints* 2018, 2018110546
- [3] Sakshi, K.V., "Image Segmentation Techniques: Statistical, Comprehensive, Semi-Automated Analysis and an Application Perspective Analysis of Mathematical Expressions", *Archives of Computational Methods in Engineering* 2023, 30, 457-495
- [4] Niño, S.B.; Bernardino, J.; Domingues, I. "Algorithms for Liver Segmentation in Computed Tomography Scans: A Historical Perspective". *Sensors* 2024, 24, 1752. doi: 10.20944/preprints202402.0464.v1
- [5] Encord, "Medical Image Segmentation: A Complete Guide", 2023. Retrieved October 30, 2023 from <https://encord.com/blog/medical-image-segmentation>
- [6] Dou, Q.; Chen, H.; et al., "3D deeply supervised network for automatic liver segmentation from CT volumes", *19th MICCAI*, (2016), 149-157
- [7] Christ, P.F.; Ettlinger, F.; Grün, F.; Elshaera, M.E.A.; Lipkova, J.; Schlecht, S.; Ahmady, F.; Tatavarty, S.; Bickel, M.; Bilic, P.; et al. "Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks", *arXiv* 2017, arXiv: 1702.05970
- [8] Hiraman, A., "Liver segmentation using 3D CT scans", *University of Kwazulu-Natal*, 2018
- [9] Wang, K.; Mamidipalli, A.; et al., "Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network", *Radiology: AI*, 1(2) (2019), 180022
- [10] Almotairi, S.; Kareem, G.; et al., "Liver tumor segmentation in CT scans using modified SegNet", *Sensors*, 20(5) (2020), 1516
- [11] Ayalew, Y.A.; Fante, K.A.; Mohammed, M.A., "Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method", *BMC Biomedical Engineering*, 3(1) (2021), 4
- [12] Scicluna, D., "Automatic segmentation of healthy liver in abdominal computed tomography scans", *University of Malta*, 2022
- [13] Ezzat, K.A.; Omran, L.N.; El Seddawy, A., "Automatic liver segmentation in computed tomography scans using deep semantic segmentation", *Bulletin of Electrical Engineering and Informatics*, 12(1) (2023), 250-256
- [14] Shao, J.; Luan, S.; et al., "Attention Connect Network for Liver Tumor Segmentation from CT and MRI Images", *Technology in Cancer Research and Treatment*, 23, 2024

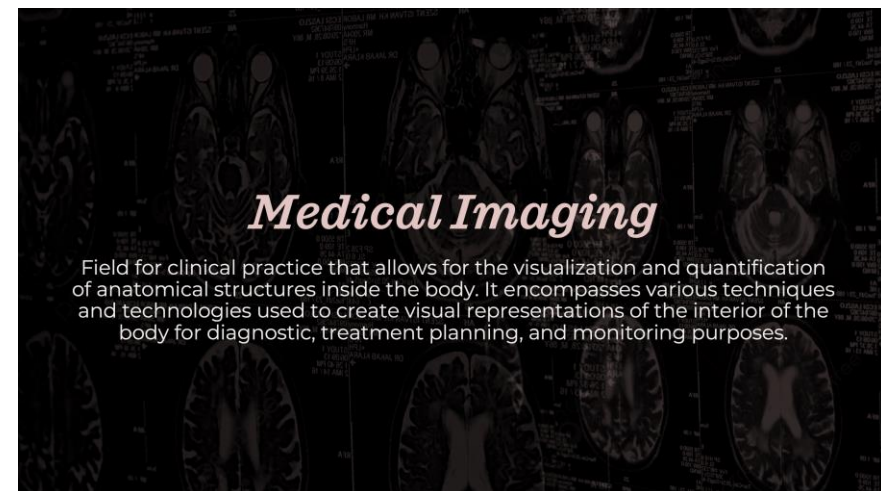
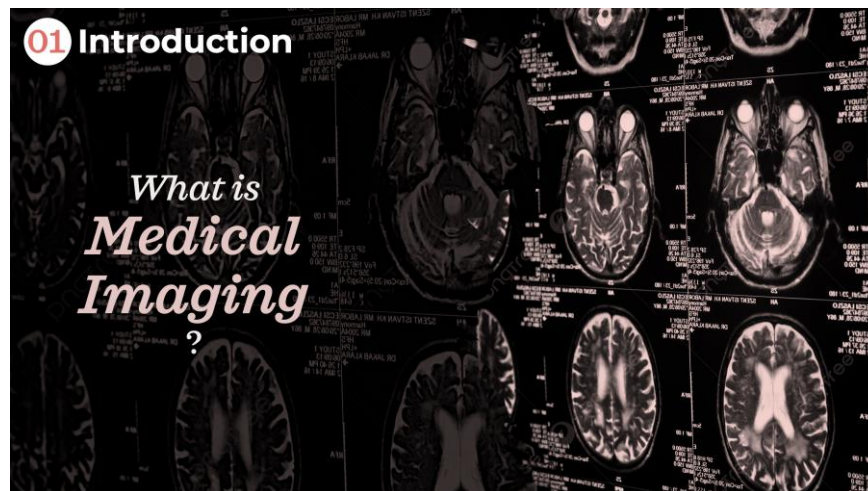
2D Liver Segmentation in Abdominal Computed Tomography Scans

Appendix K – Bachelor of Science in Biomedical Engineering Presentation

Presentation to the second-year students of the Bachelor of Science in Biomedical Engineering at ISEC within the scope of the Introduction to Clinical Databases course:



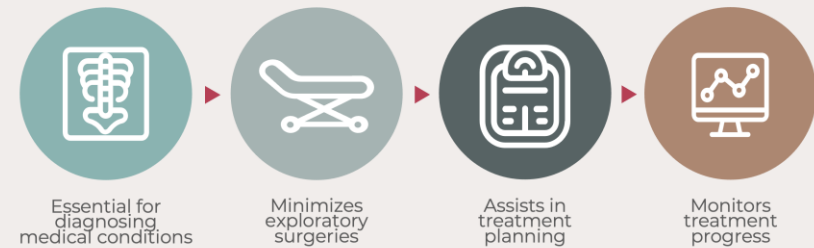
01	02	03	04	05	06
Introduction • Medical imaging • Importance • Main challenges	Data acquisition • Computed Tomography • DICOM images • Hospital Information Systems	Ethics and GDPR • Hospital's Ethics Committee • Compliance with GDPR	Data collection • DICOM metadata • Hospital's database access • Anonymization process	Data preparation • Radiotherapy Structure Set files • Pre-processing techniques • Data augmentation	Conclusions • Main challenges of the process • Some potential applications • Q&A



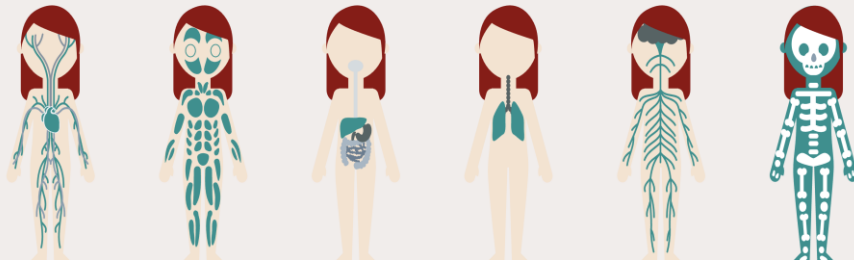
Medical Imaging Modalities

- | | |
|---|--|
|  Radiography (x-ray) |  Magnetic Resonance Imaging (MRI) |
|  Mammography |  Computed tomography (CT) |
|  Angiography |  Positron emission tomography (PET) |
|  Ultrasound |  Fluoroscopy |

Medical Imaging Importance



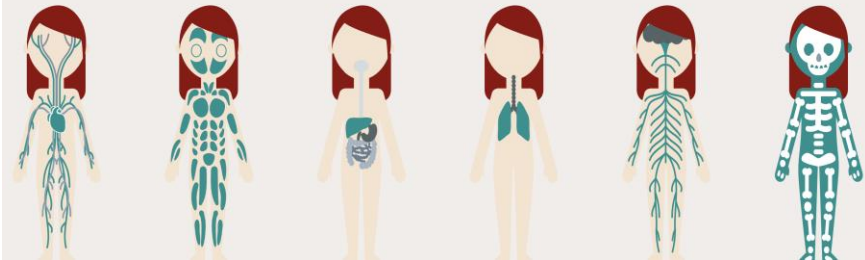
How many scans do you estimate were performed in **2021**?



12.900.000

radiological images in 2021

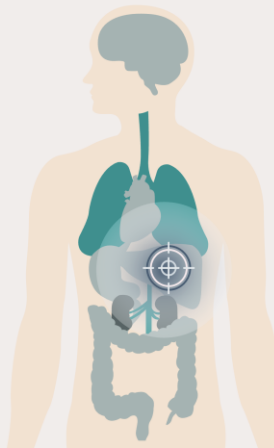
- According to the National Statistics Institute's Health Statistics 2021 (Portugal) -



2D Liver Segmentation in Abdominal Computed Tomography Scans

⚠ *However, medical imaging comes with some challenges....*

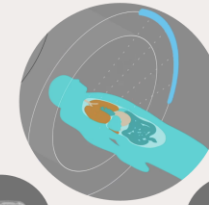
- Image quality
- Data management and storage
- Radiation exposure
- Image interpretation and analysis
- Data privacy and security



02 Data Acquisition

Computed Tomography

CT scanners use **X-rays** to create cross-sectional body images. The gantry **rotates around the patient**, emitting X-rays from multiple angles. These beams are attenuated differently by various tissues as they pass through the body.



Detectors measure the **attenuated X-rays**, and a computer reconstructs detailed images of internal structures. These images are combined into 3D representations, enabling thorough visualization of organs, bones, and soft tissues.

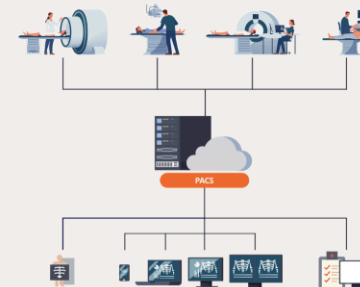


Standard that facilitates the communication of medical information using digital images and associated data.



Picture Archiving and Communication System (PACS)

System for **storing, managing and distributing** medical images.

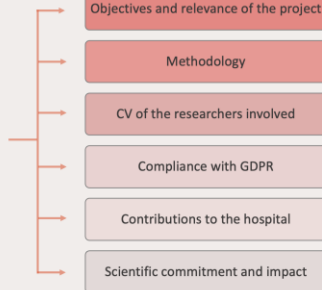


Allows healthcare professionals to **access and view images** electronically.

03 Ethics and GDPR

Submission to the Hospital's Ethics Committee

Composed by a set of documents to assess the compliance with several criteria:



General Data Protection Regulation

- ♥ Ensures protection of patients' personal and sensitive data
- ♥ Safeguards confidentiality and prevents unauthorized access

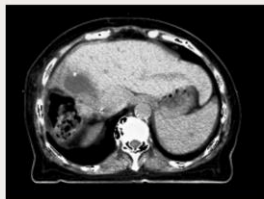


To comply with the GDPR, a **data anonymization** phase must be implemented.

04 Data Collection

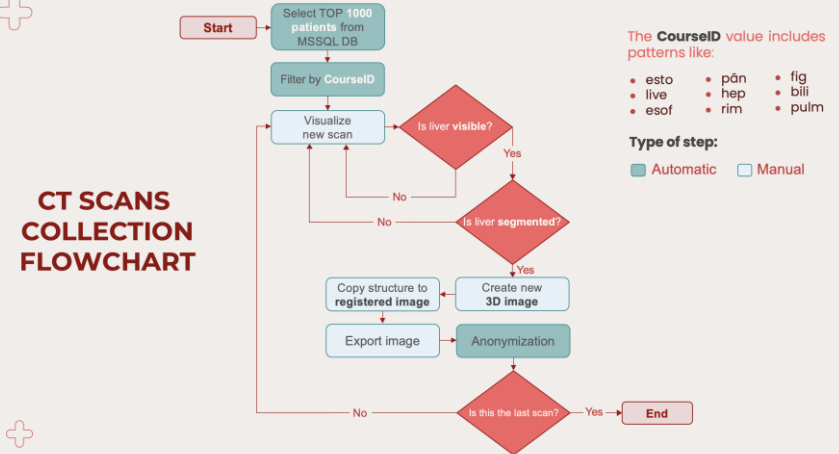
DICOM Metadata

Set of **standardized tags** that accompany medical images, containing crucial information about the patient, study, and image acquisition parameters, ensuring the identification, organization, and interpretation of medical images across different healthcare systems and devices.



—

—

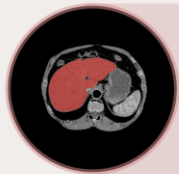
[illegible]

2D Liver Segmentation in Abdominal Computed Tomography Scans

Medical Image Data Anonymization

DICOM standards provide strict guidelines for anonymizing medical imaging data to protect patient privacy and comply with **GDPR**. These guidelines detail specific procedures for removing or replacing sensitive information within DICOM datasets.

Main commands for **data anonymization**:

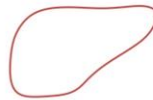


- "-m" **modifies** a **specific** DICOM tag
- "-ma" **modifies all instances** of a tag in the header
- "-ea" **removes all tags** from the DICOM image header



Radiotherapy Structure Set (RSST)

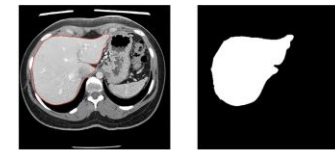
Specialized type of DICOM object that contains information about the **anatomical and clinical structures** of a patient, typically derived from CT scans.



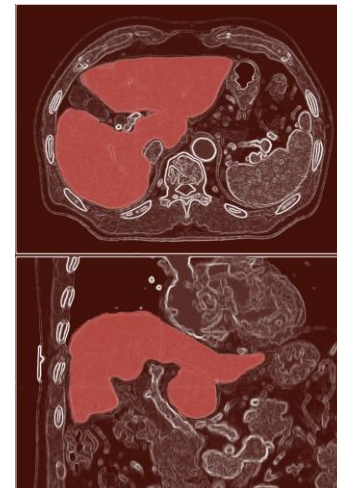
05 Data Preparation

Masks Creation

Based on the **RSST contours**, converting abstract contour data into specific pixel coordinates, translating the contours into a format that directly corresponds to the CT image's pixel grid.



The binary mask assigns a value of **1** to pixels inside the contour (**ROI**) and **0** to those outside.



Importance of Segmentation

- Precise detection
- Tumor dimension identification
- Vascular evaluation
- Identification of additional lesions
- Treatment planning



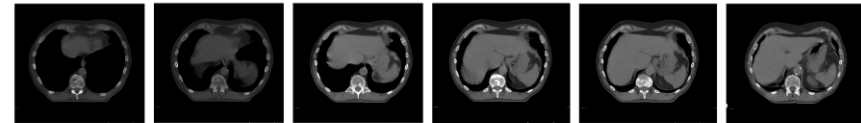


MANUAL SEGMENTATION CHALLENGES

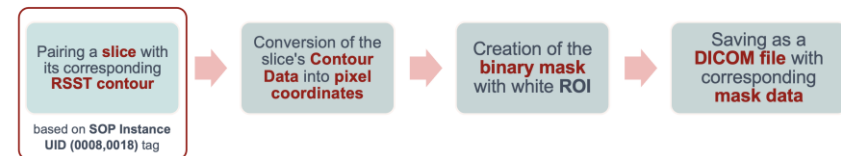
- Time-consuming
- Subject to human error
- Complexity of liver anatomy
- Variability in liver appearance
- Fatigue and eye strain
- Limited precision

Masks Creation

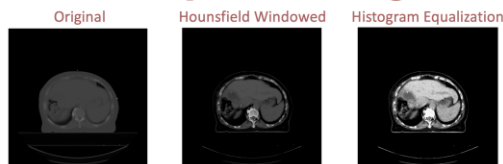
- Sort all slices in a scan by **Instance Number (IS) (0020, 0013)** tag



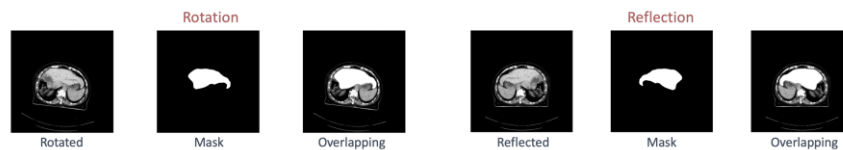
- Mask creation:



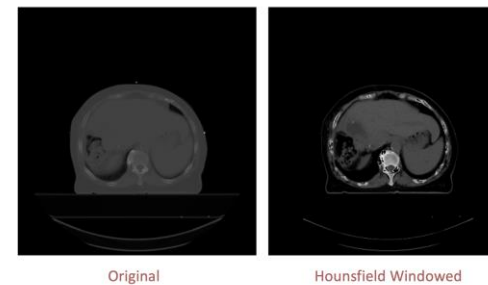
Pre-processing



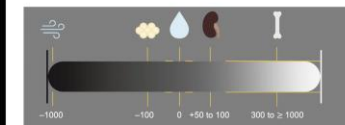
Data Augmentation



Hounsfield Windowed



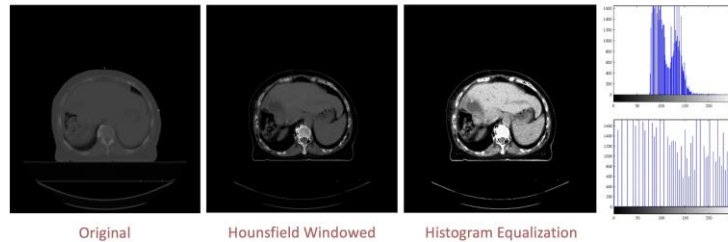
Technique used in CT imaging to enhance the visibility of specific tissues or structures. By focusing on a particular range of **Hounsfield Units (HU) values**, clinicians can better visualize **ROIs**.



2D Liver Segmentation in Abdominal Computed Tomography Scans

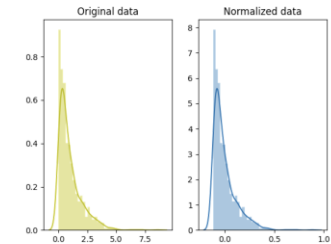
Histogram Equalization

Technique to improve **image contrast** by spreading out the most frequent **intensity values**. This process makes hidden features more visible, enhancing the quality of images.



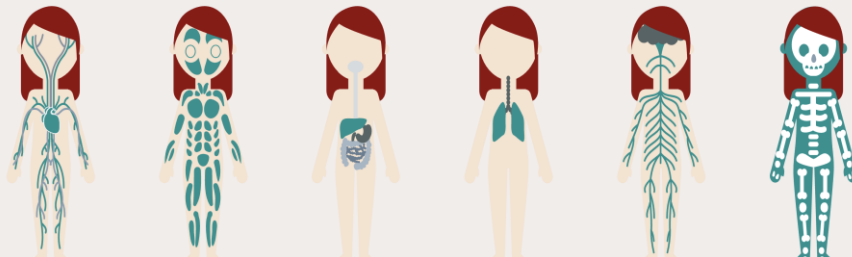
Data Normalization

Technique to scale the numerical data features to a uniform range, ensuring that all features have similar scales, transforming each feature so they fall within a common range, such as **0** to **1**, normalization prevents any single feature from dominating the learning process due to its larger scale.



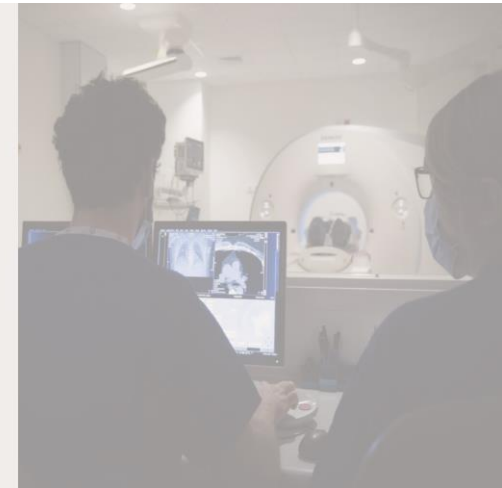
06 Conclusions

- Continuous improvement in **data management** and processing techniques is essential to maximize the value of hospital-acquired medical datasets.
- Standardized systems and ethical protocols** ensure high-quality, usable datasets for engineering applications.
- Collaboration between medical and technical experts** is crucial for successful medical data preparation and utilization.



Main challenges

- Ethical approval
- Data access and retrieval
- Data selection
- Data quality
- Data anonymization
- Data standardization



Appendix L – Liver Segmentation Application Packages

The table below present the packages needed to run the application.

Package	Version	Package	Version	Package	Version
contourpy	1.3.0	mpmath	1.3.0	PySide6_Essentials	6.7.2
cycler	0.12.1	networkx	3.2.1	python-dateutil	2.9.0.post0
darkdetect	0.7.1	numpy	1.23.5	PyWavelets	1.6.0
filelock	3.15.4	packaging	24.1	scikit-image	0.24.0
fonttools	4.53.1	pillow	10.4.0	scipy	1.9.3
fsspec	2024.6.1	pip	24.2	setuptools	72.1.0
imageio	2.35.1	pydicom	2.2.2	shiboken6	6.7.2
importlib_resources	6.4.4	pyparsing	3.1.4	six	1.16.0
intel-openmp	2023.2.0	PyQt6	6.2.2	sympy	1.13.2
jinja2	3.1.4	PyQt6-Qt6	6.7.2	tbb	2021.10.0
kiwisolver	1.4.5	PyQt6_sip	13.8.0	tifffile	2024.8.30
lazy_loader	0.4	pyqtdarktheme	2.1.0	torch	2.2.2
markupSafe	2.1.5	pyqtgraph	0.12.3	torchvision	0.17.2
matplotlib	3.9.2	PySide6	6.7.2	typing_extensions	4.12.2
mkl	2023.2.2	PySide6_Addons	6.7.2	wheel	0.43.0

Appendix M– Training and Validation Results

This appendix presents the quantitative results obtained by the segmentation models for the different experiments defined for the training and validation datasets.

• Models

This section compares the results obtained for the experiments for training and validation of the U-Net, ResUNet++, FCN and BerDiff models.

○ U-Net

Table 8.1 shows the results obtained from evaluating the training dataset for the five experiments carried out with the U-Net model for *Private Data Incrementalization* (E1-E5).

Table 8.1. Performance of the U-Net experiments on training dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E1	0.0470	0.9978	0.9617	0.9410	0.0704	0.0115	4.4712
E2	0.0297	0.9973	0.9798	0.9644	0.0370	0.0170	4.4708
E3	0.0396	0.9971	0.9710	0.9506	0.0535	0.0163	4.4714
E4	0.0391	0.9975	0.9701	0.9534	0.0544	0.0147	4.4713
E5	0.0426	0.9975	0.9671	0.9481	0.0606	0.0139	4.4712

Table 8.2 presents the evaluation outcomes of the validation dataset for the five experiments conducted using the U-Net model with *Private Data Incrementalization* (E1-E5).

Table 8.2. Performance of the U-Net experiments on validation dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E1	0.0655	0.9971	0.9395	0.9212	0.0941	0.0112	4.4708
E2	0.0925	0.9917	0.9353	0.9138	0.0974	0.0251	4.4720
E3	0.0774	0.9949	0.9426	0.9129	0.0961	0.0173	4.4717

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E4	0.0841	0.9945	0.9379	0.9121	0.1073	0.0177	4.4718
E5	0.0876	0.9948	0.9121	0.8973	0.1143	0.0162	4.4715

○ **ResUNet++**

Table 8.3 shows the results obtained from evaluating the training dataset for the five experiments carried out with the ResUNet++ model for *Private Data Incrementalization* (E6-E10).

Table 8.3. Performance of the ResUNet++ experiments on training dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E6	0.0504	0.9978	0.9616	0.9373	0.0745	0.0055	4.4721
E7	0.0282	0.9976	0.9830	0.9666	0.0332	0.0095	4.4721
E8	0.0347	0.9974	0.9782	0.9514	0.0444	0.0081	4.4721
E9	0.0297	0.9980	0.9789	0.9604	0.0406	0.0075	4.4721
E10	0.0580	0.9967	0.9628	0.9261	0.0740	0.0069	4.4721

Table 8.4 presents the evaluation outcomes of the validation dataset for the five experiments conducted using the ResUNet++ model with *Private Data Incrementalization* (E6-E10).

Table 8.4. Performance of the ResUNet++ experiments on validation dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E6	0.0752	0.9968	0.9379	0.9164	0.1033	0.0056	4.4721
E7	0.0919	0.9929	0.9127	0.9442	0.0866	0.0103	4.4721
E8	0.0893	0.9958	0.9536	0.9211	0.0834	0.0080	4.4721
E9	0.0776	0.9952	0.9545	0.9222	0.0951	0.0076	4.4721

2D Liver Segmentation in Abdominal Computed Tomography Scans

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E10	0.0849	0.9956	0.9409	0.8935	0.1042	0.0068	4.4721

○ FCN

Table 8.5. Performance of the FCN experiments on training dataset. shows the results obtained from evaluating the training dataset for the 5 experiments carried out with the FCN model for *Private Data Incrementalization* (E11-E15).

Table 8.5. Performance of the FCN experiments on training dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E11	0.0788	0.9964	0.9395	0.9032	0.1107	0.0149	4.4720
E12	0.0719	0.9937	0.9565	0.9164	0.0828	0.0272	4.4719
E13	0.1424	0.9898	0.9148	0.8461	0.1670	0.0301	4.4721
E14	0.0517	0.9967	0.9644	0.9324	0.0684	0.0170	4.4720
E15	0.0499	0.9971	0.9639	0.9359	0.0685	0.0154	4.4719

Table 8.6 presents the evaluation outcomes of the validation dataset for the five experiments conducted using the FCN model with *Private Data Incrementalization* (E11-E15).

Table 8.6. Performance of the FCN experiments on validation dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E11	0.0974	0.9957	0.9234	0.8869	0.1360	0.0151	4.4721
E12	0.1652	0.9853	0.9112	0.8317	0.1678	0.0370	4.4721
E13	0.2064	0.9855	0.8823	0.7804	0.2395	0.0318	4.4721
E14	0.1197	0.9918	0.9115	0.8767	0.1433	0.0232	4.4721
E15	0.1216	0.9924	0.9091	0.8636	0.1453	0.0194	4.4721

○ **BerDiff**

Table 8.7 shows the results obtained from evaluating the training dataset for the 5 experiments carried out with the BerDiff model for *Private Data Incrementalization* (E16-E20).

Table 8.7. Performance of the BerDiff experiments on training dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E16	0.3303	0.9839	0.6813	0.6574	0.4102	0.0175	4.4721
E17	0.4281	0.9578	0.6758	0.6204	0.4358	0.0408	4.4721
E18	0.4137	0.9650	0.6621	0.6236	0.4423	0.0342	4.4721
E19	0.4204	0.9675	0.6406	0.6083	0.4601	0.0317	4.4721
E20	0.4032	0.9716	0.6399	0.6206	0.4540	0.0277	4.4721

Table 8.8 presents the evaluation outcomes of the validation dataset for the five experiments conducted using the BerDiff model with *Private Data Incrementalization* (E16-E20).

Table 8.8. Performance of the BerDiff experiments on validation dataset.

Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
E16	0.3138	0.9851	0.7005	0.6639	0.3921	0.0163	4.4721
E17	0.4735	0.9503	0.6618	0.5794	0.4664	0.0446	4.4721
E18	0.4360	0.9634	0.6236	0.6017	0.4630	0.0361	4.4721
E19	0.4362	0.9646	0.6386	0.6208	0.4606	0.0327	4.4721
E20	0.4043	0.9715	0.6419	0.6207	0.4456	0.0270	4.4721

● **Datasets**

This section compares the results obtained for the experiments with each training and validation dataset for the datasets from 1 to 5, for the U-Net, ResUNet++, FCN and BerDiff models.

2D Liver Segmentation in Abdominal Computed Tomography Scans

○ Dataset 1

Table 8.9 presents the results of experiments with dataset 1 during training.

Table 8.9. Comparison of the performance from each model on training dataset 1.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E1	0.0470	0.9978	0.9617	0.9410	0.0704	0.0115	4.4712
ResUNet++	E6	0.0504	0.9978	0.9616	0.9373	0.0745	0.0055	4.4721
FCN	E11	0.0788	0.9964	0.9395	0.9032	0.1107	0.0149	4.4720
BerDiff	E16	0.3303	0.9839	0.6813	0.6574	0.4102	0.0175	4.4721

Table 8.10 compares the results of experiments with dataset 1 during validation.

Table 8.10. Comparison of the performance from each model on validation dataset 1.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E1	0.0655	0.9971	0.9395	0.9212	0.0941	0.0112	4.4708
ResUNet++	E6	0.0752	0.9968	0.9379	0.9164	0.1033	0.0056	4.4721
FCN	E11	0.0974	0.9957	0.9234	0.8869	0.1360	0.0151	4.4721
BerDiff	E16	0.3138	0.9851	0.7005	0.6639	0.3921	0.0163	4.4721

○ Dataset 2

Table 8.11 shows the results of experiments with dataset 2 during training.

Table 8.11. Comparison of the performance from each model on training dataset 2.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E2	0.0297	0.9973	0.9798	0.9644	0.0370	0.0170	4.4708
ResUNet++	E7	0.0282	0.9976	0.9830	0.9666	0.0332	0.0095	4.4721
FCN	E12	0.0719	0.9937	0.9565	0.9164	0.0828	0.0272	4.4719
BerDiff	E17	0.4281	0.9578	0.6758	0.6204	0.4358	0.0408	4.4721

Table 8.12 compares the results of experiments with dataset 2 during validation.

Table 8.12. Comparison of the performance from each model on validation dataset 2.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E2	0.0925	0.9917	0.9353	0.9138	0.0974	0.0251	4.4720
ResUNet ++	E7	0.0919	0.9929	0.9127	0.9442	0.0866	0.0103	4.4721
FCN	E12	0.1652	0.9853	0.9112	0.8317	0.1678	0.0370	4.4721
BerDiff	E17	0.4735	0.9503	0.6618	0.5794	0.4664	0.0446	4.4721

○ Dataset 3

Table 8.13 presents the results of experiments with dataset 3 during training.

Table 8.13. Comparison of the performance from each model on training dataset 3.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E3	0.0396	0.9971	0.9710	0.9506	0.0535	0.0163	4.4714
ResUNet ++	E8	0.0347	0.9974	0.9782	0.9514	0.0444	0.0081	4.4721
FCN	E13	0.1424	0.9898	0.9148	0.8461	0.1670	0.0301	4.4721
BerDiff	E18	0.4137	0.9650	0.6621	0.6236	0.4423	0.0342	4.4721

Table 8.14 compares the results of experiments with dataset 3 during validation.

Table 8.14. Comparison of the performance from each model on validation dataset 3.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E3	0.0774	0.9949	0.9426	0.9129	0.0961	0.0173	4.4717
ResUNet ++	E8	0.0893	0.9958	0.9536	0.9211	0.0834	0.0080	4.4721
FCN	E13	0.2064	0.9855	0.8823	0.7804	0.2395	0.0318	4.4721
BerDiff	E18	0.4360	0.9634	0.6236	0.6017	0.4630	0.0361	4.4721

2D Liver Segmentation in Abdominal Computed Tomography Scans

○ Dataset 4

Table 8.15 contrasts the results of experiments with dataset 4 during training.

Table 8.15. Comparison of the performance from each model on training dataset 4.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E4	0.0391	0.9975	0.9701	0.9534	0.0544	0.0147	4.4713
ResUNet++	E9	0.0297	0.9980	0.9789	0.9604	0.0406	0.0075	4.4721
FCN	E14	0.0517	0.9967	0.9644	0.9324	0.0684	0.0170	4.4720
BerDiff	E19	0.4204	0.9675	0.6406	0.6083	0.4601	0.0317	4.4721

Table 8.16 compares the results of experiments with dataset 4 during validation.

Table 8.16. Comparison of the performance from each model on validation dataset 4.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E4	0.0841	0.9945	0.9379	0.9121	0.1073	0.0177	4.4718
ResUNet++	E9	0.0776	0.9952	0.9545	0.9222	0.0951	0.0076	4.4721
FCN	E14	0.1197	0.9918	0.9115	0.8767	0.1433	0.0232	4.4721
BerDiff	E19	0.4362	0.9646	0.6386	0.6208	0.4606	0.0327	4.4721

○ Dataset 5

Table 8.17 contrasts the results of experiments with dataset 5 during training.

Table 8.17. Comparison of the performance from each model on training dataset 5.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E5	0.0426	0.9975	0.9671	0.9481	0.0606	0.0139	4.4712
ResUNet++	E10	0.0580	0.9967	0.9628	0.9261	0.0740	0.0069	4.4721
FCN	E15	0.0499	0.9971	0.9639	0.9359	0.0685	0.0154	4.4719
BerDiff	E20	0.4032	0.9716	0.6399	0.6206	0.4540	0.0277	4.4721

Table 8.18 compares the results of experiments with dataset 5 during validation.

Table 8.18. Comparison of the performance from each model on validation dataset 5.

Model	Experiment	Loss	Accuracy	Precision	DSC	AOE	ASSD (mm)	MSSD (mm)
U-Net	E5	0.0876	0.9948	0.9121	0.8973	0.1143	0.0162	4.4715
ResUNet ++	E10	0.0849	0.9956	0.9409	0.8935	0.1042	0.0068	4.4721
FCN	E15	0.1216	0.9924	0.9091	0.8636	0.1453	0.0194	4.4721
BerDiff	E20	0.4043	0.9715	0.6419	0.6207	0.4456	0.0270	4.4721

Appendix N – Data Augmentation Techniques

Data augmentation is a technique used to increase the diversity and volume of the training dataset, which improves the robustness and generalization of ML models. It involves generating new training samples by applying various transformations to the original images. The data augmentation techniques that were initially thought to be used are presented in this appendix.

- **Rotation**

Rotation involved rotating the original images by a specified angle. This technique can be used to increase the variability of the dataset and to help the model learn invariant features to rotations. For instance, the original CT image and mask were rotated to create new training samples, ensuring that the model could recognize the anatomical structures regardless of their orientation (Figure 81) [135].

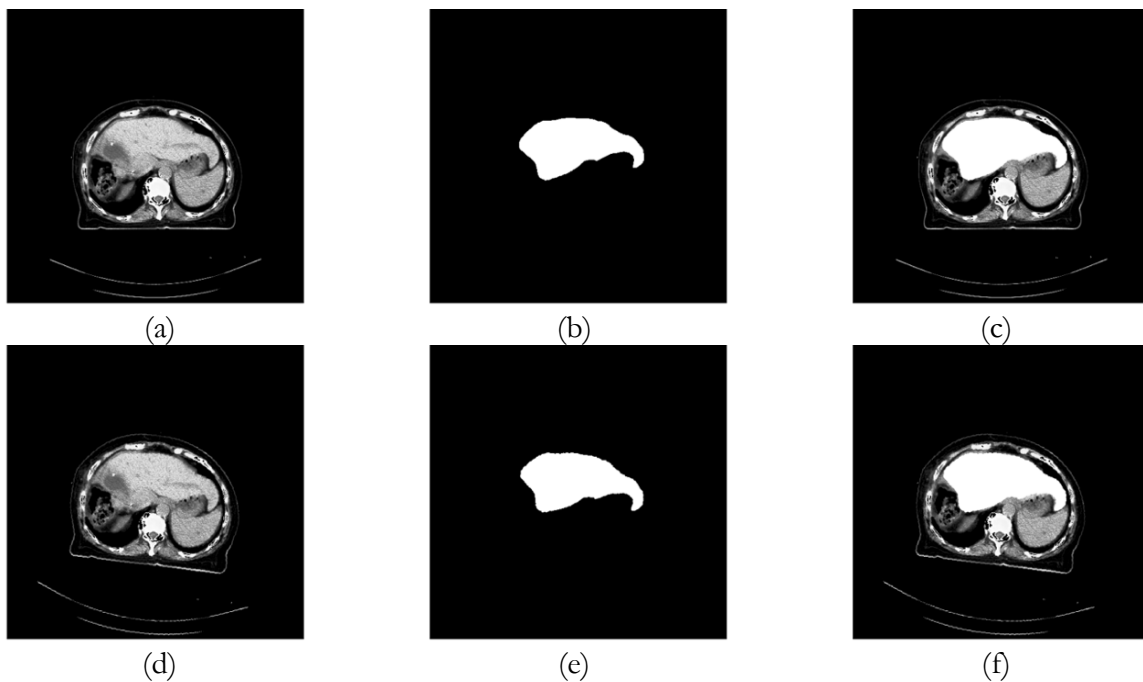


Figure 81. Rotation operation. (a) Original image. (b) Original mask. (c) Original image with mask overlay. (d) Rotated image. (e) Rotated mask. (f) Rotated image with mask overlay.

- **Reflection**

Reflection, also known as flipping, involved mirroring the images horizontally or vertically. This method was particularly useful in medical imaging to simulate different anatomical variations and to enhance the ability of the model to generalize across different patient orientations. By reflecting the original CT images and their corresponding masks, new augmented data were generated (Figure 82) [135].

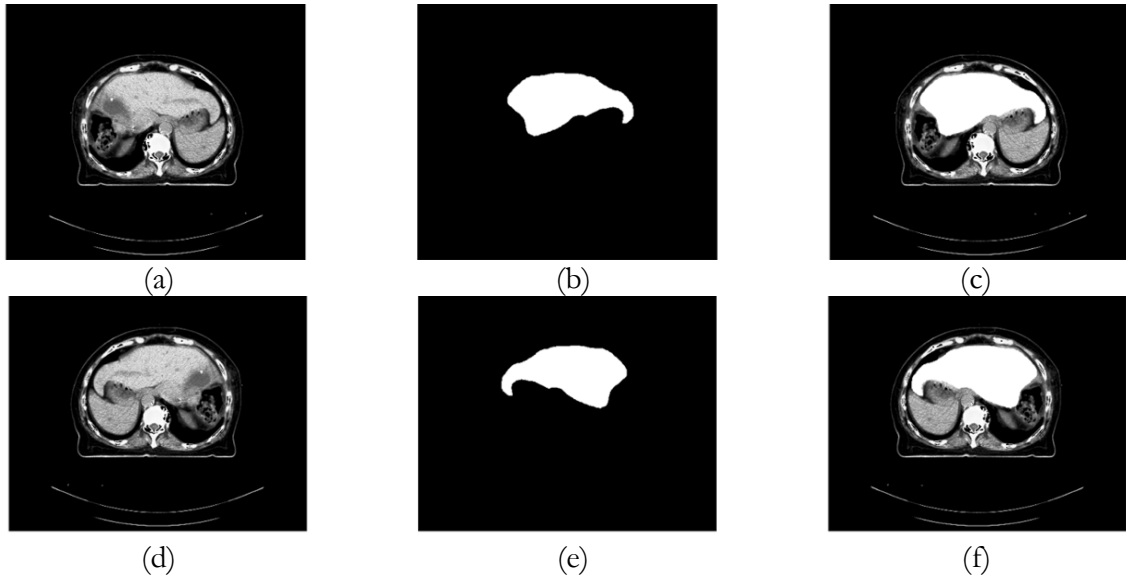


Figure 82. Reflection operation. (a) Original image. (b) Original mask. (c) Original image with mask overlay. (d) Reflected image. (e) Reflected mask. (f) Reflected image with mask overlay.

- **Combined Rotation and Reflection**

Combining rotation and reflection further increased the diversity of the dataset. By applying both transformations to the original images, the augmented dataset captured a wide range of variations, making the model more robust to different imaging conditions and anatomical variations. This combined approach supported the model to accurately interpret medical images despite differences in orientation and reflection (Figure 83).

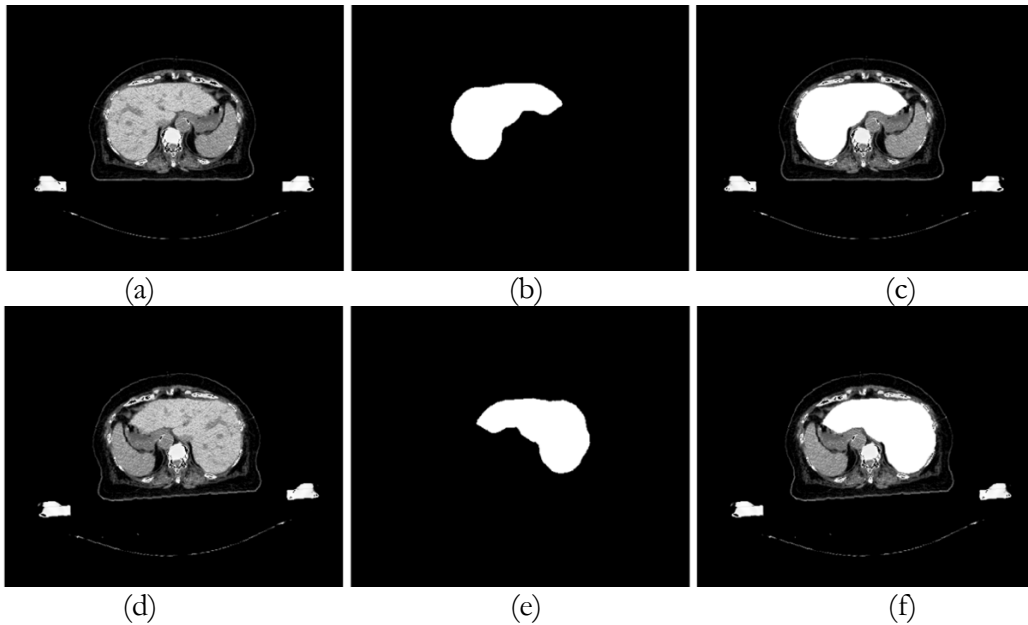


Figure 83. Combination of rotation and reflection operations. (a) Original image. (b) Original mask. (c) Original image with mask overlay. (d) Rotated and reflected image. (e) Rotated and reflected mask. (f) Rotated and reflected image with mask overlay.



**Instituto Superior
de Engenharia**

Politécnico de Coimbra