POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Development of Deep Learning-based Algorithm for Segmentation of Bowel Lesions in Crohn's Disease Patients on MRE

Laurea Magistrale in Biomedical Engineering - Ingegneria Biomedica

Author: Mattia Cazzolla

Advisor: Prof. Riccardo Barbieri

Co-advisor: Riccardo Levi

Academic year: 2022-2023

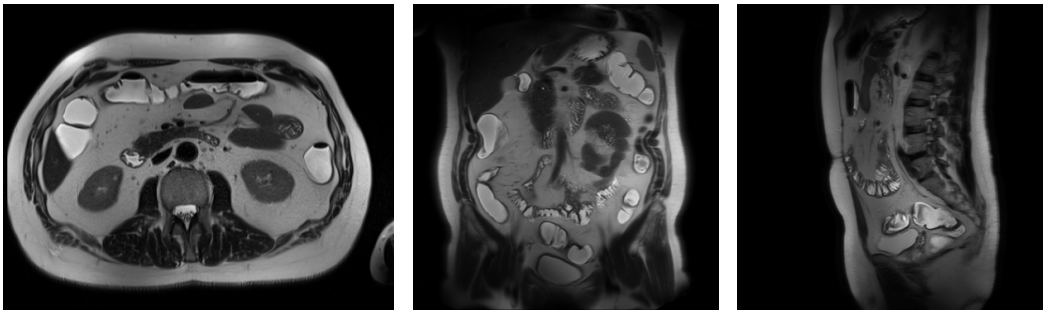
1. Introduction

Crohn's disease (CD) is a type of inflammatory bowel disease (IBD). It is characterized by transmural inflammation and skip lesions within the intestinal walls, often leading to significant bowel damage, surgical intervention, and disability [1].
The current diagnostic approach for CD relies on a non-strictly defined combination of clinical presentation, endoscopic appearance, radiology, histology, surgical findings, and serology [2].
Computed tomography enterography (CTE) and magnetic resonance enterography (MRE), have emerged as the standards for assessing the small intestine, providing insights into disease extension and activity through parameters such as wall thickness and contrast enhancement. While both modalities exhibit comparable accuracy in evaluating active inflammatory CD, to minimize radiation exposure MRE is preferred over CTE when available [2].
Nevertheless, the analysis of bowel MRE poses challenges due to the intricate nature of bowel wall structure and susceptibility to subjectivity and inter-observer variation. Computer-aided analysis based on AI techniques could help overcome such obstacles [3].
Accurate automatic detection of CD could help in rapid diagnosis, patient monitoring, and treatment planning [3]. Deep learning algorithms have been proven to have good feasibility in medical image segmentation and could be a powerful tool to detect CD efficiently.
While existing research has predominantly focused on applying deep learning to CTE images [4], there remains a notable gap in its application to MRE segmentation for CD.
The purpose of this work is to develop a deep learning automatic segmentation model for Crohn's Disease detection from MRE images. Specifically, our aims include the detection of the bowel wall affected by CD in the gastrointestinal tract utilizing state-of-the-art UNet deep learning architectures. Furthermore we aim to explore of the potential clinical utility of such technology as an aid to diagnosis.

2. Methods

2.1. Participants

The patients included in this study underwent MRE examinations at IRCCS - Istituto Clinico Humanitas (ICH) in Rozzano (Milan) between January 2013 and March 2022.



(a) Axial plane (b) Coronal plane (c) Sagittal plane

Figure 1: Example of axial, coronal, and sagittal center slices from T2-weighted MRE HASTE images

Specifically, inclusion criteria comprised patients who (1) received a diagnosis of Crohn's Disease, (2) underwent MRE, and (3) had at least one available T2 weighted image.
Exclusion criteria were applied based on (1) magnetic field strength, excluding cases with a strength different from 3 Tesla, and (2) the presence of incomplete imaging data, excluding cases lacking images in one or more anatomical planes.

2.2. MRE images

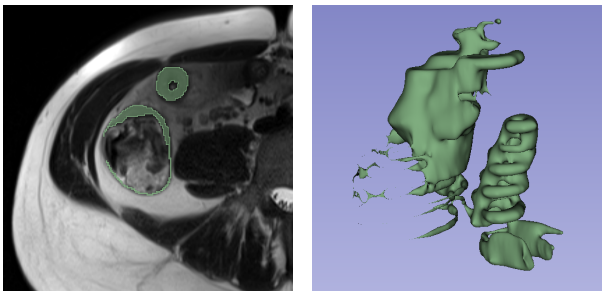
Each patient underwent an MRE imaging protocol, which included three volumetric 2D T2-weighted HASTE images, corresponding to the axial, coronal, and sagittal planes (Figure 1). The MRE scans were acquired using Siemens 3T *Verio* and Siemens 3T *Skyra fit* scanners.
Due to the use of different imaging protocols over the years, the acquisition parameters such as echo time, repetition time, and flip angle varied across the images.

2.3. Labels

The cross-sectional images obtained were manually segmented by a junior radiologist in each plane using *3D Slicer* software. Specifically the bowel wall of the disease-affected portions of the gastrointestinal tract were outlined. Figure 2 depicts a typical label in the dataset.

2.4. Segmentation Model

nnU-Net [5] is a deep learning biomedical segmentation model that automatically configures itself. The configuration includes data preprocessing, network architecture, training, and postprocessing, and is informed from domain knowledge and dataset fingerprinting.
nnU-Net follows a 3-dimensional (3D) U-Net-like architecture consisting of an encoder path and a decoder path connected by skip connections. nnU-Net has achieved state-of-the-art performance in many public biomedical datasets and competitions.
In this study, we applied nnU-Net for the 3D segmentation of CD with minimal modifications, primarily focusing on reducing the training duration to accommodate time and computational constraints.



(a) MRE image (b) 3D render

Figure 2: Example of a manual segmentation from MRE image (a), and its 3D rendering (b).

2.5. Training

Three separate nnU-Net models were trained, one for each dataset: axial, coronal, and sagittal. Each model was trained and validated with a 5-fold cross-validation. In each fold the models underwent training for 250 epochs, instead of the 1000 epochs proposed in the original paper.
During training, nnU-Net applies, by default, a range of online data augmentation techniques. Additionally, an option for a more intensive data augmentation was explored and compared against the default settings to assess its potential benefits in enhancing segmentation perfor-

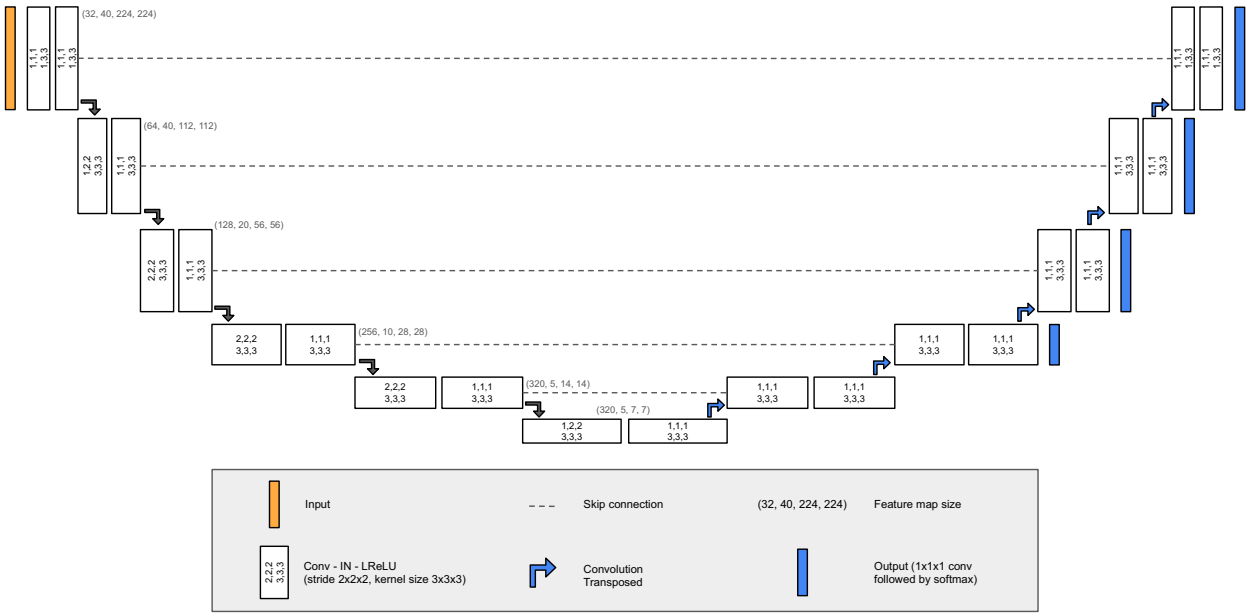


Figure 3: nnU-Net architecture for the axial dataset

mance, particularly considering the scarcity of the datasets.
All the models were trained on a Nvidia RTX 3090 GPU.

2.6. Validation Metrics

The models' performances were assessed using the following metrics:

- **Dice coefficient:** This metric quantifies the strict overlap between voxels in the predicted mask and the ground truth. It is defined as:

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

- **Dilated Dice coefficient:** Unlike the Dice coefficient, this metric allows for a small degree of uncertainty by dilating the segmentation masks by one voxel before comparison. It is defined as:

$$dDSC(A,B) = \frac{A \cap d(B) + B \cap d(A)}{A + B} \quad (2)$$

where $d(\cdot)$ represent dilation by one voxel.

For labels with a high ratio between surface and volume voxels, minor differences in segmentations can significantly impact overall overlap. The Dilated Dice coefficient offers greater flexibility and is therefore considered a more meaningful and reproducible metric for assessing complex shapes.

The overlap statistics were computed using the *Nighres* toolbox.

3. Results

3.1. Patient Cohort

The final cohort comprised 60 patients, consisting of 34 men and 26 women. The median age was 41 years, ranging from 19 to 80 years.
The dataset included a total of 61 MRE Series, as two MRE series were available from one patient. For some Series the label was not available in all three planes.
Table 1 presents the distribution of available images for training in each dataset.

Images available

Dataset	N of images
Axial	59
Coronal	59
Sagittal	60

Table 1: Number of images in each dataset

3.2. Segmentation Performances

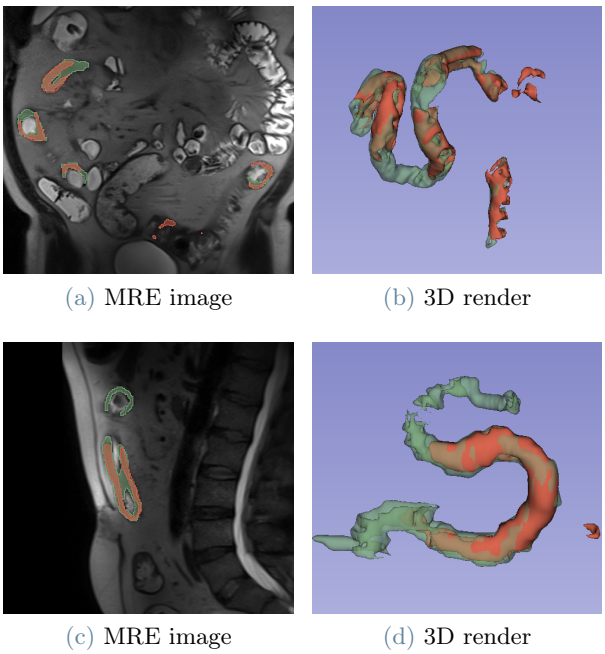
An example of architecture generated by nnU-Net is illustrated in Figure 3.
The overlapping metrics computed on the cross-validation test folds are summarized in Table 2.

Overlap metrics (mean ± std)

Set	Aug	DSC	dDSC
Ax	D	0.245 ± 0.281	0.328 ± 0.352
	I	0.243 ± 0.274	0.329 ± 0.339
Cor	D	0.201 ± 0.243	0.300 ± 0.337
	I	0.219 ± 0.252	0.321 ± 0.340
Sag	D	0.247 ± 0.287	0.340 ± 0.374
	I	0.260 ± 0.281	0.357 ± 0.368

Table 2: Dice and Dilated Dice Coefficient results. D and I refer to the Default and Intensive data augmentation.

The predictions obtained from models trained with intensive data augmentation were 17%, 40%, and 15% larger compared to the default one in the axial, coronal, and sagittal planes, respectively. The bigger prediction sizes led to a reduction in the number of cases where CD was not detected by 60%, 47%, and 13% in the three datasets, respectively. However, the difference in terms of overlapping metrics between the two data augmentation strategies is never statistically significant ($p > 0.05$).



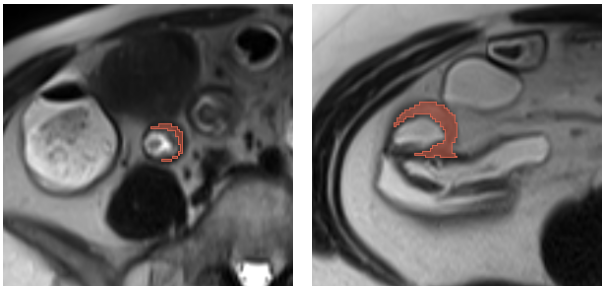
(a) MRE image (b) 3D render

(c) MRE image (d) 3D render

Figure 4: Examples of CD segmentation. In green the ground truth, in red the model prediction. (a,b) DSC=0.612, dDSC=0.817. (c,d) DSC=0.632, dDSC=0.789

3.3. Predictions

The predictions from patients with the 11 lowest segmentation accuracies underwent review by a senior radiologist. In eight of these cases, the senior radiologist confirmed that the models correctly segmented the disease in at least one of the three planes (Figure 5).



(a) Example 1 (b) Example 2

Figure 5: Examples of correctly segmented regions not present in the labels.

For these cases, the overlap metrics were recalculated (Table 3) by modifying the original labels to include the predictions as ground truth. This adjustment represents the most optimal scenario, assuming a perfect segmentation for the additional portion of the disease included in the original label.

Overlap metrics (mean ± std)

Set	Type	DSC	dDSC
Ax	O	0.243 ± 0.274	0.329 ± 0.339
	R	0.323 ± 0.286	0.413 ± 0.330
Cor	O	0.219 ± 0.252	0.321 ± 0.340
	R	0.279 ± 0.272	0.384 ± 0.339
Sag	O	0.260 ± 0.281	0.357 ± 0.368
	R	0.295 ± 0.284	0.397 ± 0.364

Table 3: Dice and Dilated Dice Coefficient results. O and R refer to the Original coefficients and the Reviewed ones.

4. Discussion

This study utilized nnU-Net neural network to develop automatic segmentation models for Crohn's Disease. The resulting segmentation achieved a Dice coefficient of 0.323, 0.279, and 0.295 and a Dilated Dice coefficient of 0.413,

0.384, and 0.397 on the axial, coronal, and sagittal datasets, respectively. Training with intensive data augmentation reduced the number of cases where no disease was detected, making it the preferred approach.
Other works present in literature are often intricate, slow or rely on semi-automated approaches that require physician intervention. The only work that used the same UNet approach as ours [4] achieved high segmentation accuracy, with a resulting Dice coefficient of 0.824 on CTE images. However, CTE does not represent a viable imaging modality for CD, as patients, often young, require frequent acquisitions to monitor disease extent and treatment response.
To our knowledge, no other studies have been conducted on developing a deep learning model for the automatic segmentation of CD from MRE images.
Radiologists would be the primary beneficiaries of the automatic segmentation of CD. The use of such technology could serve as an aid to diagnosis, particularly in clinical centers not specialized in Crohn's Disease and for junior radiologists, who may lack sufficient experience to identify all affected portions of the gastrointestinal tract, as highlighted in our study. *Gao et al.* [4] demonstrated the time-saving benefits such technology can offer junior radiologists during diagnosis.
However, this work faces limitations. The heterogeneity of image acquisition sequences and resultant variation in image quality pose challenges for such a small dataset. Additionally, addressing the quality of labels is crucial. Without reliable ground truth annotations, effectively training the models, assessing their performance, and diagnosing potential sources of error become challenging.

5. Conclusions

We have developed advanced deep learning models based on the nnU-Net neural network for automatic segmentation of Crohn's Disease in MRE images.
This technology can assist radiologists in quickly and accurately detecting the extension of the disease. Further exploration in this direction holds significant potential for enhancing segmentation accuracy and efficiency.
Moreover, integrating radiomics techniques with the segmentation task presents an opportunity to broaden the scope of CD diagnosis and management. By providing valuable insights into disease severity evaluation and treatment response, this approach can improve personalized management strategies for CD patients.

References

[1] J. Torres, S. Mehandru, J.-F. Colombel, and L. Peyrin-Biroulet, "Crohn's disease," *The Lancet*, vol. 389, pp. 1741–1755, Apr. 2017.

[2] F. Gomollón, A. Dignass, V. Annese, H. Tilg, G. Van Assche, J. O. Lindsay, L. Peyrin-Biroulet, G. J. Cullen, M. Daperno, T. Kucharzik, F. Rieder, S. Almer, A. Armuzzi, M. Harbord, J. Langhorst, M. Sans, Y. Chowers, G. Fiorino, P. Juillerat, G. J. Mantzaris, F. Rizzello, S. Vavricka, P. Gionchetti, and on behalf of ECCO, "3rd european evidence-based consensus on the diagnosis and management of crohn's disease 2016: Part 1: Diagnosis and medical management," *Journal of Crohn's and Colitis*, vol. 1, pp. 3–25, 9 2016.

[3] G. Grassi, M. E. Laino, M. C. Fantini, G. M. Argiolas, M. V. Cherchi, R. Nicola, C. Gerosa, G. Cerrone, L. Mannelli, A. Balestrieri, J. S. Suri, A. Carriero, and L. Saba, "Advanced imaging and Crohn's disease: An overview of clinical application and the added value of artificial intelligence," *European Journal of Radiology*, vol. 157, p. 110551, Dec. 2022.

[4] Y. Gao, B. Zhang, D. Zhao, S. Li, C. Rong, M. Sun, and X. Wu, "Automatic Segmentation and Radiomics for Identification and Activity Assessment of CTE Lesions in Crohn's Disease," *Inflammatory Bowel Diseases*, p. izad285, Nov. 2023.

[5] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203–211, Feb. 2021.