POLITECNICO DI TORINO
Repository ISTITUZIONALE

# A Deep Learning model to segment liver metastases on CT images acquired at different time-points during chemotherapy

*Original*
A Deep Learning model to segment liver metastases on CT images acquired at different time-points during chemotherapy / Defeudis, Arianna; Panic, Jovana; Guzzinati, Walter; Pusceddu, Laura; Vassallo, Lorenzo; Regge, Daniele; Giannini, Valentina. - ELETTRONICO. - (2022), pp. 1-5. (Intervento presentato al convegno 17th IEEE International Symposium on Medical Measurements and Applications (IEEE MeMeA 2022) tenutosi a Giardini Naxos - Taormina, Messina (Italy) nel 22 June - 24 June 2022) [10.1109/MeMeA54994.2022.9856589].

*Availability:*
This version is available at: 11583/2971148 since: 2022-09-09T08:56:49Z

*Publisher:*
Institute of Electrical and Electronics Engineers Inc.

*Published*
DOI:10.1109/MeMeA54994.2022.9856589

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

12 October 2024

# A Deep Learning model to segment liver metastases on CT images acquired at different time-points during chemotherapy


Arianna Defeudis*
*Dept. of Surgical Science*
University of Turin and
Candiolo Cancer Institute, FPO-IRCCS
Torino, Italy
arianna.defeudis@unito.it

Jovana Panic
*Dept. of Electronics and Telecommunications*
Polytechnic of Turin
Torino, Italy
jovana.panic@polito.it

Walter Guzzinati
*Dept. of Electronics and Telecommunications*
Polytechnic of Turin
Torino, Italy
walterguzzinati@gmail.com

Laura Pusceddu
*Radiology Unit*
Candiolo Cancer Institute, FPO-RCCS
Candiolo, Italy
laura.pscd@gmail.com

Lorenzo Vassallo
*Radiology Unit*
SS Annunziata Savigliano Hospital,
Savigliano, Italy
lorenzovassallo1987@gmail.com

Daniele Regge
*Dept. of Surgical Science*
University of Turin and
Candiolo Cancer Institute, FPO-IRCCS
Torino, Italy
daniele.regge@unito.it

Valentina Giannini
*Dept. of Surgical Science*
University of Turin and
Candiolo Cancer Institute, FPO-IRCCS
Torino, Italy
valentina.giannini@unito.it



***Abstract*—The aim of this study is to present a fully automatic deep learning algorithm to segment liver Colorectal cancer metastases (LmCRC) on CT images, based on a U-Net structure, comparing net with an without the transfer learning approach. This is a bi-centric study, enrolling patients who underwent CT exam before (baseline) and after first-line therapy (TP1). Patients were divided into training (using a portion of baseline sequences from both centers), to train the DL model, and two validation sets: one with baseline (valB), and one with TP1 (valTP1) sequences. The reference standard for the automatic segmentations was defined by the manual segmentations performed by an experienced radiologist on the portal phase of the baseline and TP1 CT exam. The best performing model obtained Dice Similarity Coefficient (DSC) of 0.68±0.24, Precision (Pr) of 0.74±0.27, Recall (Re) of 0.73±0.26, Detection Rate (DR) of 93% on the valB, and DSC of 0.61±0.28, Pr of 0.68±0.31, Re of 0.65±0.29 and DR of 88% on the valTP1. These encouraging results, if confirmed on larger dataset, might provide a reliable and robust tool that can be used as first step of future radiomics analyses aimed at predicting response to therapy, improving the management of LmCRC patients.**



***Keywords — liver metastases, U-Net, automatic segmentation, CT imaging, Colorectal cancer, multicentric study***


## I. Introduction

Colorectal cancer (CRC) is a malignant tumor arising from the inner wall of the colon and/or the rectum [1]. Despite improvements in screening methodologies, approximately 20% of patients are diagnosed with metastatic CRC (mCRC), which carries a 14% 5-year survival rate [2]. Due to the rich portal venous and arterial blood supply, the liver is the most frequent site of mCRC, in fact up to 30–50% of patients develop hepatic mCRC (lmCRC) during their disease course [3], [4]. The recommended treatment for lmCRC is liver resection, characterized by a 5-year survival of 20–50% of cases [3]. For patients with unresectable lmCRC, median survival has significantly increased up to 30 months due to advances in targeted therapy management in the first-line treatment [2]. Unfortunately, only 60% of patients achieve partial or complete response with first-line therapy [5].

In the last few years, Artificial Intelligence (AI) has shown promising results in the development of radiomics models, since they have recently drawn considerable interest as a potential predictive tool for treatment outcomes, allowing more personalized therapies [5]–[10]. The translation of these approaches into clinical practice is still limited since most of the studies rely on manual or semi-automatic segmentation methods, which are: time-consuming and lead to high inter-reader variability [11] . More and more efforts have been made toward the development of Deep Learning (DL) segmentation models, since they have shown significant improvement in image classification prediction and recognition tasks [12], especially with the introduction of Fully Convolutional Networks (FCNs), specifically the U-Net [12]–[16]. Moreover, the Transfer Learning (TL), an innovative training approach where a DL model developed for a specific task is used as the starting point for a DL model on a new task [17], has gained more interest thanks to its remarkable results.

To the best of our knowledge, only a few studies have presented DL-based segmentation methods of the liver lesions [18]–[22]. Most of them are developed on publicly available datasets of liver cancers, which pathologically differ from lmCRC.

The aim of this study was the development of a U-Net-based method for the automatic segmentation of the

lmCRC on Computer Tomography (CT) images before (Baseline) and after first-line treatment (TP1), acquired by two different centers. As a secondary aim, we compared different training approaches (with and without TL), and we analyzed the validation performances on both Baseline and TP1 CT sequences.

## II. Materials and Methods

### A. Dataset

The dataset consists of CT images of the abdomen from two centers: Candiolo Cancer Institute FPO-IRCCS (center A) and ASST Grande Ospedale Metropolitano Niguarda hospital (center B) of Milan. All patients have a a stage IV CRC having at least one measurable lmCRC as defined by the RECIST 1.1 Criteria (greater diameter ≥ 10 mm) and underwent a CT exam with contrast injection within 2 weeks from the start of the first-line therapy (baseline), and a CT exam after 3 months (TP1). A total of 86 baseline images (Center A=26 and Center B=60) and 79 TP1 images (Cener A=24, Center B=55) were collected.

We excluded from the analysis LmCRC that either were confluent or subdiaphragmatic or contained large vessels or were difficult to measure in the subsequent exams. Once all mts were segmented, their longest diameter was measured at baseline and TP1.

Patients were divided into training (using a portion of baseline sequences from both centers), to train the DL model, and two validation sets: one with baseline (*valB*), and one with TP1 (*valTP1*) sequences.

The study was approved by the institutional review boards (IRBs) in each institution; signed informed consent to use and analyze imaging data was obtained from all participants before entering the study.

### B. Liver Segmentation

We internally developed an automatic DL-based segmentation method of the liver, to highlight the area of interest for the subsequent analysis. For the development, we selected 40 CT sequences, chosen randomly (20 per center). In particular, 32 patients were included in the training set, the remaining 8 in the validation set.

### C. Reference Standard

For the liver segmentation model, an expert radiologist, with 5 years of experience in reading CT images, segmented the 40 liver volumes and these masks have been used as reference standard. For the lmCRC segmentation system, the same radiologist segmented lmCRC's volumes on the portal phase of the baseline and TP1 CT exam, and as before, these masks were used as reference standard. For each patient, a maximum number of 10 lmCRC were segmented. All manual segmentation were carried on using theITK-Snap software.

### D. PreProcessing

The pre-processing phase consists of three steps: the resize of the images, the liver segmentation, and the image standardization.

First, all CT sequences were resized from the original dimension (512x512 pixels), thus obtaining 256x256 images, to reduce the computational complexity and time. For the second step, the DL automatic method segmented the liver volumes. Once obtained the liver mask, the CT sequences were normalized to account the variabilities due to the different acquisition protocols of the centers. In particular, all sequences underwent the z-score standardization:

$$\mathrm{img}_{\mathrm{std}} = \frac{\mathrm{img} - \mu}{\sigma}, \quad (1)$$

where μ is the mean and σ is the standard deviation of the whole liver volume of the patient, respectively.

### E. U-Net

The U-Net is characterized by two components: the encoder, which extracts the features from the input image, and the decoder, which localizes the information related to the features, allowing to obtain a probability score map, classifying each pixel instead of the whole image [23]. The output of the model was binarized using the Otsu's Thresholding, obtaining the final prediction mask of LmCRC.

In this study we followed two training approaches:

- NO-TL

With this approach, all neurons' weights are initialized randomly from a normal distribution. More specifically, the U-Net structure is characterized by the encore's structure follows the typical Convolutional Neural Network (CNN) architecture with three Convolutional blocks, each characterized by two subsequent convolutional layers, which are all characterized by a 3x3 kernel and the Rectified Linear Unit (ReLU) activation function [24], followed by the max pooling layer. And the decoder is symmetrical to the contracting one. So, each convolutional block is characterized by two convolutional layers followed by the up-sampling layer, all of them with a 3x3 kernel and ReLU activation. The only exception is the output convolution layer, characterized by a 1x1 kernel and the sigmoid activation function [24]. The loss function used for the training is the Binary Crossentropy (2):

$$\mathrm{BC} = -\frac{1}{N}\sum_{i}^{N}\left(y_i \cdot \log\left(p(y_i)\right)\right) + (1 - y_i) \cdot \log\left(1 - p(y_i)\right). \quad (2)$$

- TL

With this approach, the U-Net's encoder is a known CNN called Residual neural network (ResNet) [25]. The ResNet-18 is an 18-layer deep CNN made up of several residual blocks. The residual blocks are characterized by the presence of skip connections that link the original input to the output of each convolutional block. For our purpose, we used the ResNet-18 with pre-trained weights on the ImageNet database [26]. The decoder is similar to the No-TL structure. For the training of the network, we used a custom loss function described by a previous study [14].

The networks learning was performed based on the training set. To avoid overfitting, the algorithm stopped the training if the loss value did not show improvement after 10 epochs. Moreover, to avoid learning stagnation, we reduced the learning rate by a factor of 2-10, if no improvement is seen for 5 epochs.

All models have been implemented on Python using the Tensorflow library, with the Adam optimizer [27] and a learning rate of 0.001, β1 of 0.9 and β2 of 0.999.

### F. Validation

The two developed networks were validated on both *valB* and *valTP1* sets. We also evaluated the performances of the best performing network related to the lmCRC dimensions, dividing between small (diameter < 20mm) and big (diameter > 20mm). We divided the *valTP1* into two subgroups: one characterized by the training set patients (TP1*tr*), the other with the remaining cases (TP1*tst*). We evaluated if there were differences in the networks' performances between TP1*tr* and TP1*tst*. The aim was to assess how the presence of the baseline in the training affected the result on the two TP1 validation subgroups.

### G. Statistical Analysis

Dice Similarity Coefficient (DSC), Precision (Pr) and Recall (Re) were used to evaluate the performance of the algorithm on LmCRC segmentation. The Dice similarity coefficient (3) relates the elements in common between two groups with respect to the total number of elements. Precision (4) is the portion of elements indicated as positive by the model that are truly positive. Recall (5) is the portion of the truly positive elements that have been correctly identified by the model.

$$\mathrm{DSC} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}, \quad (3)$$

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad (4)$$

$$\mathrm{Re} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \quad (5)$$

where TP, FP and FN are the number of true positives, false positives, and false negatives, respectively. We also evaluated the Detection Rate (DR), defined as the percentage of correctly detected LmCRC (DSC>0.2) over the total amount of LmCRC. To statistically compare the differences between baseline and TP1 validation sets, we perfomed the chi-squared (comparison of proportions) analysis.

## III. Results

### A. Dataset

For the development of the liver segmentation method, the training set included 40 patients (20 per center), while the validation set by 8 patients, as shown in fig. 1.a. For the development of the lmCRC segmentation method, the training set was characterized by 18 patients from center A and 22 from center B, *valB* by the remaining 8 and 38 baseline from center A and B, respectively. The *valTP1* is composed of 24 patients from center A and 55 from center B (fig. 1.b). Overall, there were 134, 271 and 331 LmCRCs in training, *valB* and *valTP1* sets, respectively.

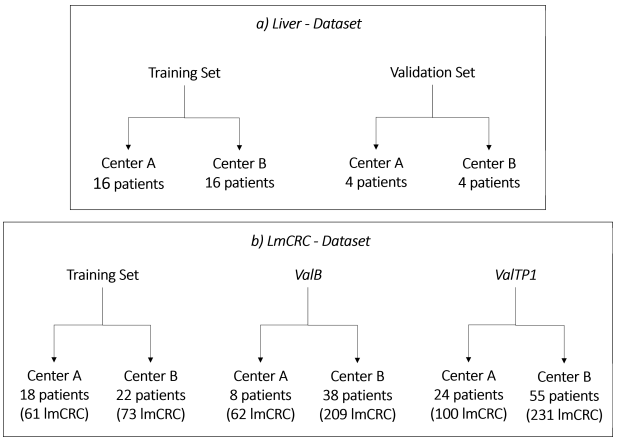


Figure 1: Datasets used for the development of the liver segmentation system (a), and the lmCRC segmentation networks (b).

### B. Liver segmentation

The performances of the DL model on the training set were DSC=0.94±0.02, Pr=0.92±0.01, Re=0.97±0.03, and DSC=0.94±0.01, Pr=0.90±0.01, Re=0.96±0.02 for the validation set.

### C. U-Net implementation and validation

Table I.a shows the results of the two developed U-Nets on the training set. Both networks show comparable performances, in particular DSC=0.80±0.19 (NoTL) vs 0.80±.024 (TL) (p-value > 0.05).

On *valB* set (Table I.b) the No-TL model reaches slightly higher performances. DSC=0.68±0.24 vs 0.64±0.25 (TL), Re=0.73±0.26 vs 0.64±0.27 (TL) and DR=93% vs 90% (TL), while the Pr is lower (0.74±0.27 No-TL vs 0.78±0.26 TL). 18/271 lmCRC were not detected by the No-TL model, and 27/271 by the TL one (p-value < 0.05). Both models did not detect the same 8 lesions, probably due to their small dimensions (average diameter of 16mm).

Table I.c shows the results on *valTP1* set. The No-TL model is confirmed to be the one with the highest performances with respect to the TL one. In particular, DSC=0.61±0.28 vs 0.55±0.29 (TL), and DR=88% vs 82% (TL), while the Pr is slightly lower. 40/331 LmCRC were not detected by the No-TL model, and 60/331 by the TL one (p-value < 0.05). We noticed that the 31 common errors were characterized by small dimensions (average diameter of 17mm) and different intensity values on the CT, which heavely affect the pixel intensity. After all these analyses, the best performing net is the No-TL U-Net model.

TABLE I. PERFORMANCES OF THE NETS

| U-Net | a) Training set | | | |
|---|---|---|---|---|
| | **DSC** (mean±std) | **Pr** (mean±std) | **Re** (mean±std) | **DR** (%) |
| ***No-TL*** | 0.80±0.19 | 0.81±0.20 | 0.85±0.16 | 97 (130/134) |
| ***TL*** | 0.80±0.24 | 0.81±0.25 | 0.82±0.23 | 94 (126/134) |
| | **b) ValB set** | | | |
| ***No-TL*** | 0.68±0.24 | 0.74±0.27 | 0.73±0.26 | 93 (253/271) |
| ***TL*** | 0.64±0.25 | 0.78±0.26 | 0.64±0.27 | 90 (244/271) |
| | **c) ValTP1 set** | | | |
| ***No-TL*** | 0.61±0.28 | 0.68±0.31 | 0.65±0.29 | 88 (291/331) |
| ***TL*** | 0.55±0.29 | 0.69±0.33 | 0.53±0.31 | 82 (270/331) |

No-TL: U-Net trained without Transfer Learning, TL: U-Net trained with Transfer Learning

Table II shows that the big lesions were more easily segmented in both *valB* and *valTP1*: the DSC values were comparable between the validation sets (DSC=0.73±0.19 and 0.72±0.24). Considering the small lmCRC, the NoTL model reached higher performances on *valB* than *valTP1* (DSC=0.71±0.30 vs 0.64±0.30).

TABLE II. RESULTS RELATED TO LMCRC DIMENSIONS

| NoTL | ***ValB*** | | | |
|---|---|---|---|---|
| | **DSC** (mean±std) | **Pr** (mean±std) | **Re** (mean±std) | **DR** (%) |
| ***small*** | 0.71±0.30 | 0.83±0.33 | 0.72±0.32 | 84 (75/89) |
| ***big*** | 0.73±0.19 | 0.78±0.24 | 0.80±0.19 | 98 (178/182) |
| | ***ValTP1*** | | | |
| ***small*** | 0.64±0.30 | 0.80±0.35 | 0.67±0.33 | 81 (116/143) |
| ***big*** | 0.72±0.24 | 0.77±0.28 | 0.78±0.22 | 93 (175/188) |

No-TL: U-Net trained without Transfer Learning, TL: U-Net trained with Transfer Learning

No statistically significant differences have been evaluated between the model's performances on *valB* and *valTP1*.

For the analysis related to how the baseline affects the TP1 results, the TP1*tr* is characterized by 35 patients (102 LmCRC), while TP1*tst* by 44 (229 LmCRC). Figure 2 shows the p-values analysis related to both TL and NoTL, comparing TP1*tr* (in blue) and TP1*tst* (in orange). In particular, there are no statistically significant differences between the two subgroups, considering both models. This means that the baseline is not helpful for the model to correctly segment the lmCRC on the TP1 sequence, considering the same group of patients. An example of morphological variability between baseline and TP1 is shown in figure 3. The first-line therapy affected the dimensions and shapes of two lmCRC (purple and yellow), and the internal textures of the third (orange), i.e.

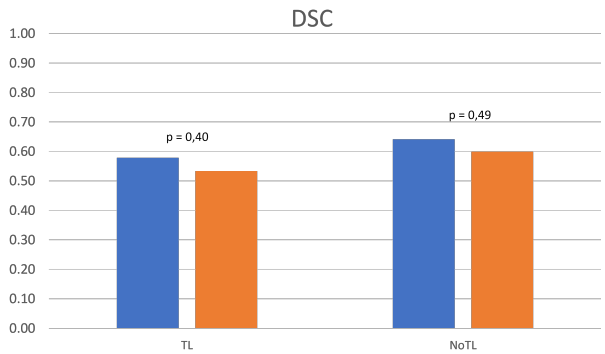


Figure 2: Comparison of TL and NoTL performances related to the DSC on TP1*tr* (in blue) and TP1*tst* (in orange). In both cases there are no statistical differences, since the p-values > 0.05.

presence of white areas.

Figure 4 shows three examples of correctly and wrongly detected LmCRC by the two models. In particular, fig. 4.a presents a *valB* patient, C1007, characterized by 9 LmCRC correclty segmented by both models (all DSC>0.2). Fig 4.b shows an example of TP1 validation patient, C2003, where three out of four LmCRC were correctly detected (DSC>0.2), while the remaining one was missed (DSC<0.2) by both models. The last example, fig. 4.c shows an example where the TL U-Net missed all four LmCRC, while the No-TL was able

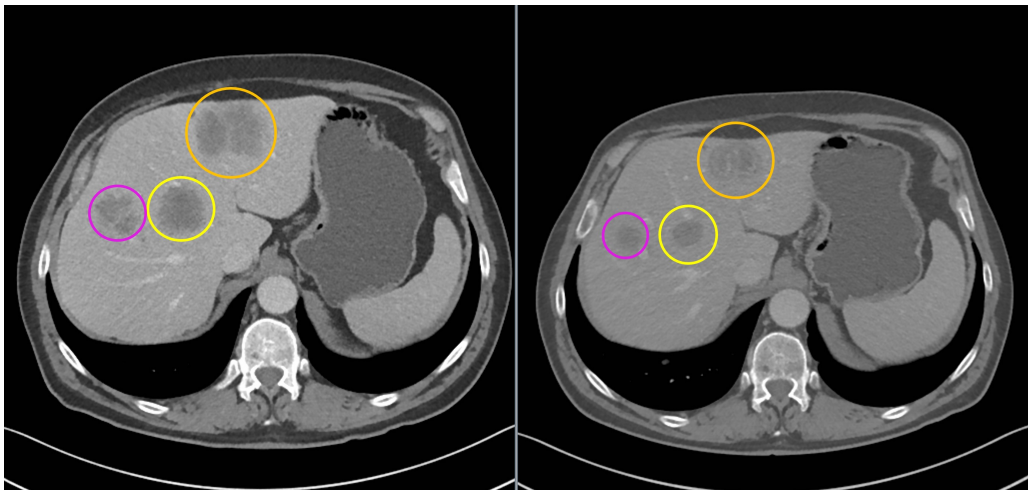

Figure 3: Comparison of the baseline (left) and TP1(right) of the C1005 patient. The three circles delimit the LmCRC. It is possible to notice that the pink and yellow ones present different shapes and dimension, while the orange is characterized by different internal texture (presence of white areas).

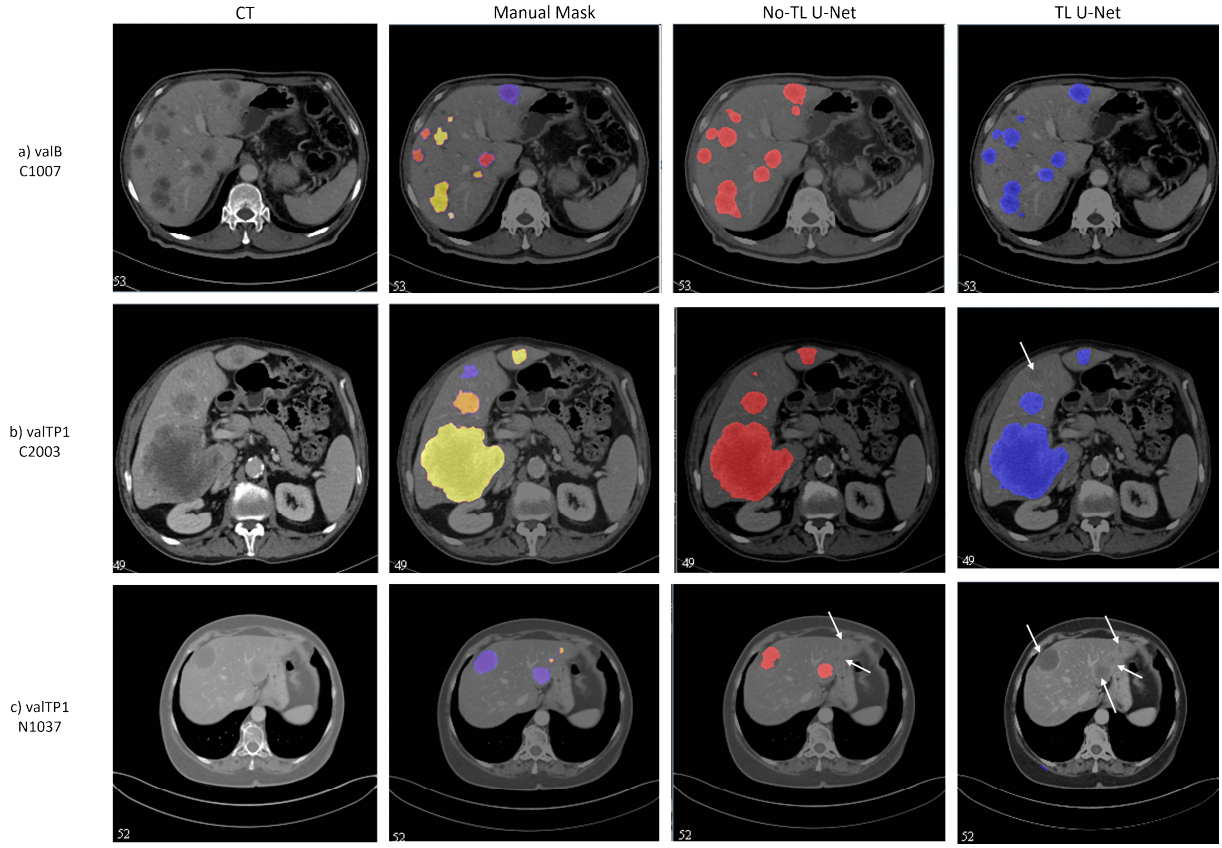


Figure 4: Examples of baseline (a) and TP1 validation (b-c) segmented by No-TL and TL U-Net models. a) shows an example of patient with 9 LmCRC correclty segmented by both models (all DSC>0.2). b) shows an example of a LmCRC missing segmentation (DSC<0.2) by both models. c) shows an example where the TL U-Net missed all 4 LmCRC, while the No-TL was able to correclty segment 2 of them. The white arrow point the missed LmCRC.

to correclty segment 2 of them.

## IV. Discussion

In this study, a U-Net-based algorithm was implemented for the automatic LmCRC segmentation, with promising results: precision of 0.74±0.27, and Detection Rate of 93% on the baseline validation set, and Pr=0.68±0.31, and DR=88% on the TP1 validation set. This model is characterized by a simple U-Net structure with three descending levels trained without the Transfer Learning approach. The computational efficiency, related to both structural complexity and timing, is one of the advantages.

Despite the success of Transfer Learning, in this work, the network trained with the TL approach did not reach adequate results (DSC=0.55±0.29, Pr=0.69±0.33, Re=0.53±0.31, DR=82%), probably due to the unsuitable pre-trained weights chosen.

As far as we know, there are few studies related to the automatic segmentation of liver metastases. Indeed, most of them are focused on liver and lesion segmentation, thanks to the publicly available multicentric database [20]–[22]. All of them present DL models with innovative structures for the segmentation of the liver, obtaining high performances. In particular, Khan et al. [21] and Li et al. [22]reached DSC of 97% and 96% for the liver segmentation, respectively. While Christ et al. [20] reached a DSC of 94% combining liver and lesion segmentation. All these performances are analogous with the DL system internally developed, reaching a DSC of 94% for the liver segmentation on the validation set.

Concerning our previous study [18], we improved the system performances thanks to: the increment of patients enrolled, the number of segmented LmCRC included, and the semplified net structure, which proved to be more suitable for the aim. Indeed, the No-TL U-Net reached a DSC of 0.61, while the previous net of 0.54.

In literature, the only comparable study on this topic is the one presented by Vorontsov et al. [19]. They presented a Fully Convolution Network for liver lesion detection and segmentation on CT examinations in patients with LmCRC, reaching DSC of 0.53 and 0.68 for small and big lesions, respectively. Our results using the No_TL net achieved higher DSC: 0.64 and 0.72 for small and big lesions on the validation set, respectively. Moreover, our model, validated on TP1, was able to generalize on morphologically different lmCRC, with respect to those used for the training.

Our study has also some limitations. First, the sample size is quite small in term of number of patients, however, the number of lmCRC is comparable with other studies. Moreover, the validation set is composed of patients acquired in the same centers that were used to train the network. We are planning to overcome this limitations by adding a larger number of patients from at least an additional center.

In conclusion, this work is one of the first to present a fully automatic segmentation algorithm of LmCRC, using

both baseline and TP1 scans. It can be an innovative starting point providing clinical support in the detection and segmentation of liver metastases, useful for future radiomics analysis aimed at personalizing treatments.

## ACKNOWLEDGMENT

This work was founded by FONDAZIONE AIRC under 5 per Mille 2018 – ID. 21091 program – P.I. Bardelli Alberto, G.L. Regge Daniele.

## REFERENCES

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer Statistics, 2021," *CA. Cancer J. Clin.*, vol. 71, no. 1, pp. 7–33, 2021.

[2] E. C. Smyth, V. Gambardella, A. Cervantes, and T. Fleitas, "Checkpoint inhibitors for gastroesophageal cancers: dissecting heterogeneity to better understand their role in first-line and adjuvant therapy," *Ann. Oncol.*, vol. 32, no. 5, pp. 590–599, 2021.

[3] D. I. Tsilimigras *et al.*, "Liver metastases," *Nat. Rev. Dis. Prim.*, vol. 7, no. 1, 2021.

[4] M. F. Müller, A. E. K. Ibrahim, and M. J. Arends, "Molecular pathological classification of colorectal cancer," *Virchows Arch.*, vol. 469, no. 2, pp. 125–134, 2016.

[5] V. Giannini *et al.*, "Delta-Radiomics Predicts Response to First-Line Oxaliplatin-Based Chemotherapy in Colorectal Cancer Patients with Liver Metastases," 2022.

[6] F. Graur and M. M. Buruian, "Pre-Treatment T2-WI Based Radiomics Features for," 2020.

[7] V. Giannini *et al.*, "Predicting locally advanced rectal cancer response to neoadjuvant therapy with 18 F-FDG PET and MRI radiomics features," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 46, no. 4, pp. 878–888, 2019.

[8] F. Coppola *et al.*, "Radiomics and Magnetic Resonance Imaging of Rectal Cancer: From Engineering to Clinical Practice," *Diagnostics*, vol. 11, no. 5, p. 756, 2021.

[9] A. Rocca *et al.*, "Early Diagnosis of Liver Metastases from Colorectal Cancer through CT Radiomics and Formal Methods : A Pilot Study," 2022.

[10] V. Giannini *et al.*, "Radiomics predicts response of individual HER2-amplified colorectal cancer liver metastases in patients treated with HER2-targeted therapy," *Int. J. Cancer*, vol. 147, no. 11, pp. 3215–3223, 2020.

[11] V. Granata *et al.*, "Radiomics in hepatic metastasis by colorectal cancer," *Infect. Agent. Cancer*, vol. 16, no. 1, pp. 1–9, 2021.

[12] J. Jin *et al.*, "Multiple U-Net-Based Automatic Segmentations and Radiomics Feature Stability on Ultrasound Images for Patients With Ovarian Cancer," *Front. Oncol.*, vol. 10, no. February, pp. 1–8, 2021.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[14] D. Barra *et al.*, "Deep learning model for automatic prostate segmentation on bicentric T2w images with and without endorectal coil," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2021, pp. 3370–3373, 2021.

[15] M. H. Soomro *et al.*, "Automated segmentation of colorectal tumor in 3D MRI Using 3D multiscale densely connected convolutional neural network," *J. Healthc. Eng.*, vol. 2019, 2019.

[16] X. Ma *et al.*, "U-Net based deep learning bladder segmentation in CT urography," *Med. Phys.*, vol. 46, no. 4, pp. 1752–1765, 2019.

[17] F. Navarro *et al.*, "Development and external validation of deep-learning-based tumor grading models in soft-tissue sarcoma patients using mr imaging," *Cancers (Basel).*, vol. 13, no. 12, pp. 1–14, 2021.

[18] V. Giannini *et al.*, "Deep learning to segment liver metastases on CT images: Impact on a radiomics method to predict response to chemotherapy," *IEEE Med. Meas. Appl. MeMeA 2020 - Conf. Proc.*, pp. 0–4, 2020.

[19] E. Vorontsov *et al.*, "Deep Learning for Automated Segmentation of Liver Lesions at CT in Patients with Colorectal Cancer Liver Metastases," *Radiol. Artif. Intell.*, vol. 1, no. 2, p. 180014, 2019.

[20] P. F. Christ *et al.*, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, pp. 415–423, 2016.

[21] R. A. Khan, Y. Luo, and F.-X. Wu, "RMS-UNet: Residual multi-scale UNet for liver and lesion segmentation," *Artif. Intell. Med.*, vol. 124, no. November 2021, p. 102231, 2022.

[22] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," *IEEE Trans. Med. Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[24] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," pp. 1–20, 2018.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.

[26] K. L. and L. F.-F. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, "ImageNet: A Large-Scale Hierarchical Image Database," *J. Vis.*, vol. 9, no. 8, pp. 1037–1037, 2010.

[27] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.