Bern University of Applied Sciences

UNIVERSITÄT BERN

Faculty of Medicine
Biomedical Engineering

Master of Science Thesis

Title: Development of a Deep Learning-Based Model for the Detection of Extracapsular Extensions in Head and Neck

by

Léandre Cuenot

of Switzerland

Supervisors
Prof. Dr. Mauricio Reyes and Dr. Daniel Schanne

Institutions
ARTORG Center for Biomedical Research, University of Bern, Medical Image Analysis Group (MIA)

Examiners
Prof. Dr. Mauricio Reyes and Dr. Daniel Schanne

Bern, October 2024

This report is confidential. (Delete this statement in the file "unibe-msc.cls" if the report is not confidential.)

Abstract

The abstract should provide a concise (300-400 word) summary of the motivation, methodology, main results and conclusions. For example:

Osteoporosis is a disease in which the density and quality of bone are reduced. As the bones become more porous and fragile, the risk of fracture is greatly increased. The loss of bone occurs progressively, often there are no symptoms until the first fracture occurs. Nowadays as many women are dying from osteoporosis as from breast cancer. Moreover it has been estimated that yearly costs arising from osteoporotic fractures alone in Europe worth 30 billion Euros.

Percutaneous vertebroplasty is the injection of bone cement into the vertebral body in order to relieve pain and stabilize fractured and/or osteoporotic vertebrae with immediate improvement of the symptoms. Treatment risks and complications include those related to needle placement, infection, bleeding and cement extravazation. The cement can leak into extraosseous tissues, including the epidural or paravertebral venous system eventually ending in pulmonary embolism and death.

The aim of this project was to develop a computational model to simulate the flow of two immiscible fluids through porous trabecular bone in order to predict the three-dimensional spreading patterns developing from the cement injection and minimize the risk of cement extravazation while maximizing the mechanical effect. The computational model estimates region specific porosity and anisotropic permeability from Hounsfield unit values obtained from patient-specific clinical computer tomography data sets. The creeping flow through the porous matrix is governed by a modified version of Darcy's Law, an empirical relation of the pressure gradient to the flow velocity with consideration of the complex rheological properties of bone cement.

To simulate the immiscible two phase fluid flow, i.e. the displacement of a biofluid by a biomaterial, a fluid interface tracking algorithm with mixed boundary representation has been developed. The nonlinear partial differential equation arising from the problem was numerically implemented into the open-source Finite Element framework *libMesh*. The algorithm design allows the incorporation of the developed methods into a larger simulation of vertebral bone augmentation for pre-surgical planning.

First simulation trials showed close agreement with the findings from relevant literature. The computational model demonstrated efficiency and numerical stability. The future model development may incorporate the morphology of the region specific trabecular bone structure improving the models' accuracy or the prediction of the orientation and alignment of fiber-reinforced bone cements in order to increase fracture-resistance.

Acknowledgements

Here you may include acknowledgements.

1. Please sign the following declaration if you did **not** use AI tools (like ChatGPT or DeepL)

„Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Ich erkläre weiter, dass ich keine unerlaubten Hilfsmittel verwendet habe, namentlich keine weiteren Personen mir beim Verfassen der Arbeit geholfen haben und ich keine Technologien der Künstlichen Intelligenz eingesetzt habe. Mir ist bekannt, dass andernfalls die Arbeit mit der Note 1 bewertet wird bzw. der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.
Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die schriftliche Arbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.“

1. Please sign the following declaration if you **did** use AI tools (like ChatGPT or DeepL). The use of AI tools must be permitted by your supervisor.

„Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Als Hilfsmittel habe ich Künstliche Intelligenz verwendet. Sämtliche Elemente, die ich von einer Künstlichen Intelligenz übernommen habe, werden als solche deklariert und es finden sich die genaue Bezeichnung der verwendeten Technologie sowie die Angabe der «Prompts», die ich dafür eingesetzt habe. Mir ist bekannt, dass andernfalls die Arbeit mit der Note 1 bewertet wird bzw. der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.
Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die schriftliche Arbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.“

Bern, October 27$^{\text{th}}$ 2024

Léandre Cuenot

Contents

Chapter 1

Introduction

Extracapsular extension (ECE) in head and neck cancers refers to the spread of metastatic tumors beyond the lymph node capsule into the surrounding connective tissue. ECE is associated with a poorer prognosis, significantly affecting treatment strategies and reducing overall survival rates in affected patients [4][7]. Accurate early detection of ECE is crucial for optimizing patient management and treatment planning. Proper identification of ECE can guide therapeutic decisions, including the potential benefits of adjunctive treatments such as postoperative concurrent chemoradiotherapy, which may improve outcomes for patients at higher risk due to ECE [6].

Currently, definitive diagnosis of ECE can only be confirmed through postoperative pathology, limiting its clinical utility [2]. In practice, contrast-enhanced computed tomography (CT) is used to detect ECE, relying heavily on physician expertise. Literature indicates that the sensitivity of CT for detecting ECE ranges from 18.8% to 72.2%, with higher sensitivity in more advanced cases. Sensitivity increases from 18.8% in early-stage (grade 1-2) ECE to 72.2% in advanced-stage (grade 4) ECE [8]. This wide range reflects considerable inter-observer variability.

Recent advancements in deep learning have shown significant potential in improving ECE detection compared to manual expertise [2][10][3]. Deep learning algorithms offer more consistent assessments and address the high inter-observer variability observed among human experts. Notably, the HECKTOR challenge, held annually from 2020 to 2022, focused on enhancing segmentation tasks through deep learning. The challenge aimed to develop models for segmenting the primary gross tumor volume (GTVp) and metastatic lymph nodes (GTVn) in the head and neck region. The results were promising, with a substantial majority of participants achieving an aggregate Dice similarity coefficient greater than 0.70 for both GTVp and GTVn, and the top-performing model achieving a Dice score of 0.788, underscoring the efficacy of deep learning approaches in complex segmentation tasks.

In this study, we propose an approach inspired by participants of the HECKTOR challenge, utilizing the same dataset, which integrates positron emission tomography (PET) imaging alongside conventional computed tomography (CT) imaging. PET imaging provides metabolic insights that complement the anatomical information from CT, potentially improving the detection of extracapsular extension (ECE) in head and neck cancers. By leveraging this multimodal approach, we aim to enhance diagnostic accuracy and provide more reliable predictions of ECE, thereby supporting more informed clinical decision-making.

The majority of the top-performing models in this domain have employed ensembles of 3D U-Net architectures. In this work, we aim to assess the robustness of 3D U-Nets under various perturbations that may occur in real-world clinical scenarios. Specifically, we will analyze how different tumor characteristics correlate with the degree to which these perturbations affect segmentation performance.

Additionally, we will link these findings to a relevant clinical application by comparing the model's performance with expert physicians' evaluations. This comparison will explore the correlation between the Dice similarity coefficient, which quantitatively assesses the segmentation accuracy of the deep learning models, and the qualitative assessment provided by physicians. Existing literature has demonstrated a moderate correlation between the Dice similarity coefficient and physician evaluations, with values ranging from 0.36 to 0.5 depending on the anatomical location of the segmented area [5].

Moreover, physicians will also evaluate cases with artificially introduced perturbations to examine whether changes in Dice scores align with variations in their clinical grading. This analysis will provide insights into the relationship between perturbation-induced segmentation errors and the clinical evaluation of ECE.

The primary focus of this study is on the segmentation aspect of the ECE prediction baseline. To enhance the overall predictive model, future work will extend the analysis to the classification component, aiming to improve the model's ability to predict ECE with greater accuracy.

Chapter 2

Methods

2.1 Hecktor dataset

For the segmentation model in this project, we utilized the dataset from the Hecktor Challenge 2022. This dataset consists of both training images with corresponding ground truth labels and test images without labels. For our purposes, we focused solely on the training subset due to the availability of ground truth annotations. This subset comprises 524 imaging cases collected from seven distinct clinical centers, including:

- CHUM: Centre Hospitalier de l'Université de Montréal, Montréal, Canada
- CHUP: Centre Hospitalier Universitaire de Poitiers, France
- CHUS: Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Canada
- CHUV: Centre Hospitalier Universitaire Vaudois, Switzerland
- HGJ: Hôpital Général Juif, Montréal, Canada
- HMR: Hôpital Maisonneuve-Rosemont, Montréal, Canada
- MDA: MD Anderson Cancer Center, Houston, Texas, USA

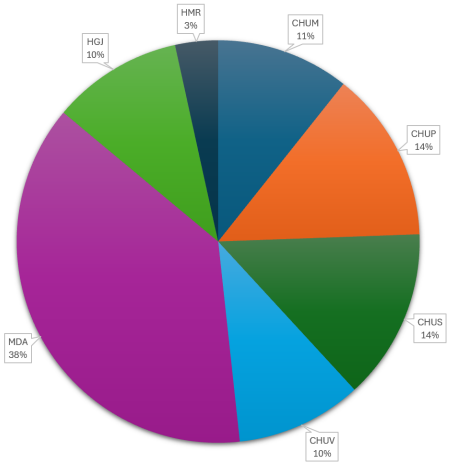
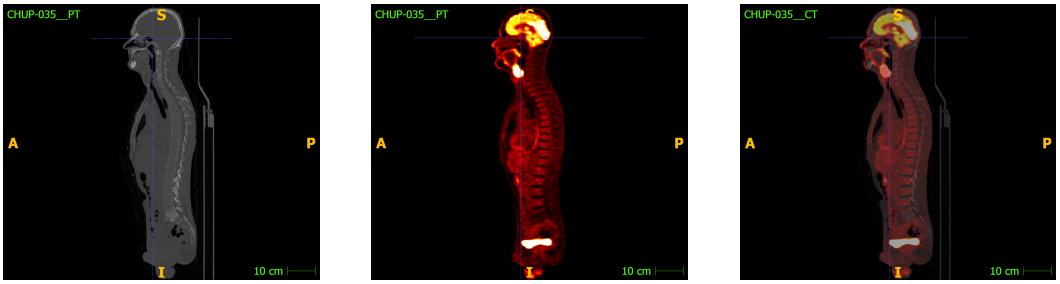


Figure 2.1. Distribution of cases across sites

Each case in the dataset contains two imaging modalities: computed tomography (CT), positron emission tomography (PET), and ground truth labels. The PET images were standardized using the Standardized Uptake Value (SUV). The CT and label images are provided in NIfTI format with a spatial resolution of 524 × 524 pixels in the axial plane, and variable depths across slices. While some CT images focus exclusively on the head and neck region, others encompass the entire body. In contrast, the PET images have a resolution of 128 × 128 pixels in the axial plane, with varying depths similar to the CT images.



(a) CT Scan **(b)** PET Scan **(c)** CT + PET and label

Figure 2.2. An example cases of the CHUP

2.2 Segmentation model

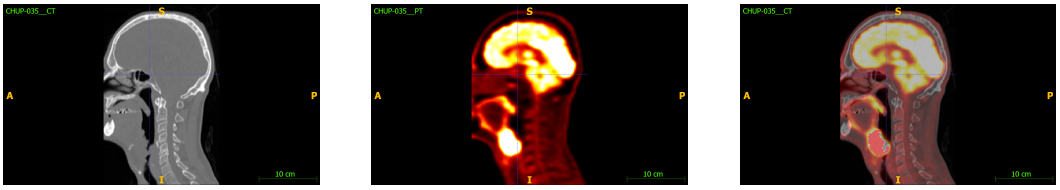
2.2.1 Preprocessing

The initial preprocessing step involved resampling the PET images to achieve uniform dimensions across modalities. Specifically, PET images were resampled to an axial plane resolution of 524×524 pixels. Some labels exhibited minor dimensional discrepancies, occasionally differing by one plane in either width or height. These discrepancies were rectified following the resampling process.

Subsequently, after ensuring that all three imaging modalities (CT, PET, and label) were aligned to the same dimensions, the images were resampled to a common isotropic voxel size of 1×1×1 mm. This resampling was performed to facilitate subsequent cropping of the head and neck region.

For the cropping procedure, the center of the head was identified using the contours of the brain on the PET scan. Based on this central reference point, a subvolume of 200×200×[maximum of 310 pixels] was extracted. This approach minimizes the computational burden associated with background regions devoid of ground truth information. Additionally, the images underwent normalization via z-score clipping to mitigate the effects of outliers.

The processed images were saved in NIfTI format following the above steps. Further preprocessing techniques were applied during the training phase, which will be detailed in subsequent sections. Aspects of the preprocessing procedure were adapted from the methods employed by the winning team of the Hecktor Challenge 2022 [7]



(a) CT Scan **(b)** PET Scan **(c)** CT + PET and label

Figure 2.3. Same example cases after preprocessing

2.2.2 Data augmentation

During the training process, data augmentation was applied to the original images using the Medical Open Network for Artificial Intelligence (MONAI) framework. Spatial augmentations were applied to both imaging modalities, including random flipping, affine transformations (translation, rotation, and scaling). For the CT images, intensity augmentations were also incorporated, such as the addition of Gaussian noise, smoothing, contrast adjustment, and intensity shifting. All augmentations were applied with an occurence probability of 20%.

Following augmentation, the images were randomly cropped into patches of size 192×192×192 voxels. Patches were centered based on labels, with a 10% probability of being centered on background, 45% on primary tumors, and 45% on nodal tumors. In cases where only one tumor type was present, the sampling probability for that tumor was increased to 90%.

For validation, a single patch of equal size was extracted, with a balanced distribution of 33% for background, primary tumors, and nodal tumors. This approach eliminates the need for predictions using sliding windows (e.g., 8 predictions per image), as a single patch is used for each validation image of size 200×200×310.

2.2.3 Model architectures

The model used for segmentation is the Dynamic UNet (DynUNet) from MONAI. It provide several parameter to configure the model. The architecture is represented below :

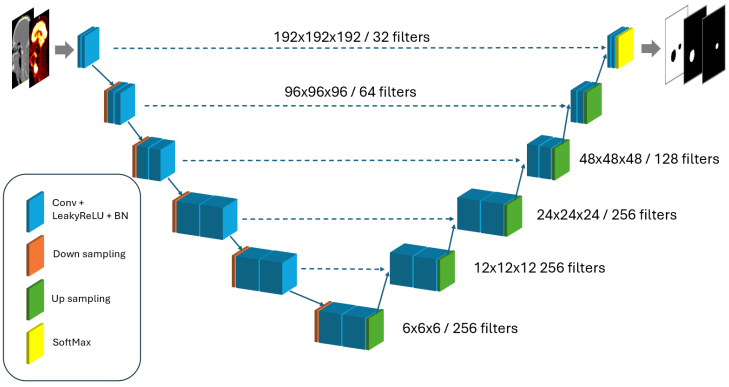


Figure 2.4. Model architecture

The model operates in three dimensions, taking as inputs two modalities: computed tomography (CT) and positron emission tomography (PET). It produces three binary segmentation outputs: background, primary tumors, and nodal tumors.

All convolutional kernels used in the model are of size 3×3×3. The architecture consists of six layers, with downsampling performed consistently by a factor of 2, reducing the input size from 192×192×192 to 6×6×6. Additionally, batch normalization is implemented in conjunction with residual blocks to enhance training stability and model performance. The probabilites are then computed through softmax that incorporate the mutual exclusivity between the 3 segmentation masks.

2.2.4 Training

The training parameters for the model are as follows:

- Optimizer : AdamW
- Learning rate : 1e-4
- Weight decay : 3e-5
- Batch size : 2
- Epochs : 100

A LambdaLR scheduler is employed, utilizing the following decay function:

$$\text{Decay Factor} = \left(1 - \frac{\text{epoch}}{\text{number\_epoch}}\right)^{0.9} \tag{2.1}$$

The loss function combines Dice loss and cross-entropy loss, defined as follows:

$$\text{Dice Loss} = 1 - \frac{2|P \cap Y|}{|P| + |Y|} \tag{2.2}$$

$$\text{Cross-Entropy Loss} = -\left(Y \log(P) + (1 - Y) \log(1 - P)\right) \tag{2.3}$$

where P is the prediction and Y is the ground truth.

Dice loss is particularly advantageous for segmentation tasks involving imbalanced datasets, such as those found in head and neck tumor imaging, where the tumors are often significantly smaller than the surrounding tissues. The Dice coefficient, which Dice loss is based on, focuses on the overlap between the predicted segmentation and the ground truth (2.2), effectively capturing the regions of interest despite their limited size.

In contrast, cross-entropy loss measures the dissimilarity between the predicted probability distribution and the true distribution (2.3). While it is easier to optimize due to its differentiable nature, cross-entropy can be more sensitive to class imbalances. This sensitivity may lead to suboptimal performance when the class of interest (e.g., tumors) constitutes only a small fraction of the total data.

Given these characteristics, a hybrid approach that combines both Dice loss and cross-entropy loss offers a promising solution. This combination leverages the strengths of both methods: Dice loss ensures that the model pays adequate attention to the small tumor regions, while cross-entropy facilitates stable and efficient optimization. Such an approach can yield improved segmentation performance, making it particularly suitable for complex tasks in medical imaging.

To enhance training performance, mixed precision training utilizing PyTorch has been implemented. This approach leverages both single-precision (32-bit) and half-precision (16-bit) floating-point formats during model training. By employing mixed precision, the training speed is significantly increased due to reduced computational overhead, while the memory footprint is minimized. This allows for the efficient use of hardware resources, facilitating larger batch sizes and improved model scalability without compromising numerical stability. The training run on a TODO with around 30Go of memory use.

To enhance model performance, the training dataset has been partitioned into five folds representing 90% (with 80% for training and 20% for validation) of the initial 524 cases, facilitating a cross-validation approach. This technique allows for the integration of predictions from each fold, thereby improving the consistency and robustness of the model's predictions. By leveraging multiple subsets of the data during training, the model can better generalize to unseen data, ultimately leading to more reliable and stable results. This method not only mitigates overfitting but also enhances the overall accuracy of the predictions.

For the validation,

2.2.5 Inference

2.3 Analysis

2.3.1 Robustness

Metrics

2.3.2 Properties

Metrics

2.3.3 Clinical evaluation

Chapter 3

Results

3.1 A Sample Section with a Table

Chapter 4

Discussion and Conclusions

4.1 Discussion

Interpret your results in the context of past and current studies and literature on the same topic. Attempt to explain inconsistencies or contrasting opinion. Highlight the novelty of your work. Objectively discuss the limitations.

4.2 Conclusions

Formulate clear conclusions which are supported by your research results.

Chapter 5

Outlook

Provide a vision of possible future work to continue and extend your thesis research.

Bibliography

[1] V. Andrearczyk, V. Oreiller, and et al. Overview of the hecktor challenge at miccai 2022: Automatic head and neck tumor segmentation and outcome prediction in pet/ct. Lecture Notes in Computer Science, 13626(2):1–30, March 2023.

[2] B. H. Kann, S. Aneja, G. V. Loganadane, J. R. Kelly, S. M. Smith, R. H. Decker, J. B. Yu, H. S. Park, W. G. Yarbrough, A. Malhotra, B. A. Burtness, and Z. A. Husain. Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. Scientific Reports, 8(5):14036, Sept. 2018.

[3] B. H. Kann and J. L. et al. Screening for extranodal extension in hpv-associated oropharyngeal carcinoma: evaluation of a ct-based deep learning algorithm in patient data from a multicentre, randomised de-escalation trial. Lancet Digital Health, 5(8):e360–e369, 2023.

[4] H. R. Kelly and H. D. Curtin. Squamous cell carcinoma of the head and neck: Imaging evaluation of regional lymph nodes and implications for management. Seminars in Ultrasound, CT, and MR, 38(1):466–478, Oct. 2017.

[5] F. Kofler, I. Ezhov, F. Isensee, and B. M. et al. Estimates of human expert perception for cnn training beyond rolling the dice coefficient. Melba, 2(9):27–71, May 2023.

[6] V. Krstevska. Evolution of treatment and high-risk features in resectable locally advanced head and neck squamous cell carcinoma with special reference to extracapsular extension of nodal disease. Journal of BUON, 20(6):943–953, July 2015.

[7] A. Myronenko, M. M. R. Siddiquee, D. Yang, Y. He, and D. Xu. Automated head and neck tumor segmentation from 3d pet/ct: Hecktor 2022 challenge report. In Proceedings of the HECKTOR 2022 Challenge, pages 31–37. Lecture Notes in Computer Science (LNCS, Volume 13626), 2023.

[8] R. S. Prabhu and K. R. M. et al. Accuracy of computed tomography for predicting pathologic nodal extracapsular extension in patients with head-and-neck cancer undergoing initial surgical resection. International Journal of Radiation Oncology Biology Physics, 88(3):122–129, Jan. 2014.

[9] T. V. Thomas, M. R. Kanakamedala, E. Bhanat, A. Abraham, E. Mundra, A. A. Albert, S. Giri, R. Bhandari, and S. Vijayakumar. Predictors of extracapsular extension in patients with squamous cell carcinoma of the head and neck and outcome analysis. Cureus, 13(4):e16680, July 2021.

[10] Y. S. e. a. Yoshiko Ariji. Ct evaluation of extranodal extension of cervical lymph node metastases in patients with oral squamous cell carcinoma using deep learning classification. Oral Radiology, 36(7):148–155, Apr. 2020.

Appendices

Appendix A

Vector and Tensor Mathematics

A.1 Introduction

...

A.2 Variable Types

...

Appendix B

Another Appendix

B.1 Section 1

...

B.2 Section 2

...