Maria Balaguer Montero

Deep Learning Tool for the Automatic Detection and Segmentation of Liver Tumors

MASTER'S THESIS

supervised by Dr. Raquel Pérez-López and Dr. Roser Sala-Llonch

Master's Degree in Biomedical Data Science

UNIVERSITAT ROVIRA i VIRGILI

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

UNIVERSITAT DE BARCELONA

UAB Universitat Autònoma de Barcelona

Universitat de Girona

Universitat de Lleida

UVIC UNIVERSITAT DE VIC UNIVERSITAT CENTRAL DE CATALUNYA

BIB BIOINFORMATICS BARCELONA

Barcelona, 7th of June of 2024

Dr. RAQUEL PÉREZ-LÓPEZ and Dr. ROSER SALA-LLONCH, certify that the student MARIA BALAGUER MONTERO has elaborated the work under their direction, and they authorize the presentation of this Master’s Thesis for its evaluation.

Advisors signature:

Roser

Dr. Raquel Pérez-López

Dr. Roser Sala-Llonch

ACKNOWLEDGMENTS

This project would not have been a success if I had not had the trust and support from people that have given their whole dedication through the development of this project, helping whenever I needed. I would like to express my sincere gratitude to all those people who have made it possible for me to be here today.

Firstly, I would like to thank my advisor Dr. Raquel Pérez-López, and also Dr. Marta Ligero and Adrià Marcos for their support and guidance given to me through this whole time. Their dedication and involvement during the whole evolution of the project has been key for its success. I would like to express my gratitude to my tutor too, Dr. Roser Sala-Llonch, who helped me find this project and has been keeping an eye on the progress of it, always available for advice and support.

I would also like to thank all the members of the Radiomics group at the Vall d'Hebron Institute of Oncology (VHIO), for their hospitality and advice at all times.

Last but not least, I would like to thank the unconditional support of my family and friends, not just for this project, but throughout the whole master. I could not have gotten this far without you by my side, thank you for believing in me. My most sincere thanks, always.

SUMMARY

In this project, we aim to develop a tool for the automatic detection and segmentation of liver tumors with existing Deep Learning models. To do so, we have collected 1516 computed tomography scans, coming from several institutions (open-source and private), accounting for a total of 4790 tumors. A review of the state-of-the-art methods has been done too, as well as a benchmarking comparison to our resulting model. Moreover, the impact in segmentation performance of several scan characteristics has been evaluated, together with a sub-study to assess the inter- and intra-radiologists’ variability. Lastly, the quantification of Total Tumor Volume has been tested as a prognostic biomarker. An extensive analysis of the results and their impact and implications on the clinical practice has also been discussed.

Keywords

Automatic Segmentation · Liver cancer · Deep Learning · Detection · Liver metastasis

LIST OF FIGURES

Figure 1. Workflow of the project 3

Figure 2. Global anatomical structure of the human liver and adjacent organs, and a close-up of its segments and lobes 6

Figure 3. Learning process of a neural network 10

Figure 4. Schematics of a complex network, undirected and weighted 12

Figure 5. Data structure of the scans included in the project....................17

Figure 6. Original U-Net architecture from 19

Figure 7. nnU-Net method configuration adapted from 20

Figure 8. TransUNet framework architecture from 21

Figure 9. All cases for the label association problem 22

Figure 10. Boxplot and violin plot of the segmentation performance....................25

Figure 11. Spider plot with all the models and all detection and segmentation metrics......27

Figure 12. Segmentation examples with the best model on external validation scans........28

Figure 13. Results of the benchmarking of the state-of-the-art model, in detection and segmentation 29

Figure 14. Effect of the CT density on the segmentation performance in both datasets.....30

Figure 15. Effect of the tumor volume on the segmentation performance in both approaches and datasets 31

Figure 16. Effect of the slice thickness on the segmentation performance in both approaches and datasets 32

Figure 17. Manual and Automatic contours performance comparison, patient-wise 33

Figure 18. Kaplan-Meier analysis on the test set using TTV quantification 35

Figure S1. Segmentation performance of the nnU-Net configurations, volume-wise..........49

Figure S2. Manual and Automatic contours performance comparison, tumor-wise 50

Figure S3. Kaplan-Meier analysis on the external validation using TTV quantification 51

LIST OF TABLES

Table 1. Population characteristics..17

Table 2. nnU-Net detection results for both approaches and datasets 26

Table 3. Best model delineation results in both approaches and datasets..........................26

Table 4. Manual and Automatic comparison, detection performance, patient-wise............33

Table 5. Manual and Automatic comparison, segmentation performance, patient-wise 34

Table 6. Manual and Automatic comparison, detection performance, tumor-wise.............34

Table 7. Manual and Automatic comparison, segmentation performance, tumor-wise.......35

Table S1. CT scan characteristics 52

Table S2. nnU-Net configurations segmentation results 53

Table S3. All models' detection results, both approaches and datasets 54

Table S4. All models' segmentation results, both approaches and datasets.......................54

Table S5. Evaluation of the benchmarking of the best model, detection 55

Table S6. Evaluation of the benchmarking of the best model, segmentation 55

GLOSSARY OF ABBREVIATIONS

AI	Artificial Intelligence
ARLD	Alcohol-related Liver Disease
BCE	Binary Cross Entropy
BTCV	Beyond The Cranial Vault
CAM	Class Activation Mapping
CNN	Convolutional Neural Networks
CRC	Colorectal Cancer
CRLM	Colorectal Liver Metastases
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DNN	Deep Neural Network
DSC	Dice Similarity Coefficient
EHR	Electronic Health Record
FDG	Fluorodeoxyglucose
GT	Ground Truth
HCC	Hepatocellular Carcinoma
HR	Hazard Ratio
HU	Hounsfield Units
IBSI	International Symposium on Biomedical Imaging
IHC	Intrahepatic Cholangiocarcinoma
IQR	Interquartile Range
JI	Jaccard Index
KiTS	Kidney Tumor Segmentation
LIME	Local Interpretable Model-agnostic Explanations
LiTS	Liver Tumor Segmentation
LR	Learning Rate
MICCAI	Medical Image Computing and Computer-Assisted Intervention
ML	Machine Learning
MLP	Multilayer Perceptron
MRI	Magnetic Resonance Imaging
MSD	Medical Segmentation Decathlon
NER	Named Entity Recognition
NIfTI	Neuroimaging Informatics Technology Initiative
NLP	Natural Language Processing
NN	Neural Network
nnU-Net	no new U-Net
OS	Overall Survival
PET	Positron Emission Tomography
RECIST	Response Evaluation Criteria In Solid Tumors
SGD	Stochastic Gradient Descent
SHAP	SHapley Additive exPlanations
TCIA	The Cancer Imaging Archive
TTV	Total Tumor Volume
VHIO	Vall d'Hebron Institute of Oncology
VHUH	Vall d'Hebron University Hospital
XAI	Explainable AI

TABLE OF CONTENTS

ACKNOWLEDGMENTS III
ABSTRACT IV
LIST OF FIGURES V
LIST OF TABLES VI
GLOSSARY OF ABBREVIATIONS VII

1. INTRODUCTION 1
- 1.1. Motivation 1
- 1.2. Objectives 2
- 1.3. Project setting 2
- 1.4. Report structure 3
- 1.5. Scope, assumptions and limitations 4
- 1.6. Location of the project 5

2. BACKGROUND 6
- 2.1. Liver 6
  - 2.1.1. Liver cancer 8
- 2.2. Deep Learning 9
  - 2.2.1. Neural Networks 10
  - 2.2.2. Complex Networks (Graphs) 11
- 2.3. Computed Tomography imaging 13

3. STATE-OF-THE-ART 14

4. METHODOLOGY 16
- 4.1. Data 16
- 4.2. Image pre-processing 18
  - 4.2.1. Non-liver tissue masking and CT scan cropping 18
- 4.3. DL architectures 19
  - 4.3.1. Vanilla U-Net 19
  - 4.3.2. nnU-Net 20
  - 4.3.3. TransUNet 21
- 4.4. Evaluation metrics 21
  - 4.4.1. Lesion association 22
  - 4.4.2. Detection metrics 23
  - 4.4.3. Segmentation metrics 24
- 4.5. Survival analysis 24

5. RESULTS 25
5.1. DL Methods Comparison 25
5.2. Model Benchmarking 29
5.3. Image variables impact on segmentation performance 30
5.3.1. CT density 30
5.3.2. Tumor volume 31
5.3.3. Slice Thickness 31
5.4. Inter- and intra-radiologist variability study 32
5.5. Prognostic biomarker 35
6. DISCUSSION 37
6.1. Technical discussion 37
6.2. Ethical-social impact, sustainability and diversity discussion 39
7. CONCLUSIONS 40
8. FUTURE WORK 41
9. REFERENCES 42
10. ANNEXES 49
10.1 Supplementary Figures 49
10.2 Supplementary Tables 52
10.3 Supplementary Equations 56

1. INTRODUCTION

1.1. Motivation

Cancer is the second top leading cause of death globally accounting for an estimated 10 million deaths, one in every six deaths in 2020 [1]. Due to its wide variety, biological complexity, and high heterogeneity, it still has no cure, although many efforts reside in the search for an effective treatment. This variability makes each type of cancer a new disease, presenting unique symptoms, diagnosis, treatments and outcomes.

In this context, liver cancer poses a significant health challenge. Primary liver cancers such as hepatocellular carcinoma (HCC) and cholangiocarcinoma are often diagnosed in advanced stages, offering limited treatment options and unfavorable prognosis. Additionally, the liver is a common site for metastases originating from other primary cancers, significantly impacting patient prognosis [2, 3]. Early and precise detection is crucial as it opens the possibility for localized and potentially curative treatments, thereby improving the overall outcome for the patients.

The evaluation of liver tumor burden is crucial at different stages of cancer treatment. For therapy decisions, especially when planning surgical interventions, accurately assessing the number and volume of liver tumors is critical [4]. This evaluation, typically performed on medical images such as Computed Tomography (CT), is essential for planning curative-intent surgeries where balancing tumor removal with preserving healthy liver tissue is critical. Currently, this task is carried out, either subjectively or in a more quantitative way, manually, which is not only time-consuming but also sensitive to variability between different observers and within the same observer [5, 6]. Furthermore, the need for a more comprehensive assessment of liver tumors extends to treatment monitoring in cancer patients. The evaluation of cancer volume changes throughout treatment on CT images, as opposed to relying solely on the maximum diameter of a few tumors (as defined by the standard Response Evaluation Criteria In Solid Tumors (RECIST) [7]), offers a more accurate response assessment and prediction of clinical outcomes [8–10].

Yet, manually delineating tumors for volume analysis (i.e., drawing tumor contours) poses practical challenges and often acts as a bottleneck in numerous research projects and clinical applications that involve volumetric assessment, as it is a complex and time-consuming task.

The motivation for this project arises from the need to find a better methodology that will reduce manual workload and variability in measurements, streamline the treatment planning process, and support a more accurate and robust response evaluation, while also enhancing accuracy in tumor detection and volume assessment. All this done automatically.

This study has been developed inside a research project and the present results will be submitted for publication in a scientific journal.

1.2. Objectives

The main objective of this project is to develop a tool for the automatic detection and segmentation of liver lesions, implementing several existing Deep Learning (DL) methods for biomedical image segmentation.

To achieve the primary goal, the following sub-studies were defined as secondary objectives:

- To do a proper **data collection** and organization of both medical images (CT scans) and clinical information (demographics and survival data), key for the correct development and evaluation of the project.
- To **benchmark** the results obtained by the different methods to the current state-of-the-art, publicly available automatic liver tumor segmentation tool, the top-scoring model of the Liver Tumor Segmentation (LiTS) challenge from the Medical Image Computing and Computer-Assisted Intervention society's (MICCAI).
- To study the effect of image-related variables (i.e. tumor size, cancer tissue density, CT slice thickness…) on the detection and segmentation performance.
- To compare the automatically generated contours to multiple manual ones and benchmark the obtained metrics with those derived from manual segmentations, assessing the inter- and intra-radiologists' variability in a sub-cohort of patients.
- To evaluate the total tumor burden of disease as a prognostic biomarker, representing an example of a potential clinical application.

1.3. Project setting

The methodology followed to develop this project is described in this section and summarized in Figure 1. The project methodology is divided into 3 separate sections, or workpackages, that are summarized below and explained in detail later:

1) The first module corresponds to the data collection and the pre-processing pipeline applied to the images obtained, necessary for the next step. The CT scans were acquired at several centers, and the manual ground truth tumor delineation was performed on each scan by an experienced radiologist. A publicly available model was used for the segmentation of the whole liver organ, and the mask was used to crop the images for scan size reduction purposes. During this step, demographic and survival analysis information was collected as well for all patients in the development and test datasets.

2) For the model development and evaluation, three different DL models were trained. Moreover, all methods were evaluated following 2 approaches: patient-wise evaluation (considering all the tumor burden, volumetric assessment) and tumor-wise evaluation (each individual lesion compared to its corresponding ground truth). To do

the second approach, graph network analysis was used to establish the correspondences among predicted and ground truth lesions.

3) Lastly, the applications include several practical uses of the tool. First, the best-performing model was compared to the state-of-the-art, publicly available liver tumor segmentation algorithm. Secondly, the impact of several scan characteristics on the model performance was studied. Then, a sub-cohort of the test set was re-segmented by three radiologists, enabling us to compare the detection and segmentation performance of the experts with those of the tool. Lastly, the Total Tumor Volume (TTV) quantification was assessed as a prognostic biomarker.

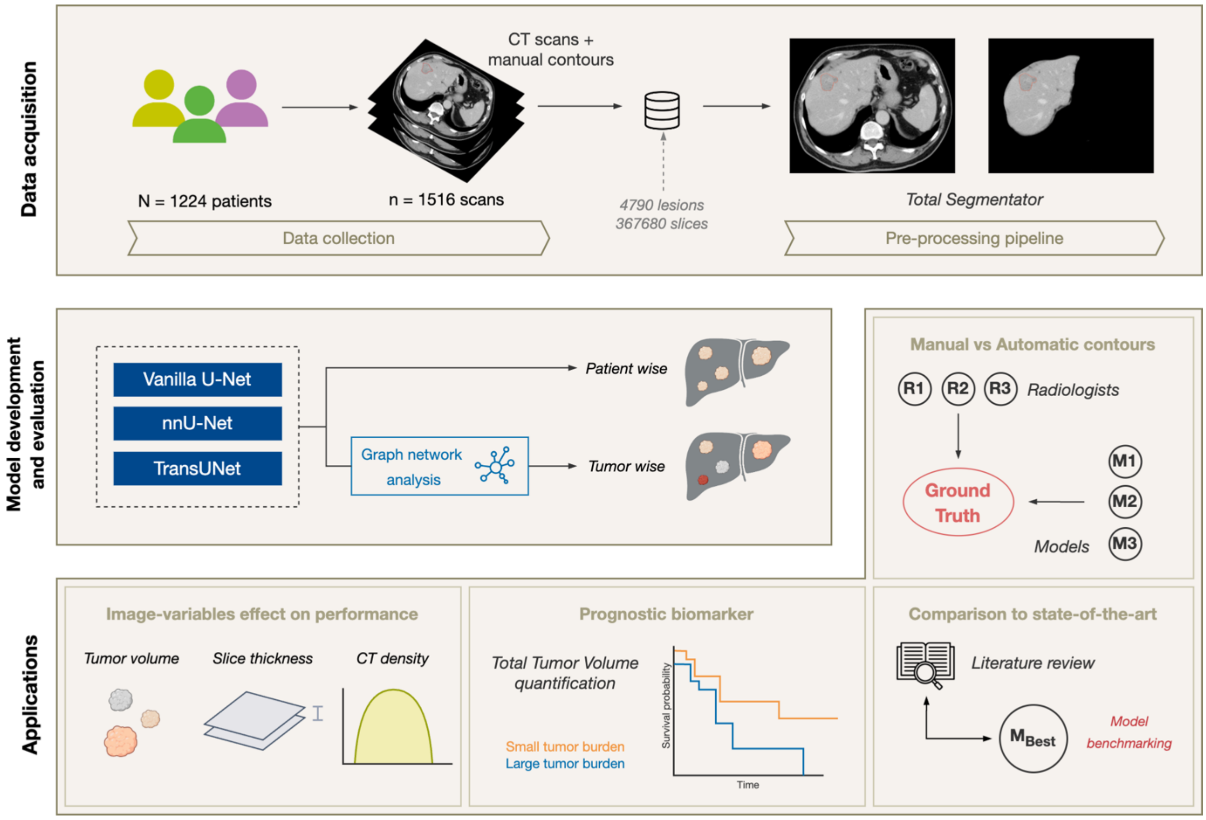

Figure 1. Workflow of the project

1.4. Report structure

The report is divided into 8 main sections. Firstly, an introduction of the chosen field is detailed, as well as the objectives and the scope and limitations that the project has. It is followed by a description of the general background surrounding the key terms of the project. On section 3, the state-of-the-art is analyzed in depth. Next, an extensive description of methodology followed is included in the design and development section. This is followed by exposing the results obtained using the methods just explained. On section 6, the discussion extracted from the results is presented, followed by the conclusions of the whole project. Finally, the future work to come for this project is proposed.

Additionally, a list of figures and tables, a glossary of abbreviations used in the project, and an abstract are included at the beginning of the project. Supplementary information is added at the end regarding, additional figures, tables and equations.

1.5. Scope, assumptions and limitations

The **scope** of the project includes the following steps:

- Literature review about the existing methodologies for biomedical imaging segmentation, including the Deep Learning architectures, performance results and participation in international challenges.
- Development of an image pre-processing pipeline to improve segmentation results.
- Comparison of DL algorithms for the detection and delineation of liver tumors.
- Comparison between manual and automatic segmentations for the assessment of inter- and intra-radiologists' variability and its comparison against the model performance.
- Implementation of the model to test Total Tumor Volume as a prognostic biomarker.

In order to be able to complete all the points that have just been defined as the scope of the project, some assumptions and limitations need to be taken into consideration. First the **assumptions** will be disclosed:

- The Ground Truth (GT) segmentations considered are perfect. Reality is, as will be proven later, the variability between manual contours exists, and even if we are considering one as an absolute truth, it may not always be the case.
- The criteria used for the ground truth segmentations was always the same, even if we know that it is not (limitation below).
- Our population is representative of all the different options of primary and metastatic liver cancer.
- Only the most recent information is needed, in such a fast-improving field. All the studies considered prior to this study are from the last 10-year period (2015-2024).

As for the **limitations**, they will be listed here and furtherly detailed on the discussion:

- Not all demographic and survival data is available. For external reasons, only 1 of the 3 datasets we have used for external validation had this information publicly available. Open-source datasets do not always contain the same type of information meaning that for some part of the analysis a subset of the scans had to be excluded. Therefore, a fully detailed representation of our cohort is not available.

- Related to one assumption above, the GT segmentations are not completely perfect, meaning that we are computing results with what may not be 100% accurate.
- Not all the external validation datasets have been segmented following the same criteria (RECIST); one of the cohorts has segmented tumors under the minimum diameter size of the RECIST guidelines (10 mm).

1.6. Location of the project

This project has been developed within the Radiomics group at Vall d'Hebron Institute of Oncology, working closely with Dr. Raquel Pérez-López (MD, PhD), who supervised this project, and Dr. Marta Ligero (PhD) and Adrià Marcos (BSc) who mentored me. This project is included in the frame of a collaborative research project with other groups from the VHIO institution and the development of a scientific article is being carried out in parallel for its submission to a scientific journal in the near future.

The Radiomics group oversaw the data analysis and model development that is presented in this project. This part of the project was developed mainly face-to-face at the Cellex and Saturnino centers in Barcelona, but there have been periods that had to be done remotely using the secure VHIO's VPN to ensure data protection. All the data analysis and training was performed using the Radiomics cluster integrated by seven machines with Intel i9-10980XE 128GB RAM and dual NVIDIA GEFORCE RTX3090 with 24GB RAM with Ubuntu Linux.