Politecnico di Torino

Politecnico di Torino

Corso di Laurea in Ingegneria Biomedica Magistrale
A.a. 2024/2025
Sessione di Laurea Dicembre 2025

Development of a deep learning method to detect and segment diverticula and colon cancer based on CT scans

Relatori:
Samanta Rosati
Valentina Giannini

Candidati:
Giovanni Sogos

ABSTRACT

The aim of this thesis is to develop a method to detect and segment diverticulitis and colon cancer in colon CT (computed tomography) images, starting by focusing in a deep learning approach by detecting the lesion through an automatic segmentation pipeline.
Diverticulitis is a very common disease in the western population, it usually occurs when the walls of the large intestine weaken and there is a formation of diverticula. Diverticula are growths of about 5-15 mm of diameter that can initially be asymptomatic, but they can inflame and eventually cause the diverticular disease. The symptoms of diverticulitis may be very similar to the ones of colon cancer, which has also a significant incidence in the western population. The standard clinical exams in these cases consist in invasive procedures such as colonoscopy and biopsy, so the Computer Tomography can be an excellent aid for the radiologist. With the CTC a 3-D reconstruction of the colon can be obtained, the physician can operate on it by highlighting the interested regions and decide whether or not to proceed with more specific and invasive exams. There can be challenging cases due to the thickening of the colonic walls, especially in the elderly ones, so morphologically the cancer and the diverticula may become even more difficult to distinguish. The image database is composed of a total of 151 CT scans, that were gathered between January 2010 and September 2023: 110 from the Luigi Sacco Hospital were manually segmented by an expert radiologist, while the remaining 41 volumes were manually segmented by a less experienced one. A first attempt of automatic segmentation was performed by using a standard UNet architecture, but since the results were far from being promising, the nnUNet, a self-configuring deep learning framework was implemented in order to automatically segment the interested lesions, using a 2-dimensional approach.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 THE COLON

The colon is a tube-shaped organ which is connected to the small intestine on the upper end and to the anus on the lower end. It is part of the long intestine and also constitute its longest component. [1]

The colon is anatomically divided into 4 main regions:

- Ascending colon: runs upwards on the right side of the body, from the end of the small intestine to the diaphragm.
- Transverse colon: passes through the abdominal cavity.
- Descending colon: runs downwards on the left side of the body.
- Sigmoid colon: an S-shaped segment which converges in the rectum.

The first three (ascending, transverse and descending) are specialized in the absorption of water and ions from the chyme. The sigmoid colon stores the leftover material in the lumen after the absorption has been completed. [2]

Even though the colon walls have the same structure as the rest of the gastrointestinal tract, the longitudinal muscle layer of the muscularis externa is not continuous, but comprised between three relatively tight bands named teniae coli, which run through the whole colon length.

When the chyme reaches the colon region, it contains only a few digestible nutrients, because the majority of them has been absorbed in the previous GI tracts. What is left is essentially represented by inorganic ions, non-digestible material, bacteria and water. The main function of the colon is to reduce the chyme volume by absorbing the leftover water in order to turn it into a semisolid material: the feces. These are stored in the colon until they're ready to be eliminated by the organism. [2]

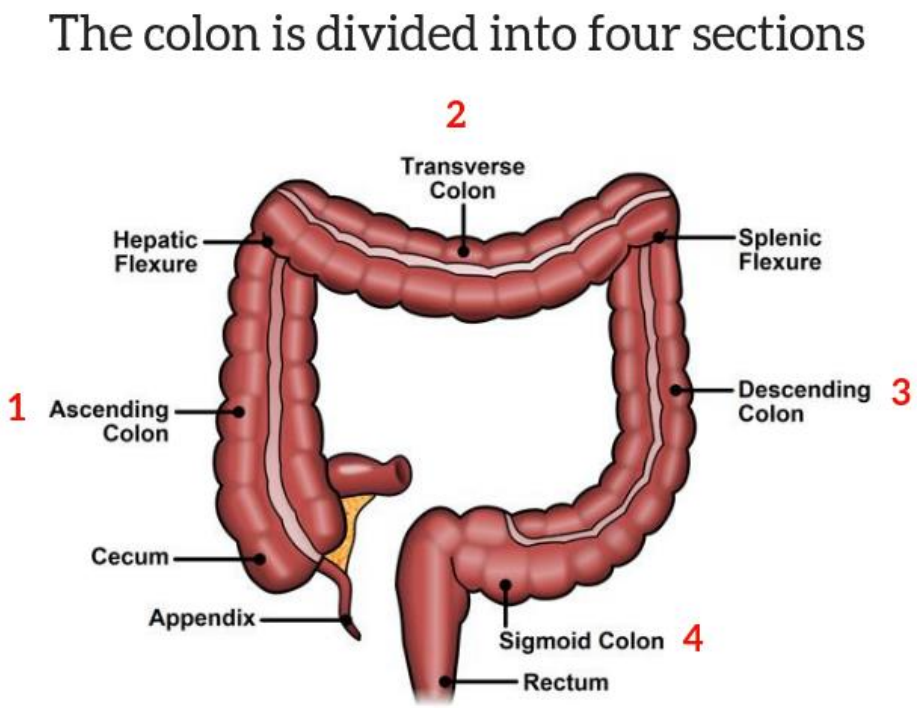

Figure 1: Main anatomic division of the colon [3]

1.1.2 DIVERTICULAR DISEASE

It has been observed that the diverticular disease hits a significant percentage of the adult western population over 40 years of age, from 10% to 30%, and approximately 40% of them is not aware of being affected by it. The diverticular disease happens when the large intestine walls weaken during time, probably due to an exasperation of the colonic walls derived by the excessive pressure after frequent constipation episodes. Diverticula are extroflexions of about 5-15 mm of diameter that appear in the mucosa and submucosa. The exact cause of the weakening of the colonic walls is not ascertained to this day, but in the majority of cases a genetic predisposition has been observed. The diet also plays a crucial role, in the pathogenesis of the diverticular disease are implied the diets with a small content of fibers and the typical occidental diets. Alterations of the gastrointestinal motility also contribute to the disease. [2]
The diverticula are observed mainly in the sigmoid colon, but can be found in other region of the intestine. Even though the diverticula may be asymptomatic, they can eventually cause an inflammatory disease called diverticulitis, which manifests with intense abdominal pain. The diverticula allow the chyme to settle for a long period of time, this causes the irritation of the colonic walls. Other symptoms correlated to the diverticular disease involve fever and shivering. The treatment for the diverticular disease involves anti-inflammatories, antibiotics and painkillers, surgical resection may be needed in very severe cases. [2]

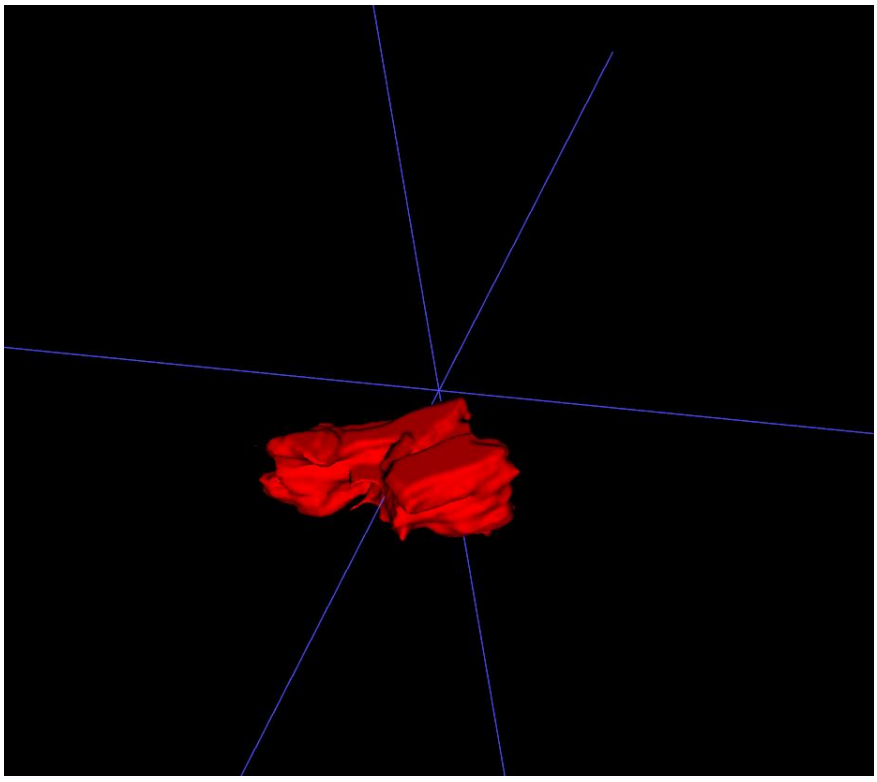

Figure 2: Example of Diverticula

1.1.3 COLON CANCER

Colon cancer is one of the tumors with the highest incidence worldwide after lung cancer, breast cancer for women and prostate cancer for men. In the United States, colon and rectal cancers, which are considered separately, are the second most common cause of mortality caused by cancer. [4] Even though it is one of the most diffused, colon cancer's incidence rates during the years have been reduced in Western countries thanks to the colonoscopy screening, but its incidence among younger subjects has been increasing.The main risk factors for developing a colon tumor have been identified in: age (the median is over 65 years of age), family history, adenomas on colonoscopy, inflammatory bowel disease (IBD) history and environmental factors (such as alcohol consumption, smoking and obesity). [4]

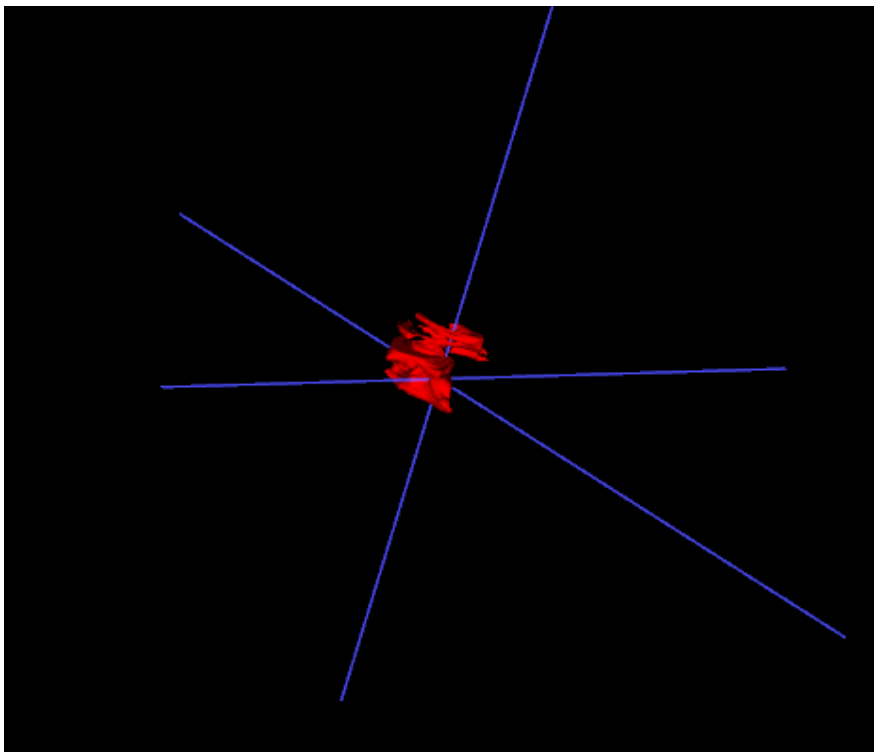

Figure 3: Example of colon cancer

1.2 IMAGE SEGMENTATION

1.2.1 U-NET

The UNet architecture can be summarized in a total of 23 convolutional layers, a contracting path on the left followed by an extensive path on the right. The standard convolutional network scheme can be observed on the left side: two 3x3 unpadded convolutions are applied repeatedly, each of them is followed by a rectified linear unit (called ReLu) and a 2x2 max pooling operation with stride 2 used for downsampling. Whenever a downsampling occurs, the number of channels dedicated to the features is duplicated. As the extensive path on the right side, we can observe that each block is initiated by an upsampling of the feature map with a 2x2 convolution afterwards, also referred as "up-convolution", which results in the halving of the feature channels. Right after that, a concatenation occurs with the correspondent cropped feature map on the left side, in this step the cropping is mandatory since each convolution causes a loss of border pixels. The remaining steps inside each block of the extensive path are two 3x3 convolutions, followed by a rectified linear unit (as the unpadded ones on the left side). As for the final layer, a 1x1 convolution is applied in order to map each 64-component feature vector to match the corresponding classes. [5]

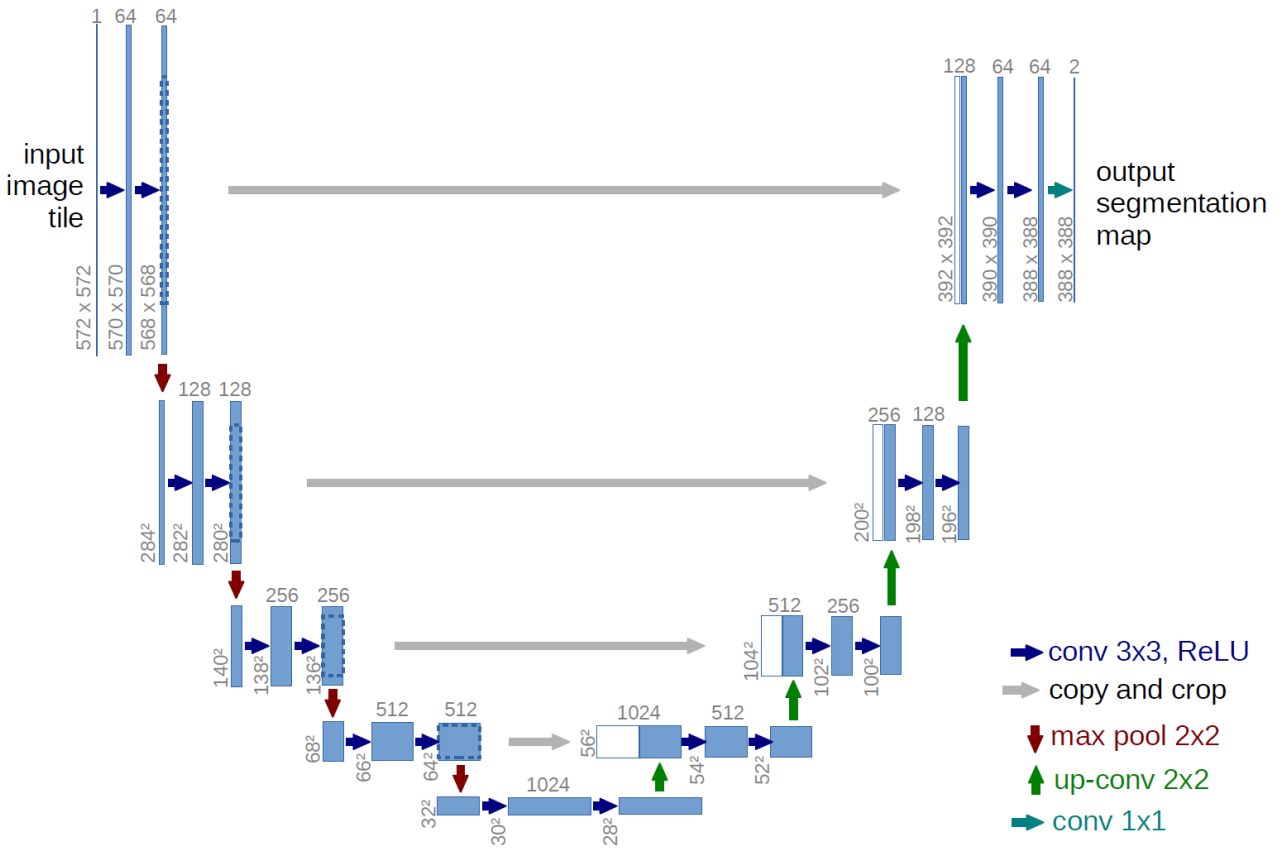

Figure 4: U-Net Architecture [5]

1.2.2 NNUNET

The nnUNet is a self-configuring segmentation method, by adapting to the given dataset to automatically generate a proper network architecture and training system. If needed, also pre-processing and post-processing tasks can be fulfilled thanks to this method. [6]
When a new dataset if given, its properties are extracted as a "dataset fingerprint", from image size, voxel spacing and class ratios. After that nnUNet automatically creates 3 different U-Net architectures, a 2-Dimensional and two 3-Dimensional configurations (one at full resolution, and the other at lower resolution), then after a cross-validation it empirically chooses the best performing one.[6]

1.3 THE AIM OF THE THESIS

The main focus of the work was to develop a deep learning segmentation method in order to detect and segment the diverticula and colon cancer in CTC images. Computed Tomography constitutes an optimal diagnostic element for physicians but it needs to be followed by more invasive procedures such as colonoscopy to confirm a diagnosis, so an instrument that could already detect the lesion prior to uncomfortable procedures for the patient would improve the procedure. The task presents itself as challenging since the diverticula are complicated to segment and morphologically are very similar to the surrounding tissues in CT scans, as for colon cancer it differentiates more since the mass has a more irregular form than the diverticula.

2 MATERIALS AND METHODS

2.1 IMAGE DATASET

The dataset is composed of a total of 151 CTC (Computed Tomography Colonography) volumes, the images are in the Nifti format.
The 151 patients that underwent CTC were 75 males and 76 females, with an age varying from 25 to 90 years old and a median age of 69.81.
As for the final diagnosis, 75 patients were diagnosed with diverticular disease while 76 were diagnosed with sigmoid colon cancer, and among these last ones 28 were diagnosed with both cancer and diverticular disease.
There are also 36 patients labeled as "challenging cases", 16 of these were finally diagnosed with cancer, while the other 20 with diverticulitis. The challenging attribute has been given to the cases in which the image features of the lesions were not classifiable as benignant or malignant.
CTC was performed positioning the patient in 3 different ways: supine (76), prone (38) and lateral decubitus (38). Each CTC slice is a 512x512 image, while the number of slices per volume varies for each patient. The resolution on the x and y axis vary from patient to patient.
The segmentation task was performed by a single radiologist with the an open-source software (ITK-SNAP). The areas where the diverticulitis and the colon cancer developed were identified starting from the CTC image and a segmentation mask was created by the respective radiologist. A summary of the final diagnosis was collected in an Excel file, along with the patients' sex, age and positioning during the screening. The challenging cases were also highlighted in the same spreadsheet. [7]

2.1.1 INCLUSION AND EXCLUSION CRITERIA

The main inclusion criteria for the multicenter study can be summarized as follows:

- Diseases that involved wall thickening of at least 5 mm observed in the rectum or sigmoid colon
- A final diagnosis obtained from surgery, histological or visual endoscopy which confirmed the presence of Colorectal Cancer or Diverticulitis.

On the other hand, the exclusion criteria were:

- Factors that compromise the measurement of the wall thickening such as not optimal bowel preparation and not sufficient colonic distension
- Final diagnosis of polyps or other colonic pathologies and patients excluded in the follow up. A total of 4 patients belonged to these categories and was excluded from the study. [7]

2.1.2 CTC PROTOCOL

The CTC exams were conducted in the two centers using different types of CT detectors:

- At Luigi Sacco Hospital, a 64-section multidetector CT scanner (Brilliance 64 from Philips Healthcare or a 16-section multidetector CT scanner (Brightspeed from General Electric Healthcare).
- At Spedali Civili Hospital, a 16-section multidetector CT scanner (Siemens Somatom Sensation 16 from Siemens Healthcare).

The patients who underwent the CTC had to attain to a bowel cleansing procedure 3 days before the exams that consisted of a sachet of mild laxative per osmosis (Movicol®) after the 3 main meals of the day. It was also asked to the patients to drink at least 1 L of water per day and to consume a low-fiber diet. 3 hours prior to the CTC the patients were administrated orally with 1 mL/kg of iodinated contrast material (Gastrografin®) diluted in 1 L of water for fecal marking.

Right before the scans, CO2 was insufflated in the colon via an automated insufflator (PROTOCO2L) by utilizing a flexible rectal catheter with a retention balloon. In order to facilitate the colonic distension, the patients also received intravenously 20 mg/ml of Hyoscine Butylbromide (Buscopan®), expect for those who could run into collateral effects. As a result of the protocol, the scanning performed in both centers were highly comparable. [7]

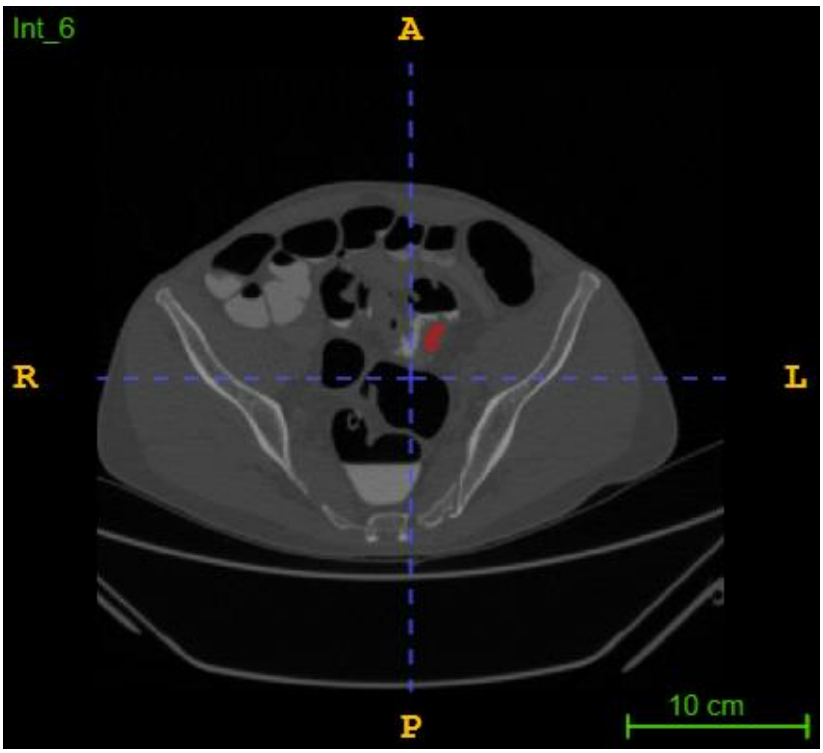

Figure 5 Slice extracted from a CT volume of a patient with diverticula and his corresponding manual segmentation

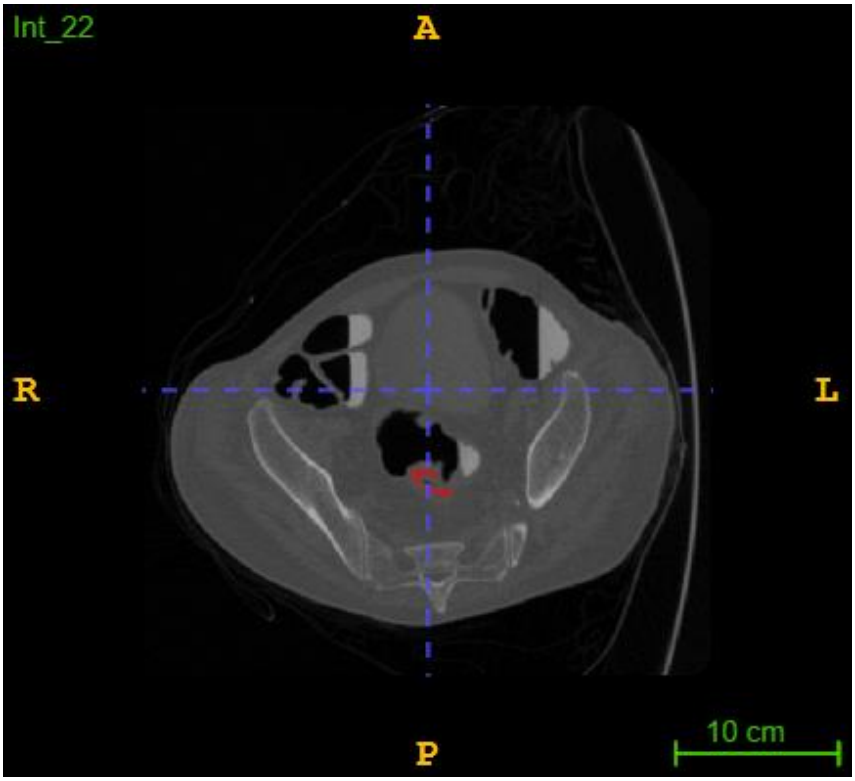

Figure 6: Slice extracted from a CT volume of a patient with colon cancer and his corresponding manual segmentation

2.1.3 DATASET DIVISION

The CTC scans obtained were segmented by the two physicians and the final dataset was obtained. In order to perform the automatic segmentation tasks the dataset was divided into Training and Validation set, which were used in the training of the model, and a Test set used to evaluate the capability of the trained model to generalize on an unknown dataset. For the first task, performed utilizing a standard U-Net, the Dataset was divided into 72 training cases, 32 validation cases and 45 test cases. A total of 149 cases, since 2 were excluded because of a discrepancy in size between the volume and the corresponding manual segmentation. The division was performed with a clustering method, utilizing the mean and standard deviation values of the lesions and the according classification (1 for tumor and 0 for diverticula) to then create a dendrogram. The dendrogram was then cut in order to obtain 4 clusters form which the subjects were extracted randomly to obtain the given sets. For the first nnU-Net task however, computation difficulties prohibited from utilizing the whole set and a significative reduction was necessary.

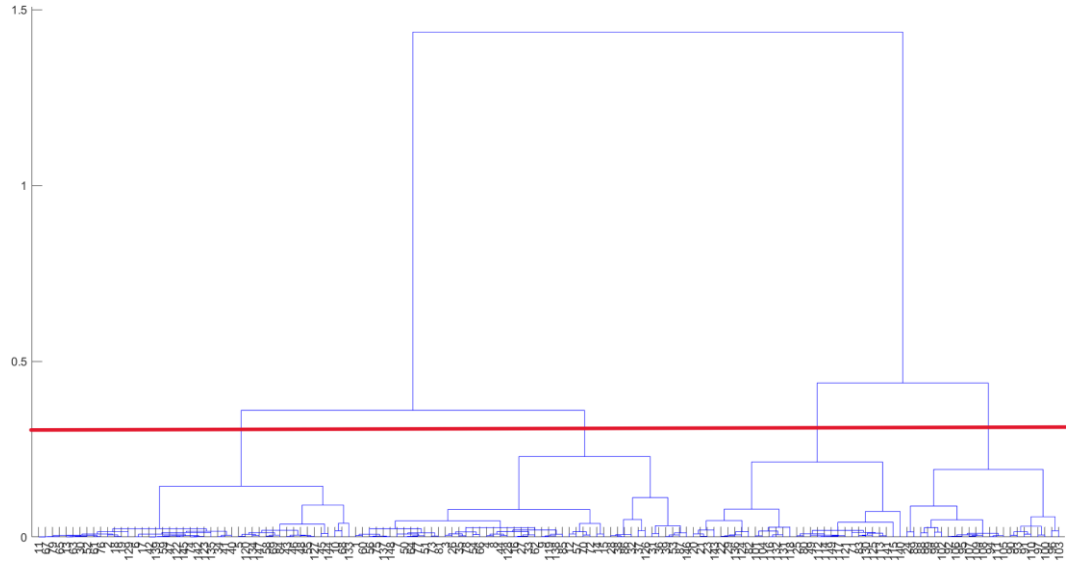

Figure 7: Dendrogram used for clustering

2.2 AUTOMATIC SEGMENTATION

2.2.1 U-NET

As a first approach to the problem, a tentative of automatic segmentation was performed utilizing the standard U-Net architecture in order to understand the first criticalities and where intervention is needed.
The Pre-Processing applied in this case involved reducing the volumes to the slices where the lesion was in order to overcome the computation difficulties. Since the background tissues constitutes a large part of the image compared to the lesion, an Hounsfield unit interval was chosen in order to highlight the interested tissues (between -400 and 1100).
The main parameter for regulating this U-Net were:

- Spatial dimensions: 2, since we wanted to perform a 2-D segmentation task.
- Input channels: 1, since we were dealing with grayscale images.
- Output channels: 2, since the output needed from the network were the image and the tumor/diverticula mask.
- Learning rate: starting at 0.001, with a weight decay of 1e-5
- Loss Function: Dice Loss function.
- Maximum number of iterations: 100.
- Batch size: 15, the number was lowered due to computational issues.

This attempt was far from successful, exposing some of the difficulties of detecting the lesions starting from the CTC images.
The standard UNet architecture was judged not advanced enough to treat this morphologically complex regions, since the lesion were mostly not detected, with DSC coefficient close to 0, as seen in the image below. One of the main problem was determined by the background tissues, that were confused with the lesion very easily, since the morphological characteristics were quite similar.

From here it was obvious the need to use a more advanced architecture to deal with the problem and to limit as much as possible the background tissues in the image.

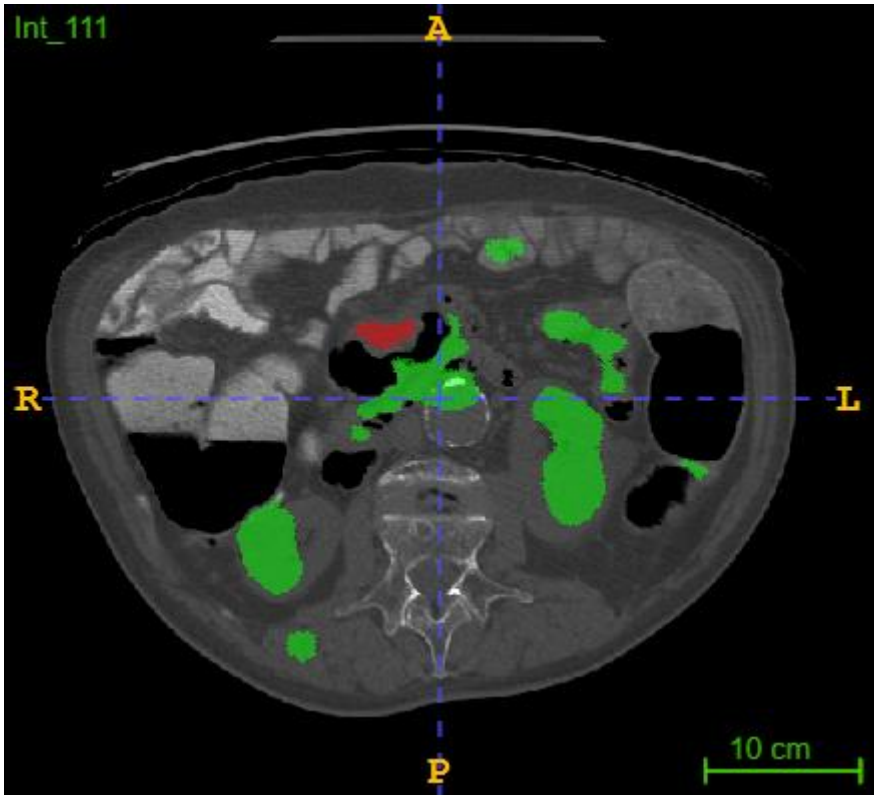

Figure 8: In green the automatic segmentation obtained from the U-Net vs the manual segmentation in red

2.2.2 TOTAL SEGMENTATOR

Total segmentator is a deep learning segmentation tool provided by the department of Research and Analysis at University Hospital of Basel, it was trained on a vast quantity of CT and MRI images, allowing to give a very robust segmentation tool for 104 anatomical structures in CT images. It is heavily based on nnUNet. All the CT images used to train this model are real-life cases. Total segmentator was used in order to obtain a segmentation of the colon area, in order to exclude the non-interesting tissues from the images. An example of the 3-D segmentation is shown in the picture below. [8]

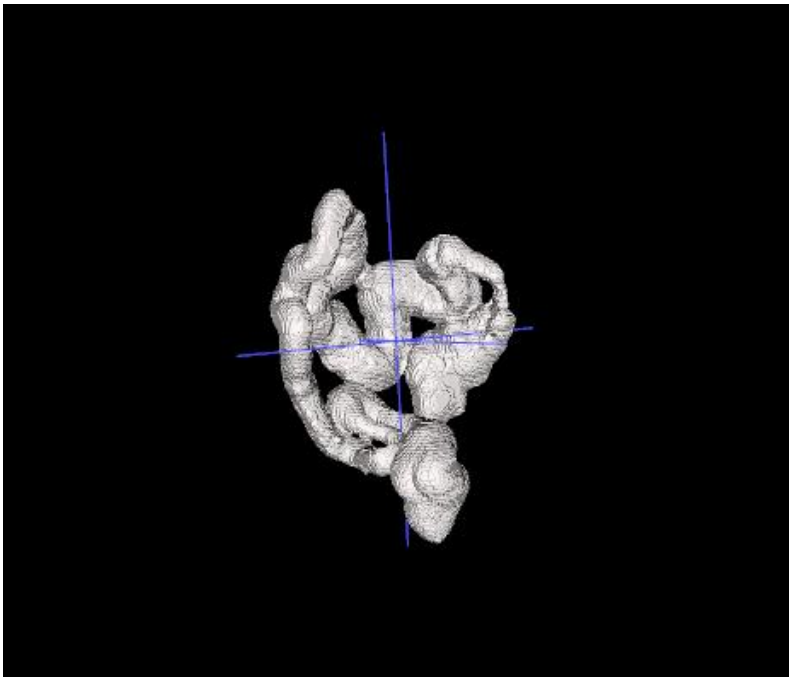

Figure 9: Colon segmentation obtained with Total Segmentator

2.2.3 PRE-PROCESSING

Since the first attempt of using an UNet was unsuccessful, a nnUNet was implemented to detect the interested lesion.
Each CT volume underwent a pre-processing phase:

- Normalization: by using the minimum and maximum value of the pixel on the image it was performed a Min-Max normalization to obtain volumes with pixel values included in an interval between 0 and 1.
- Resampling: since the resolutions on the x and y axis differed from case to case, it was crucial to give each slice of each volume a spatial 2D resolution of 1mm x 1mm, reducing the dimension of the data to be given to the nnUNet architecture. A resampling was also performed on the z-axis for 47 volumes, whose spacing was between 0,5 and 0,7 mm and was brought to 1 mm, in order to reduce the number of slices on these cases.
- Cropping and padding: after applying Total Segmentator on each volume, at first the percentage of the lesion inside the segmentation masks was calculated. In most cases the lesion was adjacent to the colon mask but not included in it, so the step performed was to calculate the center of each slice's segmentation, then calculate the maximum and minimum coordinates in each direction. A rectangle was drawn to cover the colon region and the lesion included. This rectangle ended up to become the final part of the image, while to cover the rest zero padding was applied.
  The colon segmentation was also used in order to remove some of the initial and final slices on each volume, since there was no colon to begin with.
  Thanks to this step, a great part of the background tissues was excluded from the final image in order to eliminate unnecessary data to feed to the nnUNet, as shown in the picture below.

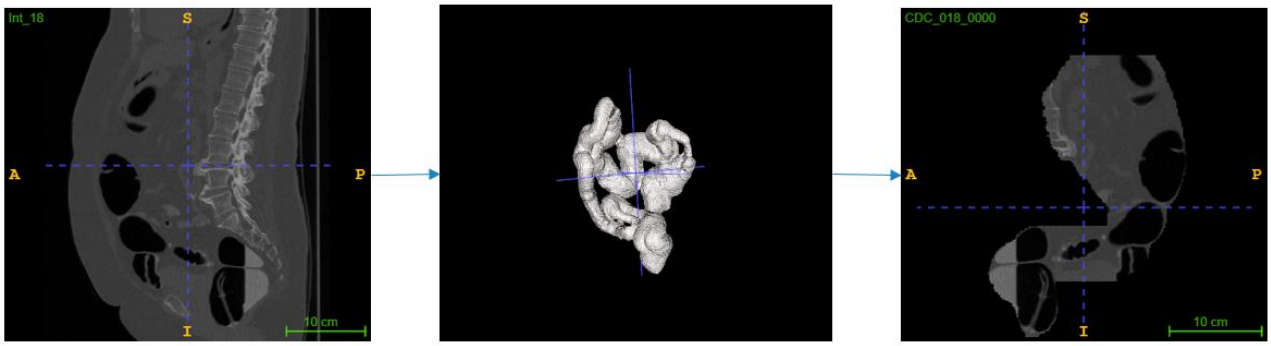

Figure 10: Pre-Processing steps starting from the original volume

2.2.4 NNUNET

The nnU-Net was chosen as a more advanced tool to execute the automatic segmentation task. Its first approach evidenced some difficulties due to computational limits, so it was first tried on a heavily reduced dataset, with a total of 39 training cases, 11 validation cases and 99 test cases .
Only with further attempt it was possible to utilize the previously defined dataset division and include more cases in the training phase.
The dataset fingerprint was extracted by utilizing the “plan and pre-process” function, although the pre-processing was reduced by selecting the “no resampling hack” since the pre-processing phase had been already performed beforehand. The U-Net like architecture and its parameters were automatically chosen in this process and are shown in the table below. From the automatically chosen parameters the batch size was changing in order to reduce the computational cost.

nnU-Net Parameters	
Initial feature map size	32
Patch size	396x396
Batch size	10
Number of convolutions	71 (64 encoder, 6 decoder, 1 final)
Activation function	Leaky ReLU
Loss function	Dice Loss + Cross-Entropy Loss
Number of stages	7 (7 encoder, 6 decoder)

Figure 11: nnU-Net architecture

3 RESULTS

3.1 EVALUATION METRICS

After applying the nnU-Net model to the dataset, evaluation metrics were needed in order to estimate the efficiency of the automatic segmentation method.

3.1.1 DICE SIMILARITY COEFFICIENT

One of the most common metric utilized to measure the performances in medical image segmentation is the Dice Similarity Coefficient (DSC). Its purpose is to evaluate the spatial overlap of the automatic segmentation on the manual segmentation. It is commonly defined by 2 times the intersection between the two segmentation masks divided by their sum, as in the formula below:

$$DSC = \frac{2|A \cap B|}{(A + B)}$$

In which A and B are the pixels of the respective automatic and manual mask, while ∩ represents the intersection. [9]
The values obtained from this metric can differ from a range of 0 to 1:

- DSC = 0: there is no overlap between the two masks, the automatic segmentation failed to recognize the correct pixels of the corresponding manual mask.
- 0 < DSC < 1: there is a partial overlap between the two mask, the mask obtained from the automatic segmentation shares a number of pixels with the manual one.
- DSC = 1: complete overlap between the two masks.

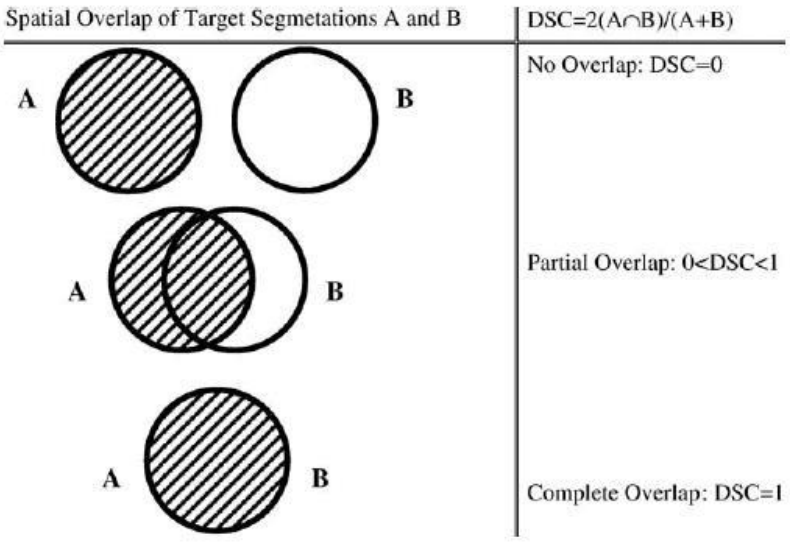

Figure 12: Graphic Representation of Dice Similarity Coefficient [9]

3.1.2 RELATIVE VOLUME DIFFERENCE

Alongside the DSC another metric was used to evaluate the performance of the automatic segmentation, the RVD or Relative Volume Difference. As the name suggests, it is a measure of the difference in size between the manual and the automatic segmentation mask as a fraction of the manual one. [10]
It was implemented by subtracting the sum of the pixels of each mask and dividing by the sum of the pixels of the manual mask created by the physician, using the following formula:

$$RVD = \frac{|sum(X)| - |sum(Y)|}{|sum(Y)|}$$

Where Y represents the ground truth (manual mask) and X the predicted segmentation (automatic mask). The RVD is useful to detect the cases of overestimation and underestimation of the interested region in the final automatic segmentation.

3.1.3 PRECISION, SENSITIVITY, SPECIFICITY, BALANCED ACCURACY

To summarize the results of the segmentation task other metrics can be employed, and they're based on 4 main categories which depend on the correct or erroneous prediction of the interested labels:

- True Positives (TP): the pixels related to the lesion of the automatic mask which coincide with the ones of the manual mask.
- True Negatives (TN): the pixels related to the healthy tissues of the automatic mask which coincide with the ones of the manual mask.
- False Positives (FP): pixels predicted as cancer or diverticula in the automatic mask but are labeled as healthy in the manual one.
- False Negatives (FN): pixels predicted as healthy tissues in the automatic mask but are labeled as cancer or diverticula in the manual one. [11]

From these values the following evaluation metrics can be extracted:

- Sensitivity: a fraction of the pixels labeled correctly as cancer or diverticula.

$$\frac{TP}{TP + FN}$$

- Specificity: a fraction of the pixels labeled correctly as healthy tissues.

$$\frac{TN}{TN + FP}$$

- Precision: a fraction of the pixels labeled as unhealthy that are actually correct.

$$\frac{TP}{TP + FP}$$

- Accuracy: a fraction of the correctly classified pixels.

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- Balanced Accuracy: accuracy metric which is more useful in an unbalanced dataset such as the object of the thesis, since the diverticula and cancer cover a small area compared to the healthy tissues.

$$\frac{SENSITIVITY + SPECIFICITY}{2}$$

- F1 Score: the equivalent of the Dice Similarity Coefficient.

$$\frac{2(SENSITIVITY * PRECISION)}{SENSITIVITY + PRECISION}$$

3.2 FINAL RESULTS

3.2.1 TRAINING

In the figures below the trend of the training variables such as training and validation loss and a pseudo dice is shown for the two nnU-Net models trained.

Figure 13: First training trend with smaller set

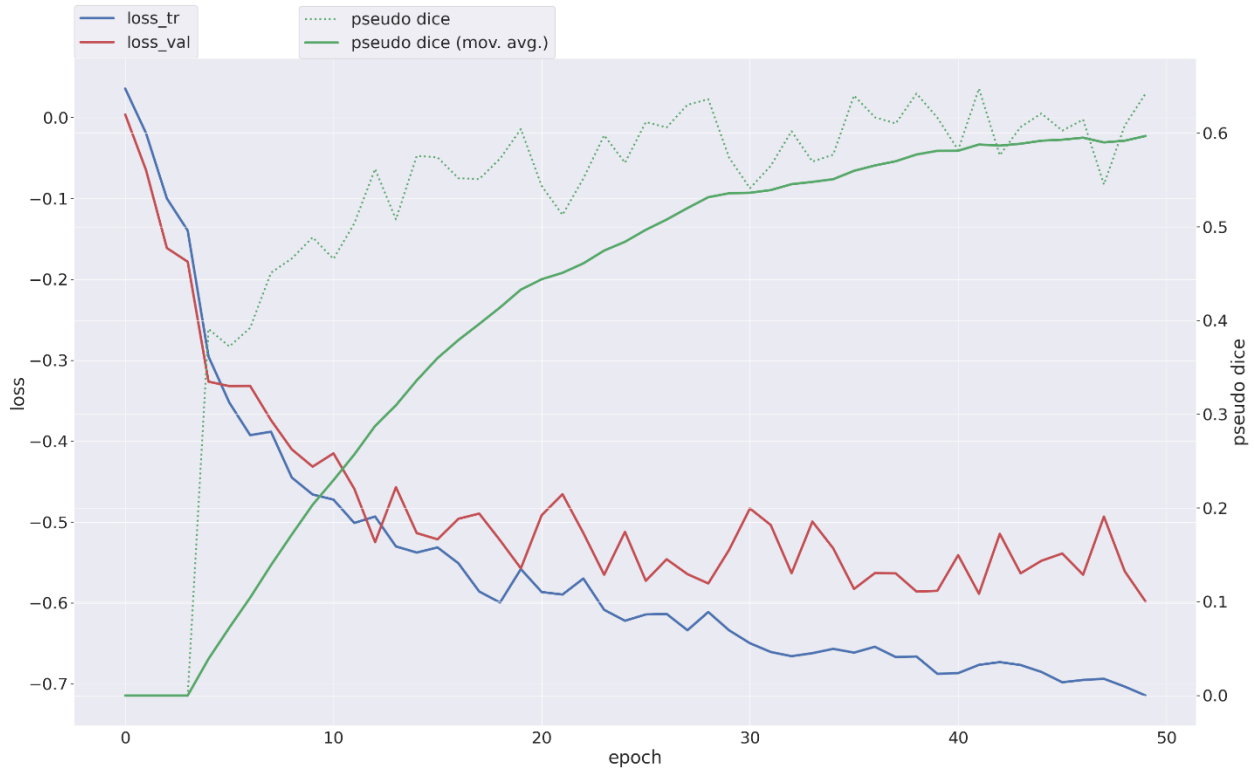

Figure 14: Second training trend with larger dataset

The values of validation loss and training loss have been reported to be better for the second training, as expected since the number of training and validation cases increased.

3.2.2 DSC VALUES

The performances of the automatic segmentation obtained with the trained nnUNet were calculated by confronting the automatic segmentation with the manual segmentation for the images of the whole dataset.
The similarities between the manual mask and the automatic ones was obtained by calculating, for each 3D volume, the Dice Similarity Coefficient (DSC) metric, applying it to the region that contained the lesion, whether it's a tumorous one or diverticula.
In the graphics below the distribution of DSC values for the Training, Validation and Test sets are shown.
For the first application with a smaller dataset in the Training Set has been observed a mean DSC value of 0.34, with a standard deviation of 0.30. As for the Validation and Test Set, a mean DSC value of 0.13 was obtained, with a standard deviation of 0.20.
As for the second application with a larger dataset in the Training Set has been observed a mean DSC value of 0.59, with a standard deviation of 0.24. As for the Validation and Test Set, a mean DSC value of 0.51 was obtained, with a standard deviation of 0.32 for Validation and 0,24 with a standard deviation of 0,22 for the Test Set. The boxplots with the distribution of values are shown below.

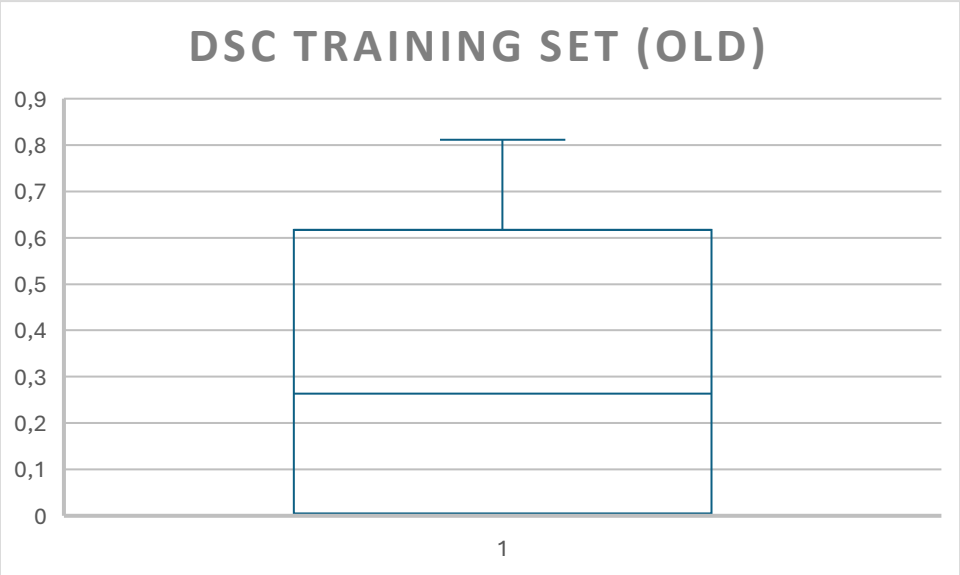

Figure 15: Dice Similarity Coefficient for the Training Set with fewer cases

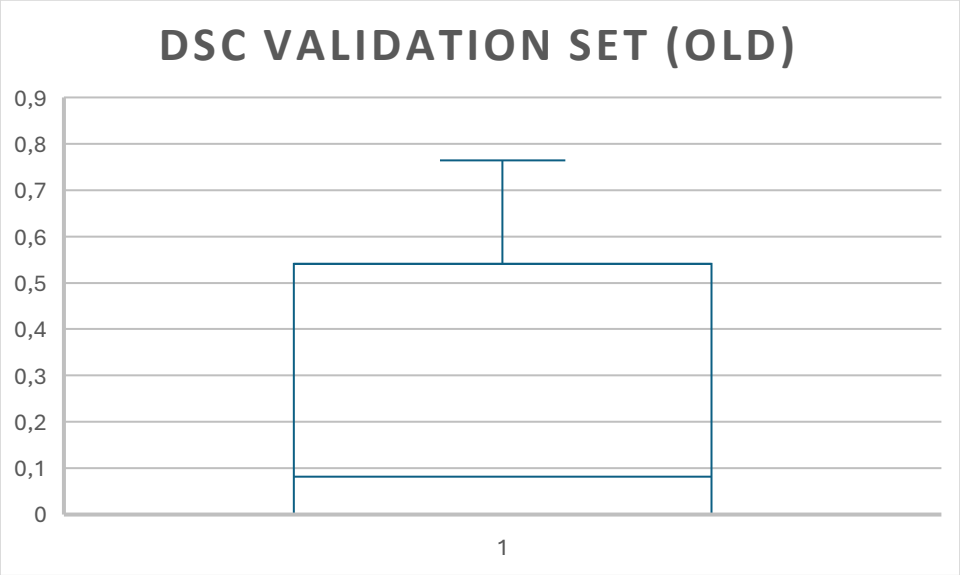

Figure 16: Dice Similarity Coefficient for the Validation Set with fewer cases

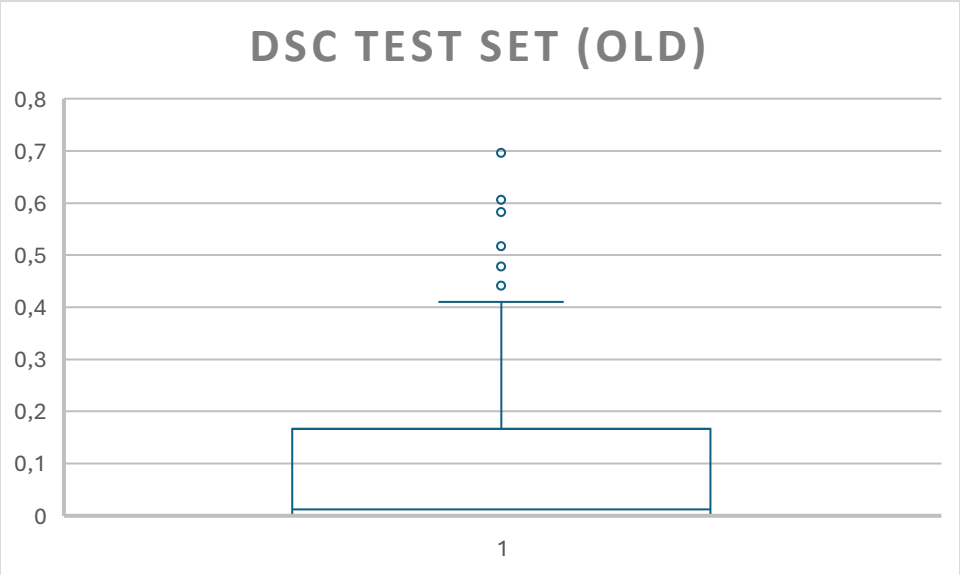

Figure 17: Dice Similarity Coefficient for the Test Set of the first nnU-Net

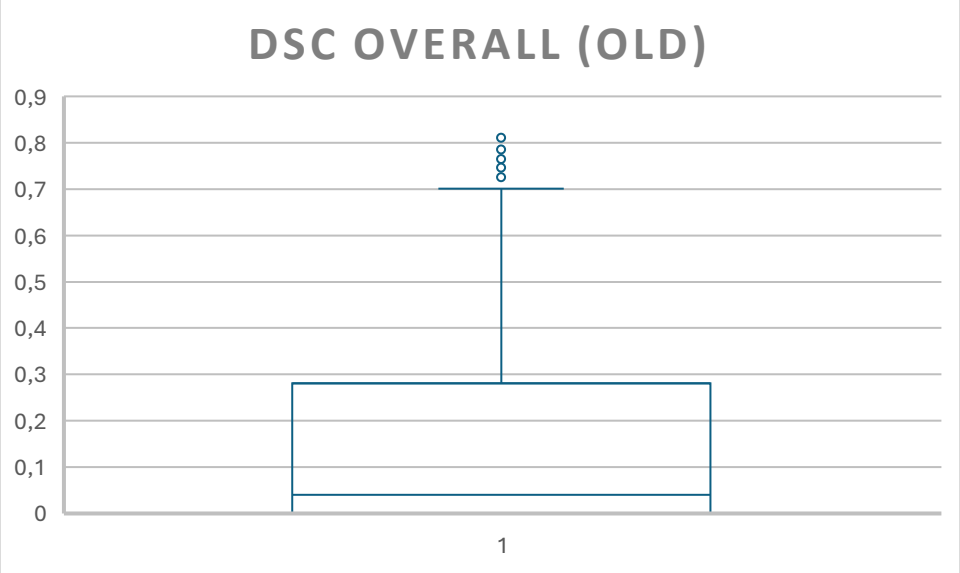

Figure 18: Overall Dice Similarity Coefficient for the first nnU-Net

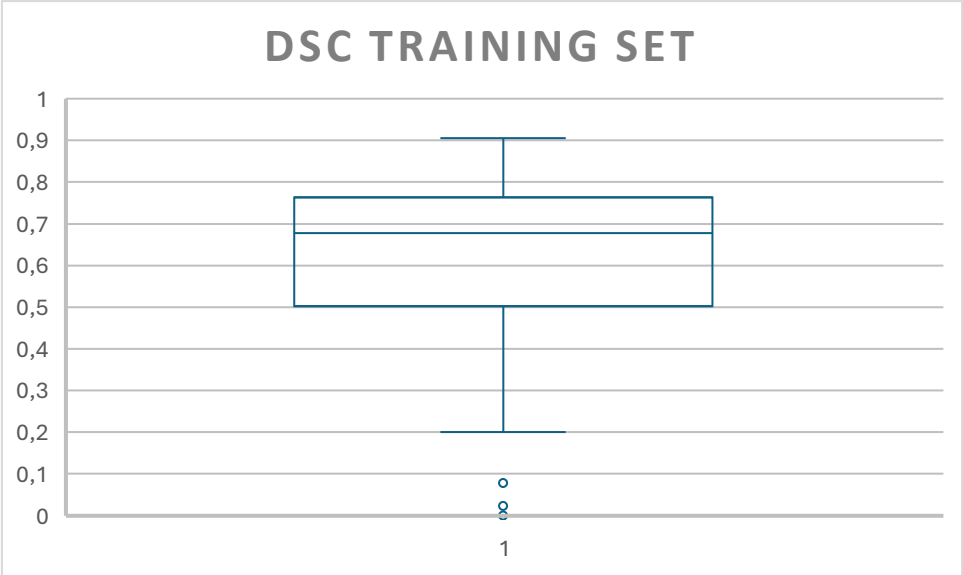

Figure 19: Dice Similarity Coefficient for the Training Set of the second nnU-Net

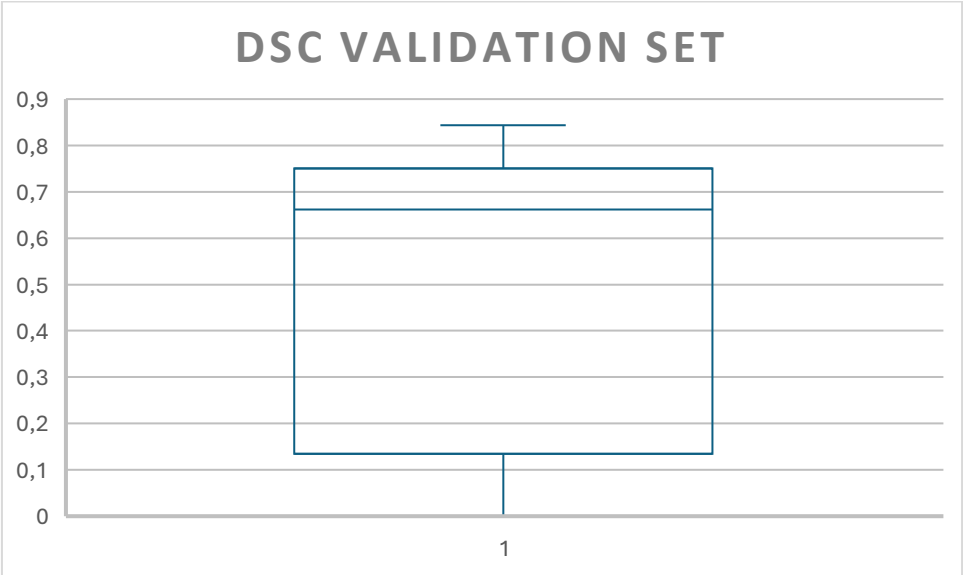

Figure 20: Dice Similarity Coefficient for the Validation Set of the second nnU-Net

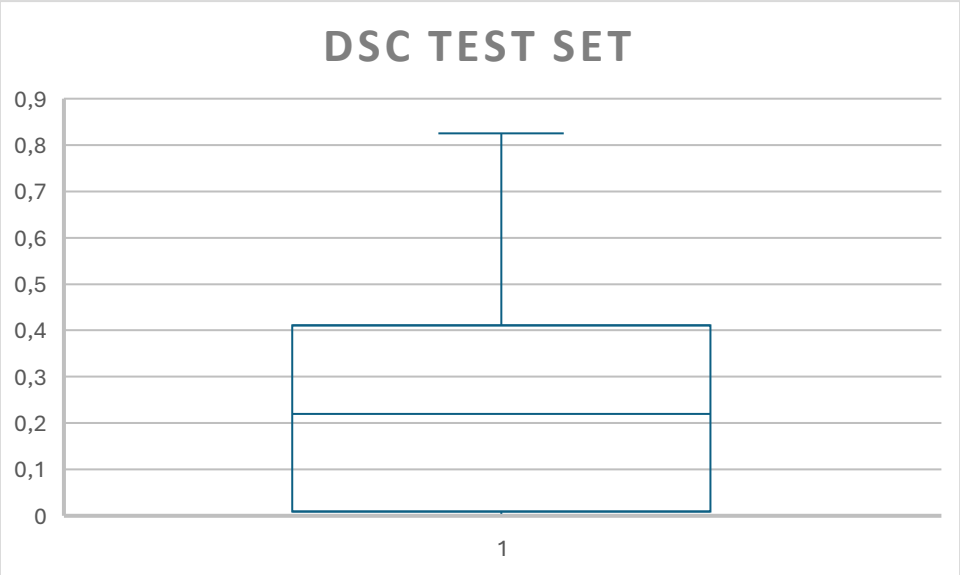

Figure 21: Dice Similarity Coefficient for the Test Set of the second nnU-Net

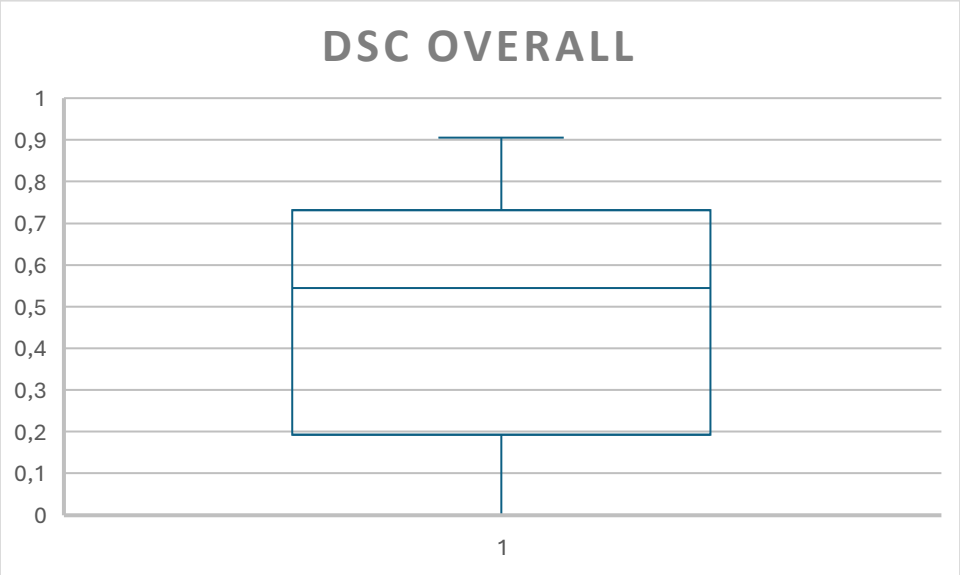

Figure 22: Overall Dice Similarity Coefficient for the second nnU-Net

3.2.3 RVD VALUES

Another confrontation between the automatic and manual mask was performed by calculating the Relative Volume Difference values for the masks obtained with both trainings. The boxplot with the distribution of RVD values are shown below.

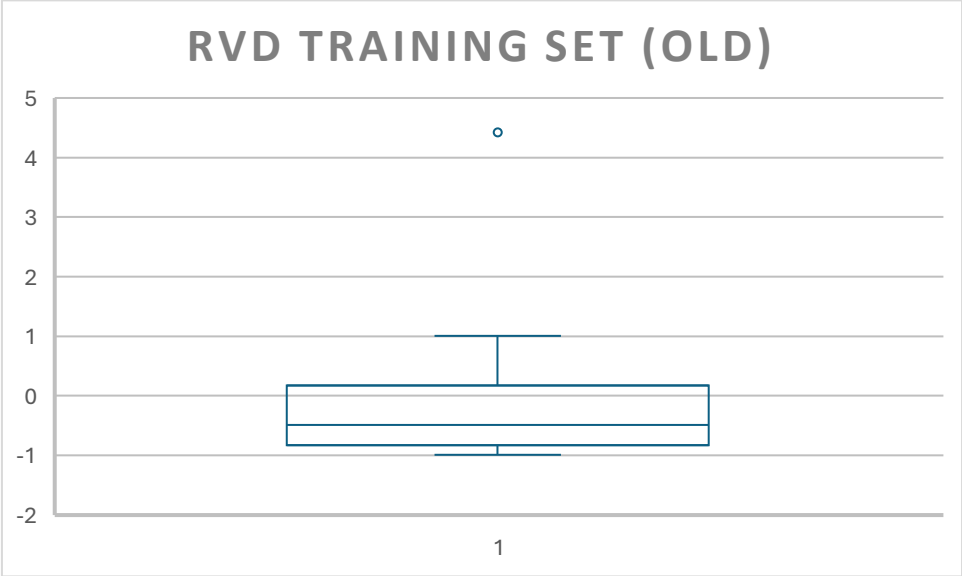

Figure 23: Relative Volume Difference for the Training Set of the first nnU-Net

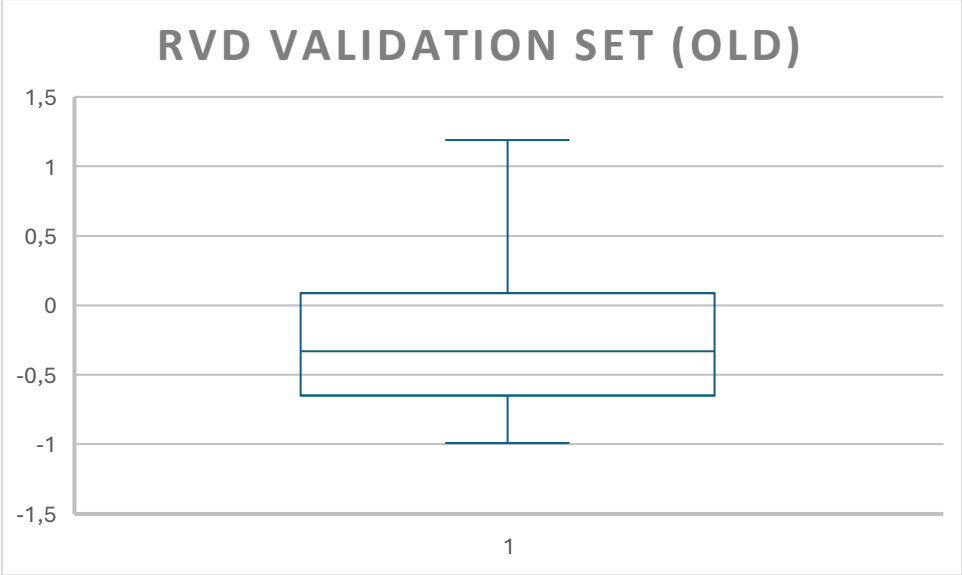

Figure 24: Relative Volume Difference for the Validation Set of the first nnU-Net

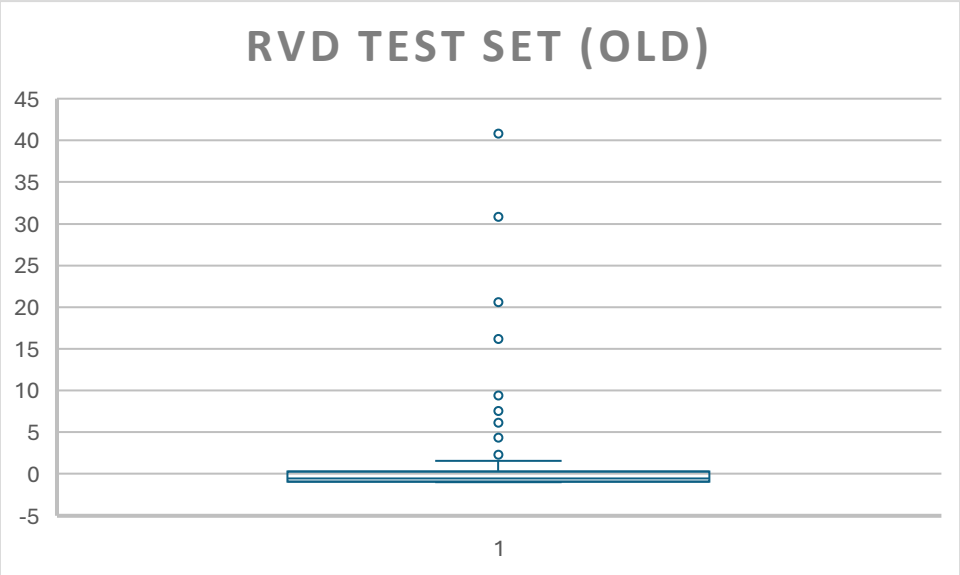

Figure 25: Relative Volume Difference for the Test Set of the first nnU-Net

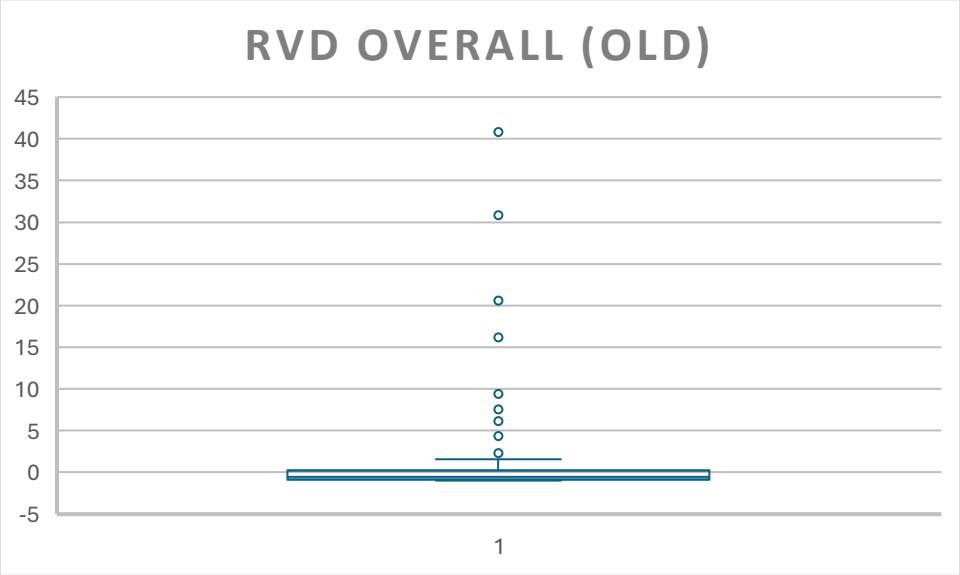

Figure 26: Relative Volume Difference overall values distribution for the first nnU-Net

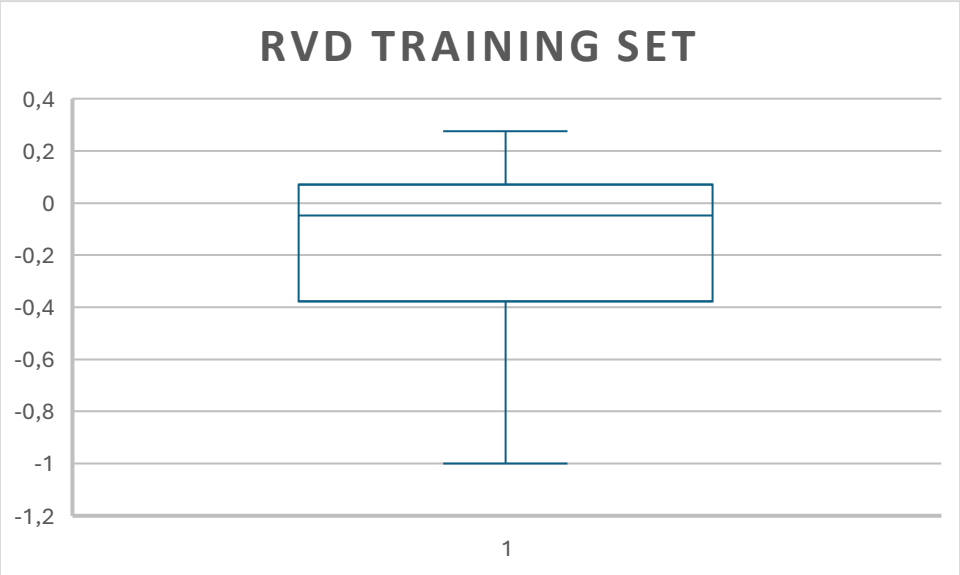

Figure 27: Relative Volume Difference for the Training Set of the second nnU-Net

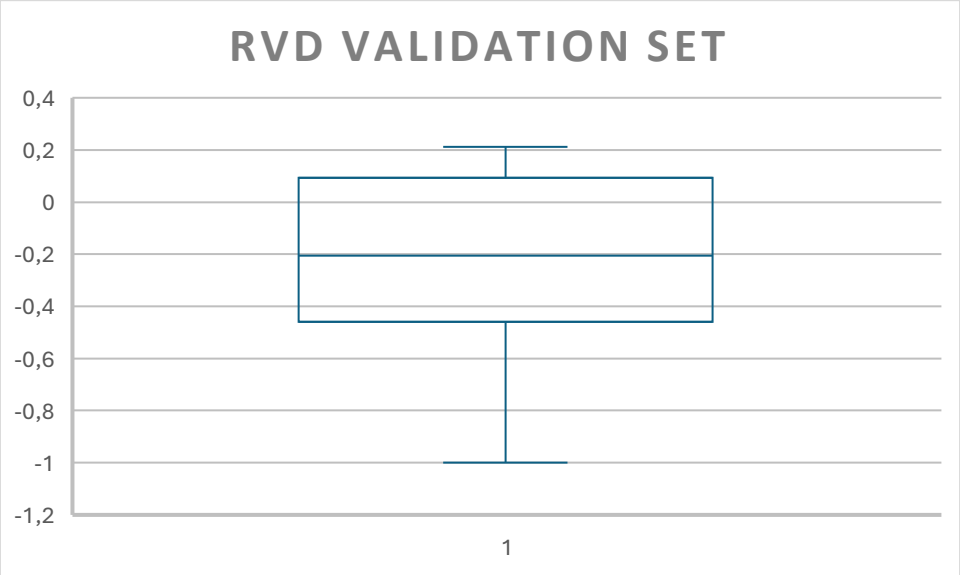

Figure 28: Relative Volume Difference for the Validation Set of the second nnU-Net

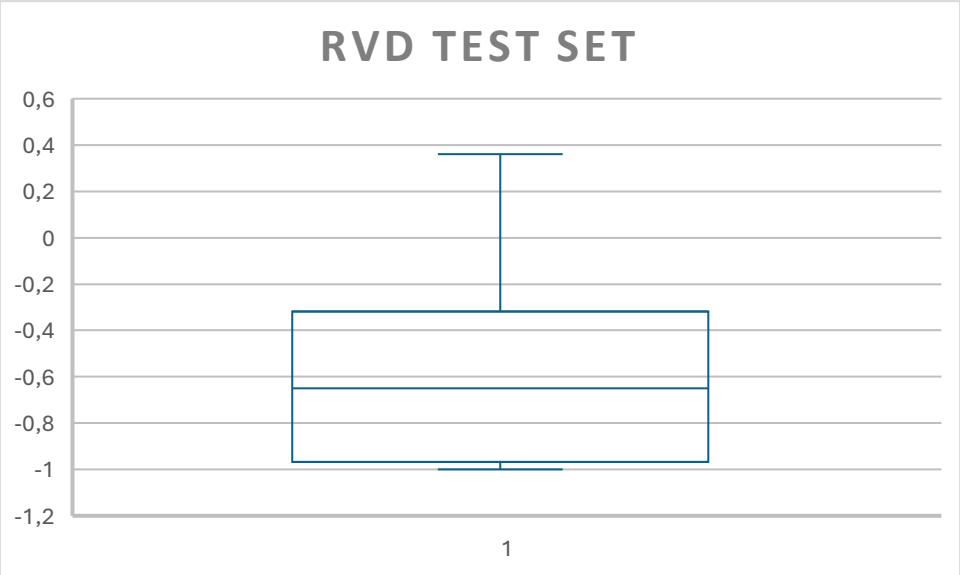

Figure 29: Relative Volume Difference for the Test Set of the second nnU-Net

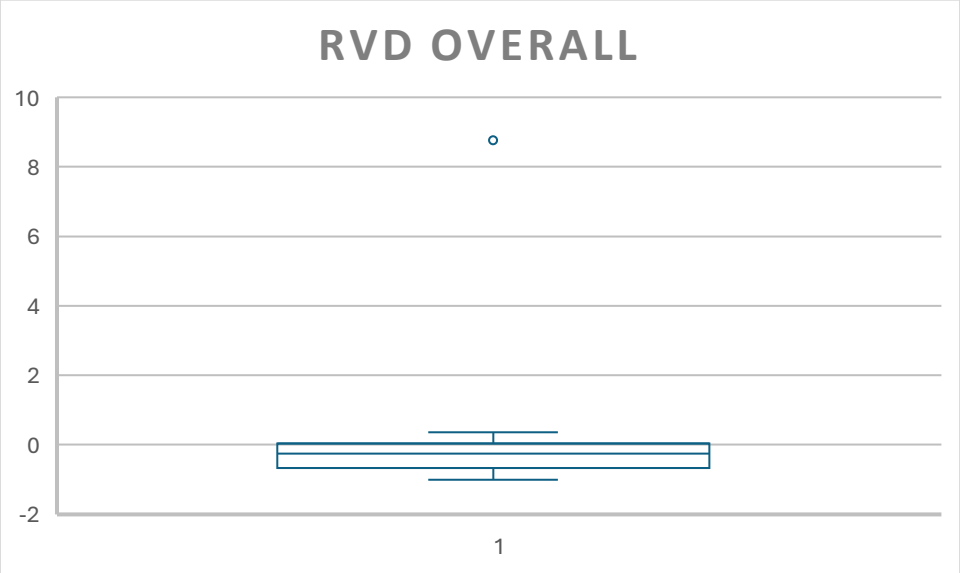

Figure 30: Relative Volume Difference overall distribution for the second nnU-Net

3.2.4 PRECISION AND RECALL

The mean values of Precision, Sensitivity, Balanced Accuracy and F1 score are reported below for Training, Validation and Test Set of the two nnU-Net trained.

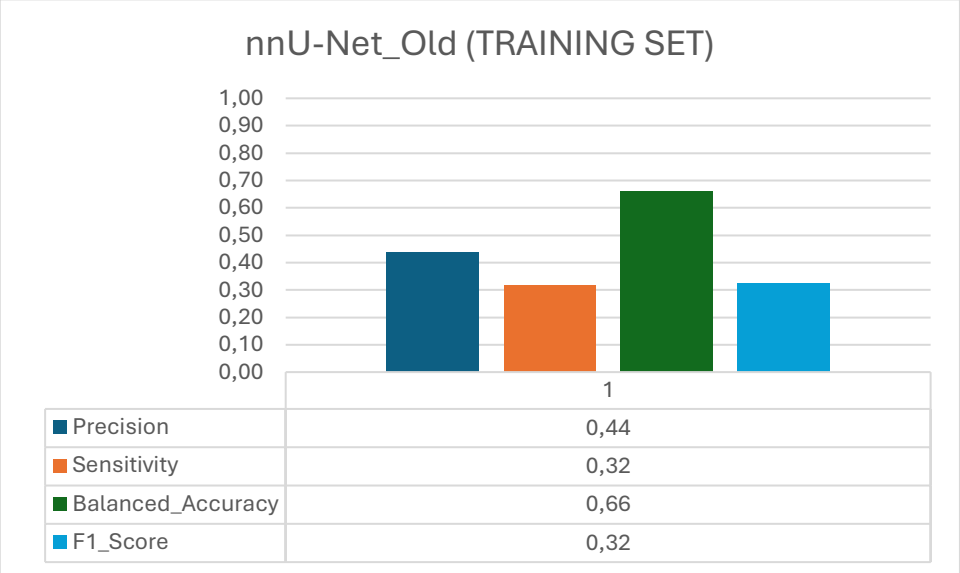

Figure 31: Evaluation metrics of the Training Set of the first nnU-Net

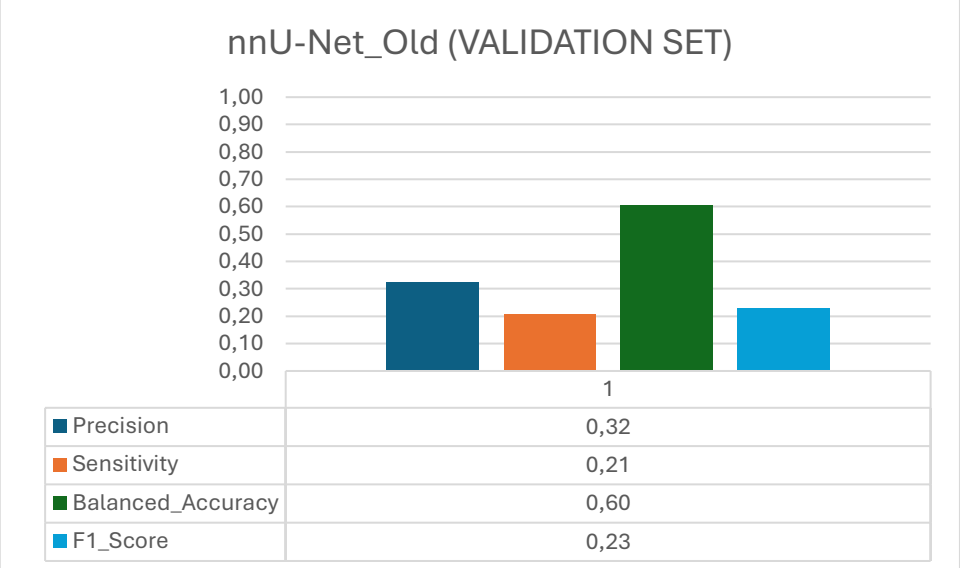

Figure 32: Evaluation metrics of the Validation Set of the first nnU-Net

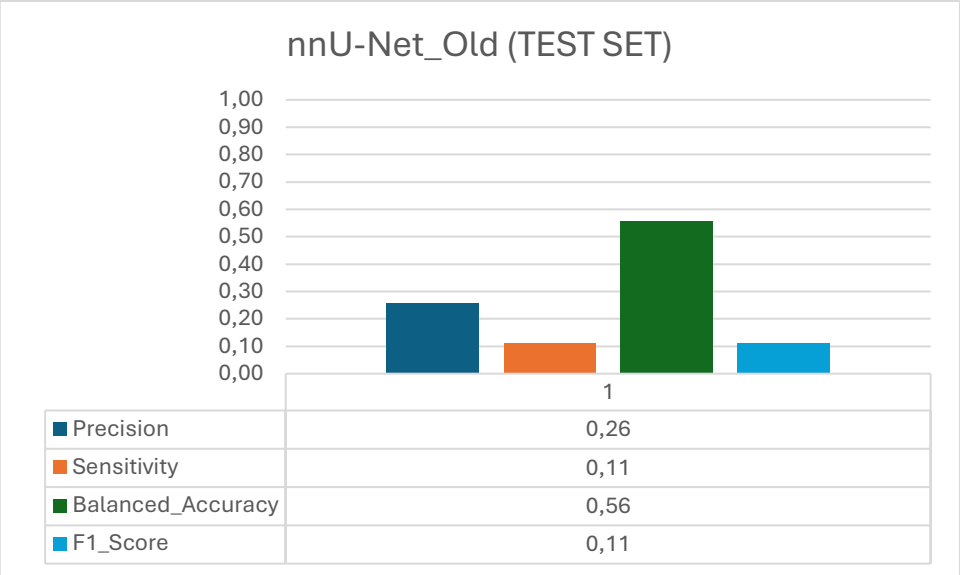

Figure 33: Evaluation metrics of the Test Set of the first nnU-Net

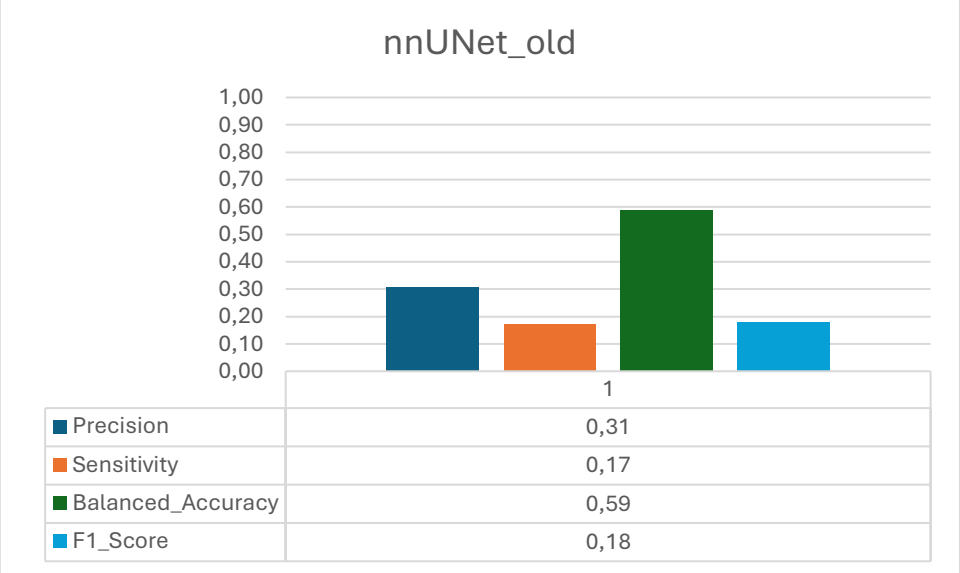

Figure 34: Overall evaluation metrics of the first nnU-Net

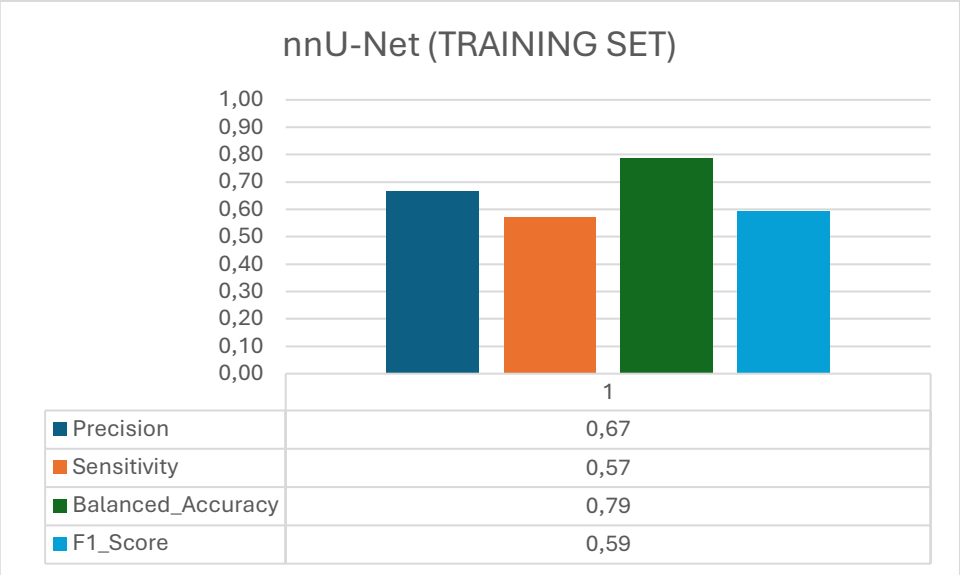

Figure 35: Evaluation metrics of the Training Set of the second nnU-Net

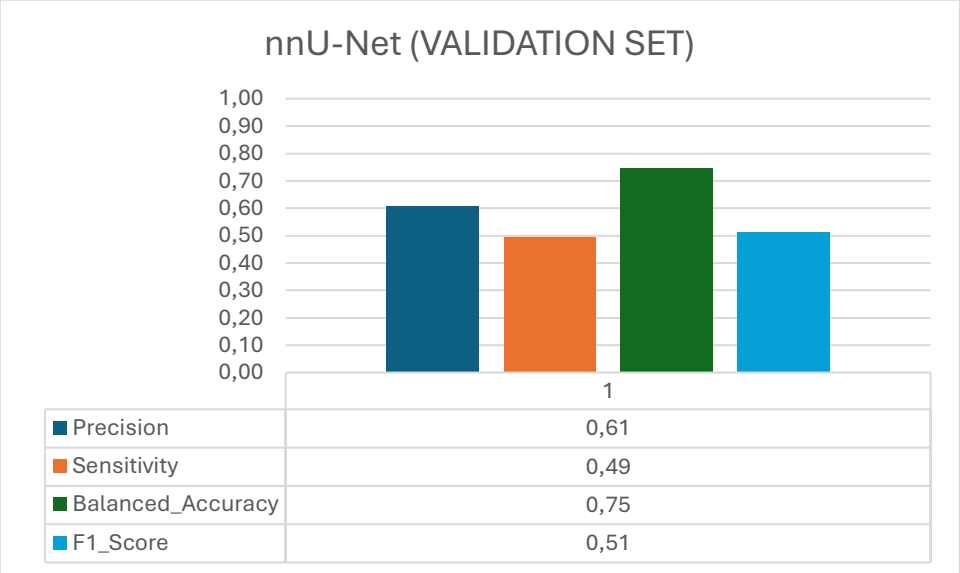

Figure 36: Evaluation metrics for the Validation Set of the second nnU-Net

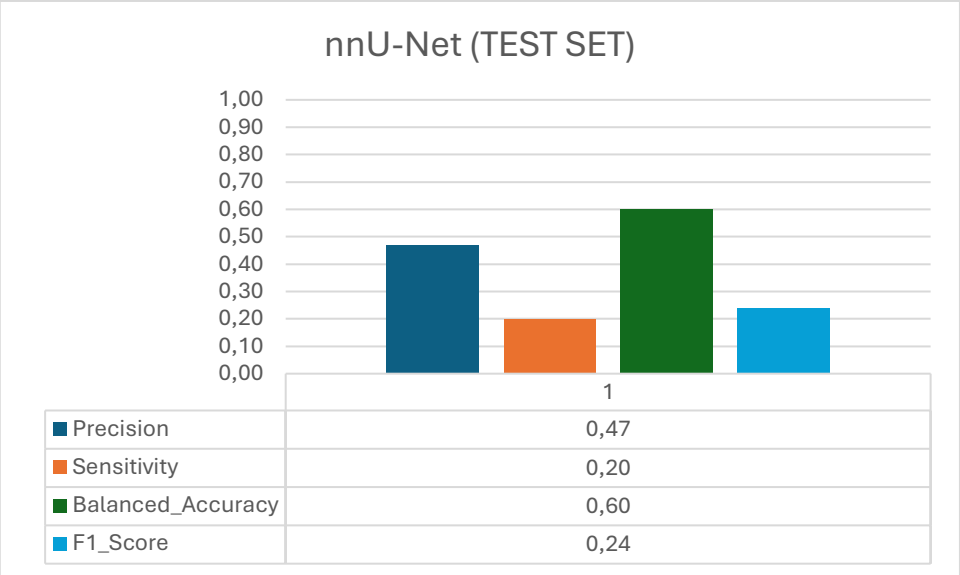

Figure 37: Evaluation metrics of the Test Set of the second nnU-Net

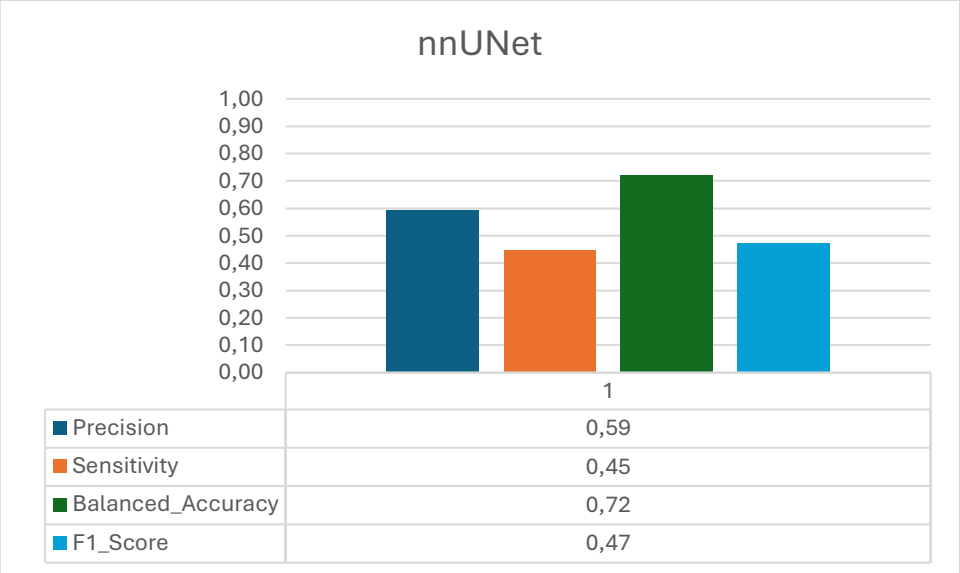

Figure 38: Overall evaluation metrics of the second nnU-Net

3.2.5 CONFRONTATION BETWEEN CANCER AND DIVERTICULA

An aspect of interest of the thesis was also to observe the differences in detection and segmentation of the two different type of lesions in the CT scans. As the diverticula is just a growth in the colonic wall the morphology is largely similar to the surrounding colon tissues, while the automatic segmentation method was slightly more performing with the colon that distinguishes himself more from the surrounding tissues. The evaluation metrics distribution for both cancer and diverticula are reported in the figures below.

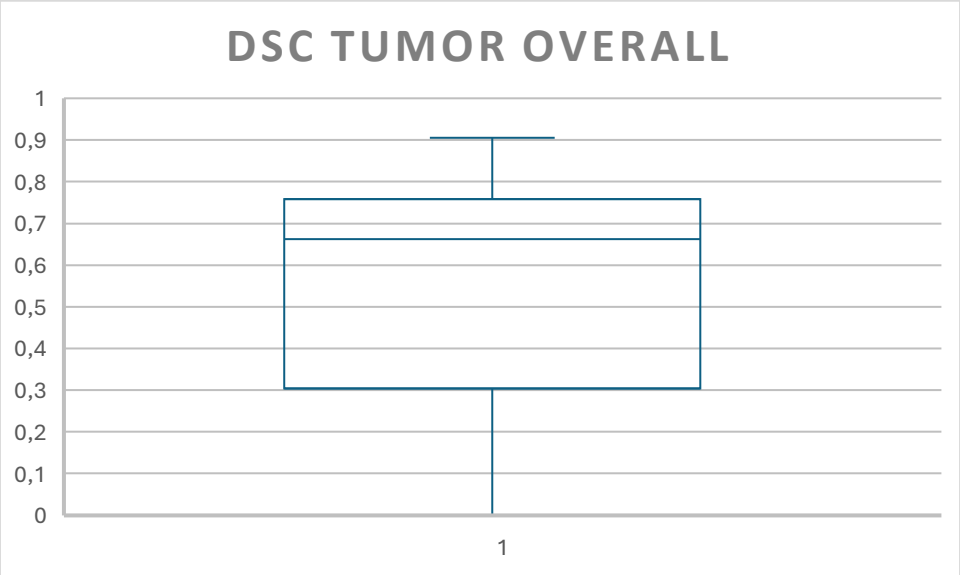

Figure 39: DSC distribution for the tumor

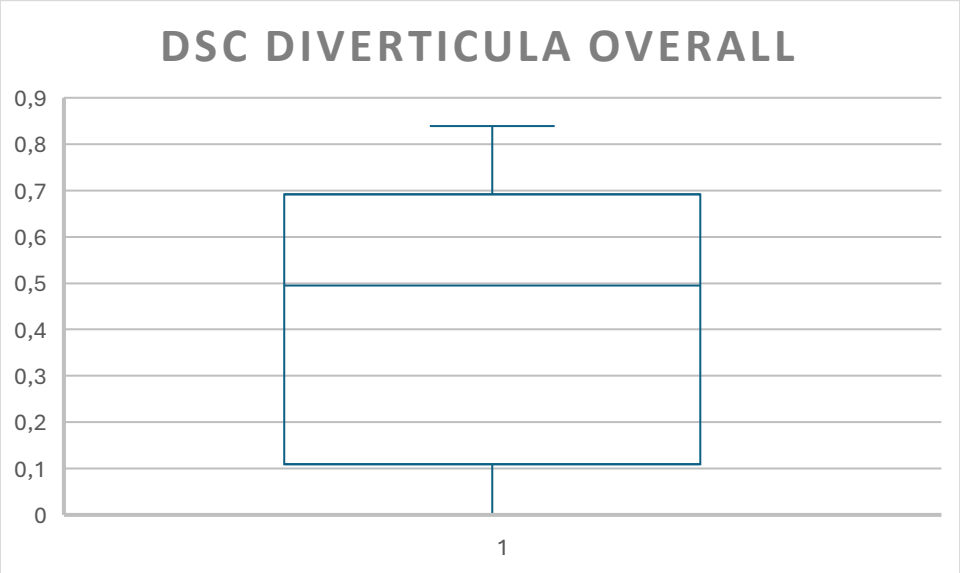

Figure 40: DSC distribution for the diverticula

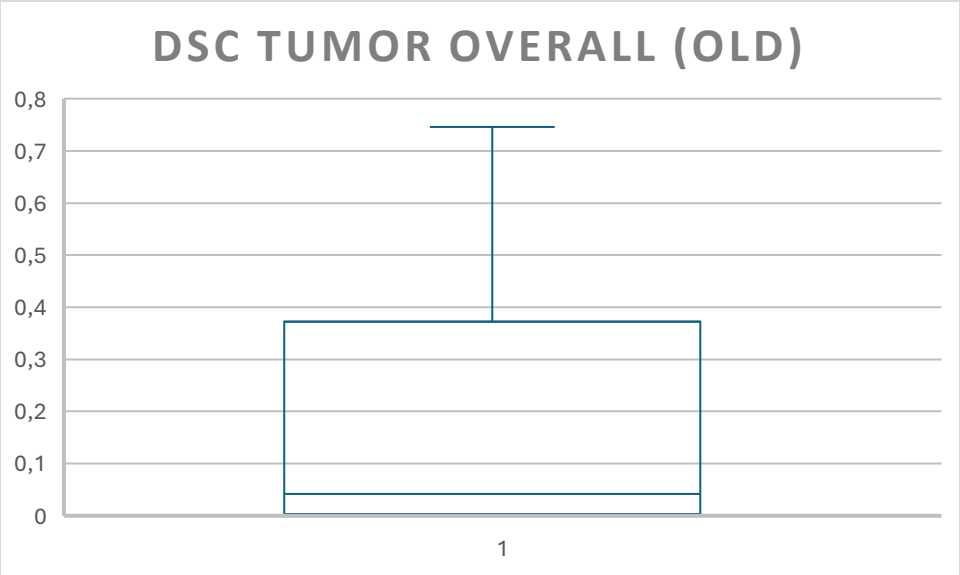

Figure 41: DSC distribution for the tumor in the first nnU-Net

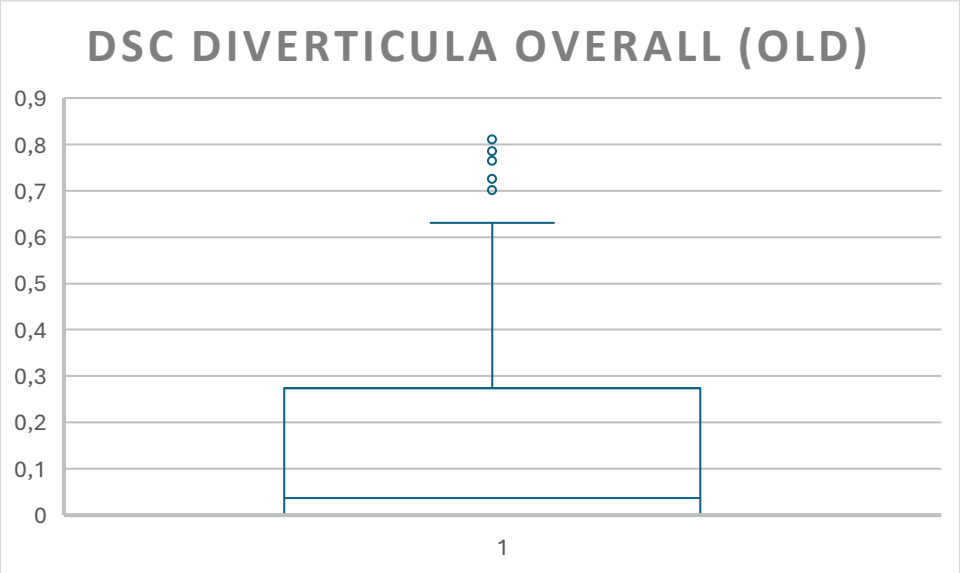

Figure 42: DSC distribution for the diverticula in the first nnU-Net

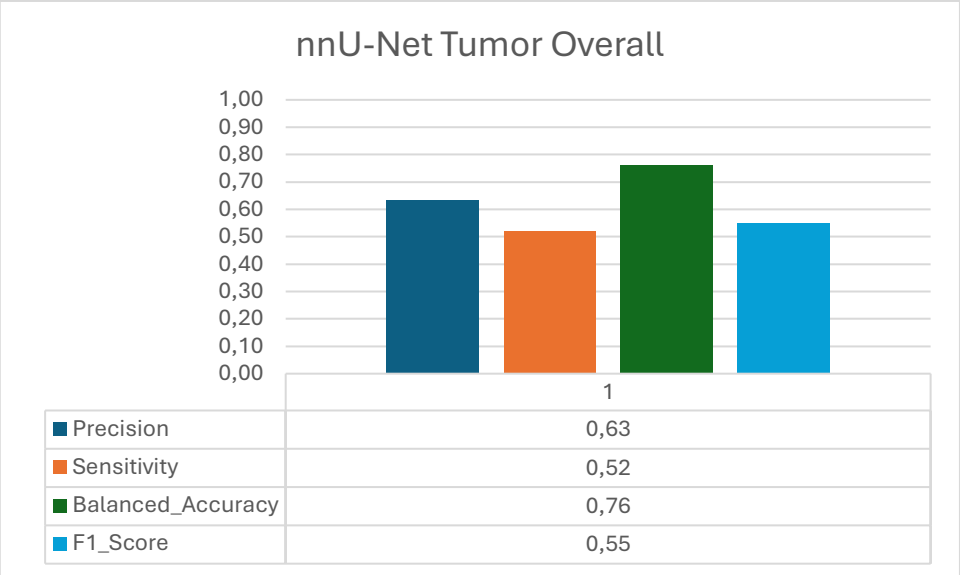

Figure 43: Evaluation metrics for the tumor

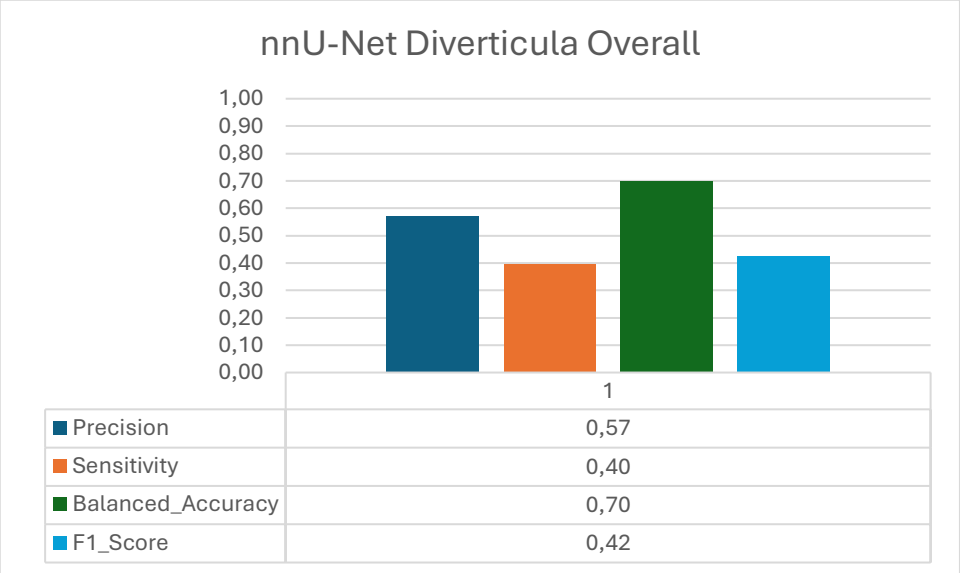

Figure 44: Evaluation metrics for the diverticula

Another aspect considered in the confrontation between the two types of lesions was the distribution of the patients for certain intervals of DSC value. For the first nnU-Net the division was made considering 3 intervals of DSC values:
- from 0 to 0.1, that corresponds to a non-identified case
- from 0.1 to 0.4, that correspond to a identified lesion but not completely segmented
- higher than 0.4, that corresponds to an acceptable segmentation

This graph highlighted the fact that no particular preference in the detection of one region in respect of another was observed.

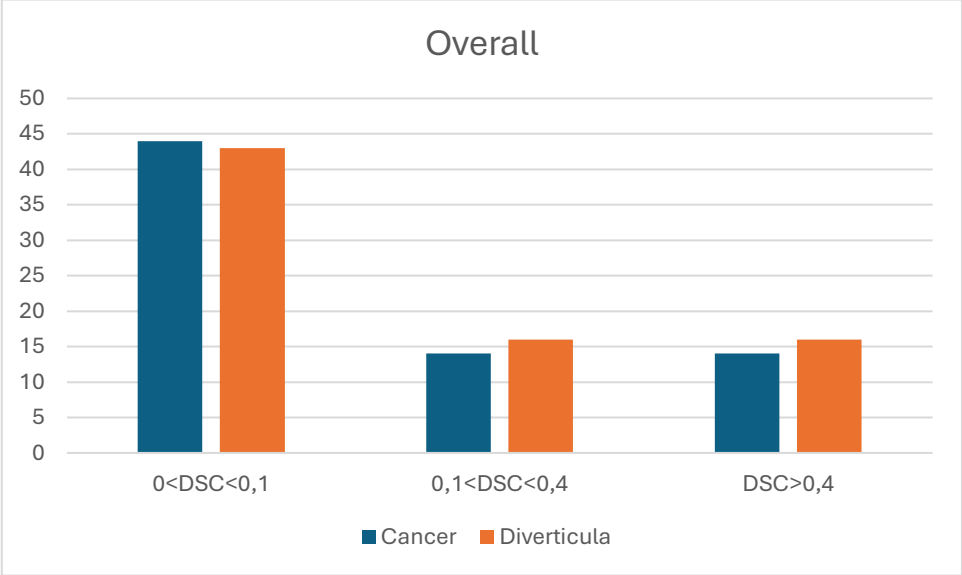

Figure 45: Patients divided by DSC and final diagnosis

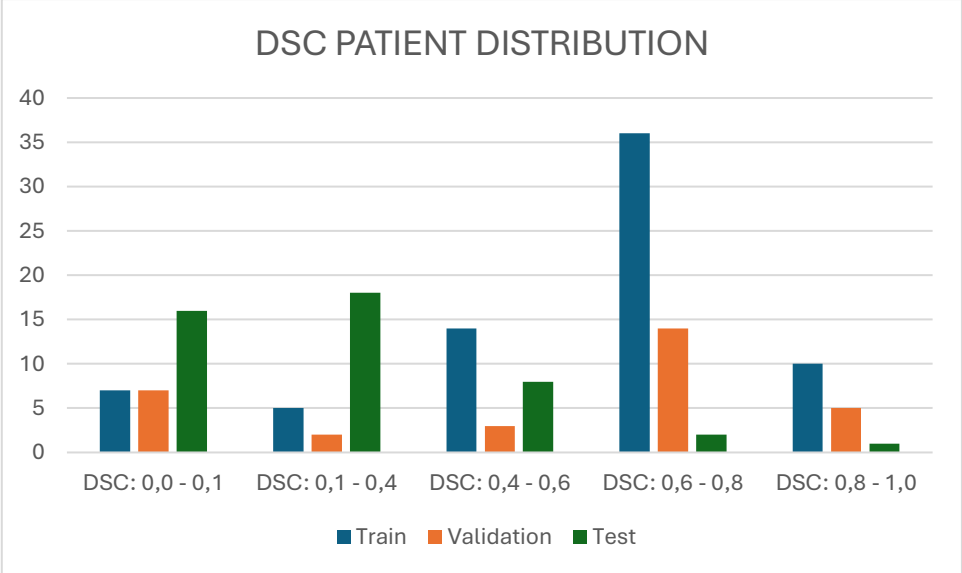

Figure 46: DSC distribution among the patients for the final nnU-Net

3.2.6 MAIN ERRORS

By evaluating the errors on the nnUNet segmentation, 2 macro categories have been considered relevant in the analysis, in terms of false positives and lesions not detected, especially when the DSC value was lower than 0.1.
The first one was observed when the network detects a false positive, by giving a segmentation of an healthy area instead of the interested lesion, as seen in the picture below.
This type of error was observed for both nnU-Net models trained, with a higher presence in the case with less patients, since the training was performed using a small set and the capability to generalize was largely impeded by that.

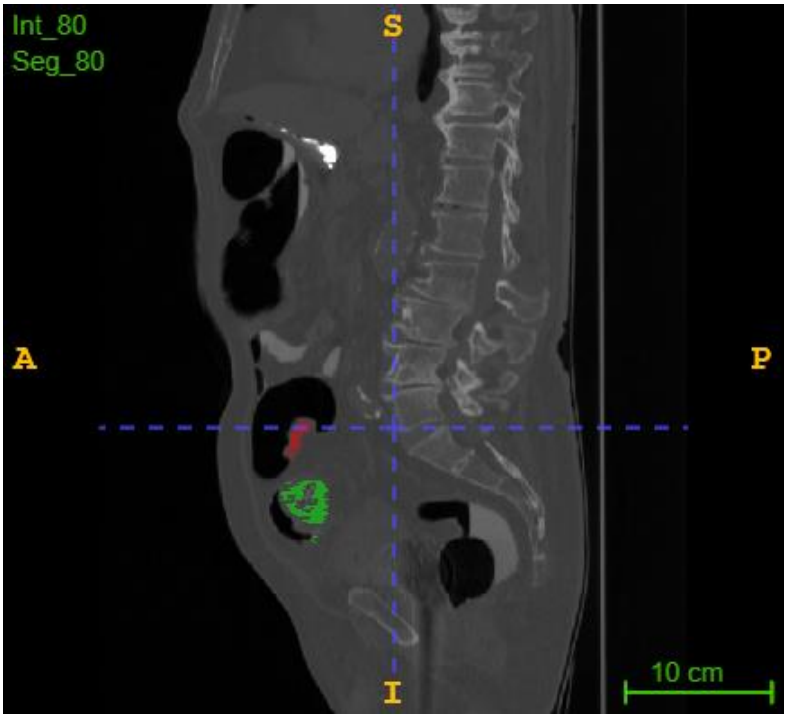

Figure 47: Sagittal view of an incorrect detection of the lesion

The second type of error was observed when the network fails to detect any lesion at all, hence the automatic segmentation results in a volume of zeros, and corresponds to a null DSC value. This was mainly observed in cases where the tumor or the diverticula occupies a really small area in the colon, as shown in the picture below.

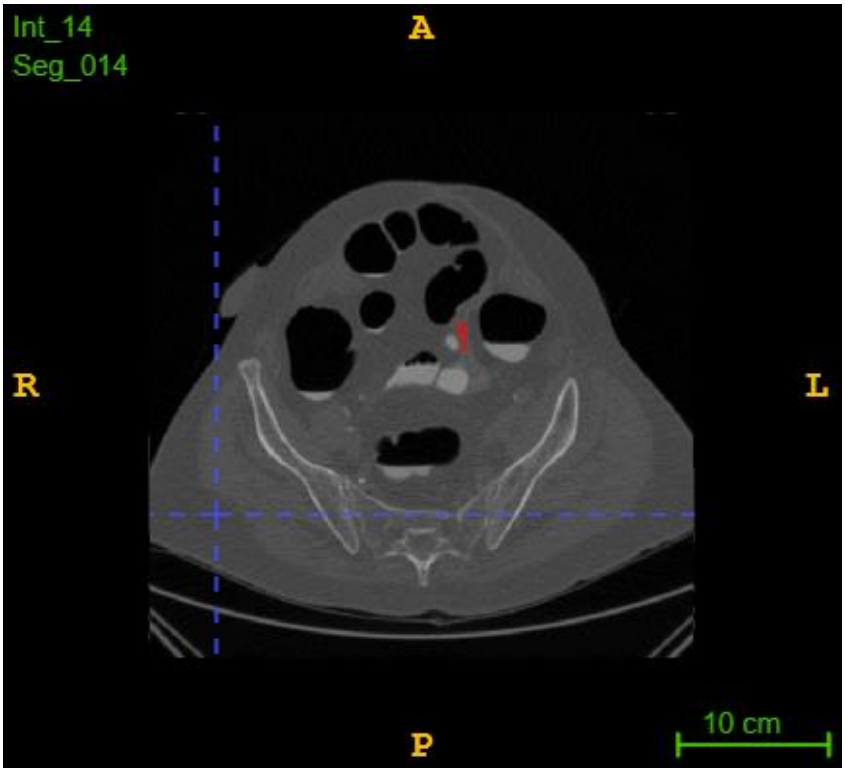

Figure 48: Example of slice with failed detection and segmentation

3.2.7 ACCEPTABLE CASES

The nnU-Net was also able to provide acceptable automatic segmentations, with DSC values above 0.4, even though in fewer cases in the Test Set and even less in the first nnU-Net. This was observed, as expected, mainly in the Training set used for learning.
In the picture below, an example of subject with diverticulitis and a DSC value of 0.78 is shown, in red the automatic segmentation and in white the manual mask. The nnUNet was able to perform a better automatic segmentation on subjects where the lesion occupied a larger area, in respect to the ones reported in the errors above.

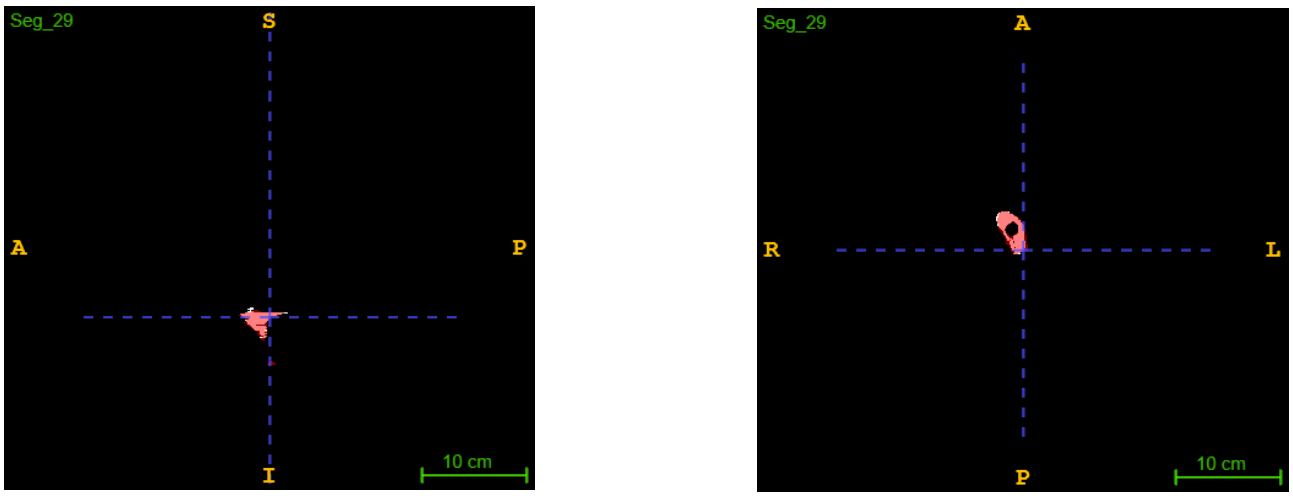

Figure 49: Sagittal and Horizontal view of an acceptable segmentation of diverticula

Cases of acceptable segmentation were also observed in subjects with colonic cancer, with a slight overestimation of the tumoral region. In the picture below, an example of a subject of the Training set with a DSC value of 0.75.

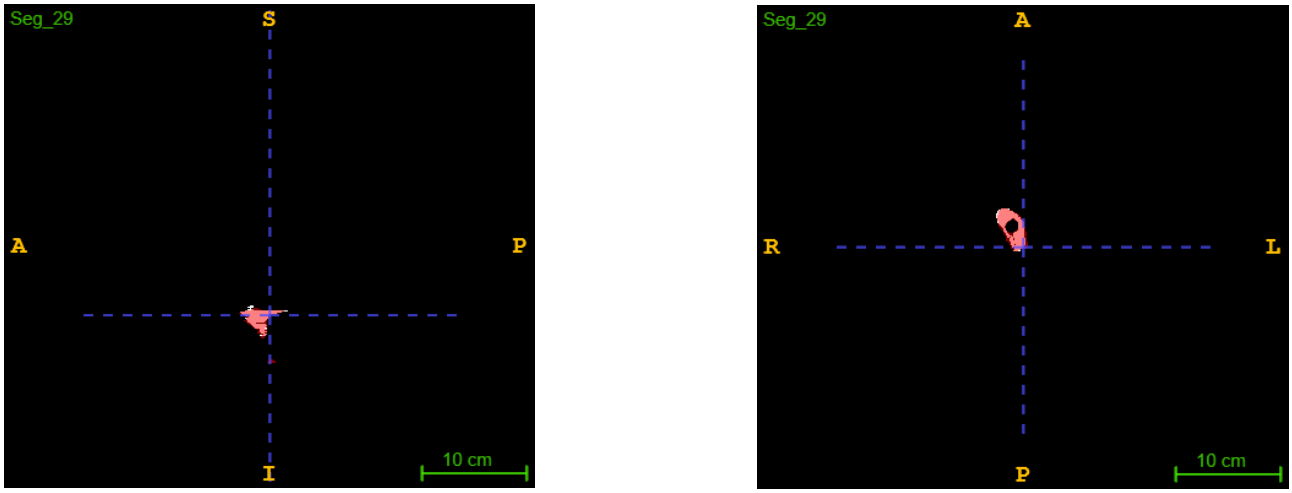

Figure 50: Sagittal and Horizontal view of an acceptable segmentation of the colon cancer

4 CONCLUSIONS

The aim of the thesis was to develop a deep learning method in order to detect and segment colon cancer and diverticula in CT scans. The task turned out to be difficult to solve due to various problems. The first one can be easily recognizable in the complexity of the segmentations, since the interested regions present many morphological similarities.
Furthermore, the difficulty to generalize, as in the capability to recognize lesions not belonging to the training set, was observed as it can be seen in the table below. The mean DSC values for the Test Set in the best attempt turned out to be very much lower than the ones obtained in the Training and Validation Set.

Set	Mean DSC	STD
Train	0,59	0,24
Validation	0,51	0,32
Test	0,24	0,22

Figure 51: Summary of mean and standard deviation of DSC values in the second nnU-Net

One of the probable causes was the use of a heavily unbalanced dataset, as the tumor or diverticula area was inferior to the area occupied by the surrounding tissues, and in addition the lesions were very small compared to the total CT volumes used for training the model, since they included lots of healthy colon slices. The work could have carried out better probably with a more balanced dataset and by focusing on a specific portion of the colon, such the sigmoid for the diverticula.
Further studies should focus on implementing more complex deep learning methods to accomplish the task, by collecting more CT volumes to train the model.

REFERENCES

[1] https://www.cancer.gov/publications/dictionaries/cancer-terms/def/colon
[2] Cindy L. Stanfield, “Fisiologia”, IV edizione, Cap.20, “Il Sistema Gastrointestinale”
[3] https://www.saintjohnscancer.org/gastrointestinal/conditions-we-treat/colorectal-cancer/about-the-colon/
[4] Menon G, Cagir B. Colon Cancer. In: StatPearls. StatPearls Publishing, Treasure Island (FL); 2025. PMID: 29262132.
[5] O. Ronneberger, P. Fischer, e T. Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation», in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, e A. F. Frangi, A c. di, in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
[6] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods, 18(2), 203-211.
[7] S. Innamorati, «DIAGNOSTIC CHALLENGE IN CT COLONOGRAPHY: ROLE OF AI IN THE DIFFERENTIAL DIAGNOSIS BETWEEN MASS-LIKE DIVERTICULAR DISEASE AND COLORECTAL CANCER», Università degli studi di Milano.
[8] Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel Boll, Joshy Cyriac, Shan Yang, Michael Bach, Martin Segeroth. TotalSegmentator: robust segmentation of 104 anatomical structures in CT images
[9] Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, Wells WM 3rd, Jolesz FA, Kikinis R. Statistical validation of image segmentation quality based on a spatial overlap index. Acad Radiol. 2004 Feb;11(2):178-89. doi: 10.1016/s1076-6332(03)00671-8. PMID: 14974593; PMCID: PMC1415224.
[10] Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. J Med Imaging (Bellingham). 2018 Jan;5(1):015006. doi: 10.1117/1.JMI.5.1.015006. Epub 2018 Feb 19. PMID: 29487883; PMCID: PMC5817231.
[11] Varoquaux G, Colliot O. Evaluating Machine Learning Models and Their Diagnostic Value. 2023 Jul 23. In: Colliot O, editor. Machine Learning for Brain Disorders [Internet]. New York, NY: Humana; 2023. Chapter 20. Available from: https://www.ncbi.nlm.nih.gov/books/NBK597473/ doi: 10.1007/978-1-0716-3195-9_20