
# LUNG CANCER SEGMENTATION IN PET-CT SCANS USING DEEP LEARNING

BRAM MEIJER

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBER

702065

COMMITTEE

dr. Sharon Ong & Martijn van Leeuwen
drs. Stijn Rotman

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

June 23, 2023

WORD COUNT

7911

ACKNOWLEDGMENTS

I would like to thank Dr. Sharon Ong and Martijn van Leeuwen for their involvement and helpfulness during the semester. They provided me with many useful resources and highly relevant feedback during our biweekly meetings. They motivated me throughout the entire process and I really enjoyed collaborating with them during this semester.

# LUNG CANCER SEGMENTATION IN PET-CT SCANS USING DEEP LEARNING

BRAM MEIJER

**Abstract**

In recent years, deep learning has become the most preferred option for the procedure of automatic medical image segmentation. The goal of this thesis is to examine the extent to which deep learning can be applied to segment lung cancer tumors in PET-CT scans. Often, data obtained from computed tomography (CT) scans are used to train a medical image segmentation model. In this thesis, the effect of the inclusion of positron emission tomography (PET) data as single input, as well as the incorporation of multimodal PET-CT input, on model performance was investigated. Additionally, the performance of two commonly used deep learning architectures, a Convolutional Neural Network (CNN) and U-Net, were compared. The main results demonstrated that PET and PET-CT data provided a beneficial effect on segmentation and detection performance, obtaining Dice coefficient and detection rate scores over 60%. The usage of CT data resulted in a poorer segmentation and detection performance, obtaining scores under 30%. Despite these promising results, some limitations such as overfitting and high false positive rates were observed. This work extensively compares the effects of CT, PET, and PET-CT data on segmentation and detection performance on a dataset that has not been widely explored. Using both PET and CT data as input for a deep learning model, clinicians can more accurately identify the presence and spread of lung cancer.

## 1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

The data have been acquired from the University Hospital Tübingen and the University Hospital of the Ludwig Maximilian University through an online request. The obtained data is anonymised. Work on this thesis did not involve collecting data from human participants or animals. The original owner of the data used in this thesis retains ownership of the data during and after the completion of this thesis. However, the institutions

were informed about the use of this data for this thesis and potential research publications.

All the figures in this work belong to the author. The thesis code can be accessed through the ETZ hospital server. In terms of writing, the author used assistance with the language of the paper: a thesaurus (Name: Thesaurus.com, source: https://www.thesaurus.com/) was used to improve the author's original content, for paraphrasing. No typesetting tools or services were used.

# 2 INTRODUCTION

## 2.1 *Project Definition*

Cancer refers to a collection of more than 100 unique diseases. In cancer cells, there is an uncontrolled reproduction of malignant cells, which can invade nearby tissue. Over time, these malignant cells can form clusters at locations within the body, known as tumors, which fatally affect the human survival by disrupting important organ processes (Weinberg, 1996). Lung cancer is the most commonly diagnosed type of cancer worldwide (Bomanji et al., 2001). Despite considerable increments in survival rate for most cancer types in the United States, there has only been a modest increase in survival rate for lung cancer patients (Schabath & Cote, 2019). This can be attributed to the fact that in the majority of cases, the diagnosis occurs in a late stage when survival chances are minimal. Effective medical imaging techniques play a crucial role in the early detection and treatment of lesions and cancerous tumors (Thapar et al., 2022).

Positron emission tomography (PET) is a functional imaging technique in which metabolic activity is quantified, making it possible to detect tumors that display a high metabolic activity (Grossiord et al., 2017). To do this, in a medical setting, positrons are inserted into a tracer known as F-fluorodeoxyglucose (F-FDG) (Townsend, 2008), which is injected into the patient body. Inside the body, the F-FDG tracer decays due to positron emission: these positrons quickly lose their energy and interact with an electron, resulting in positron-electron annihilation (Raichle, 1983). This process causes two annihilation photons to travel in opposite directions with a relatively high energy. Areas with a high rate of annihilation energy correspond to areas with high metabolic activity (Raichle, 1983). PET has the advantage of being highly sensitive and specific to lesions with a high metabolic activity (Wang et al., 2022).

In computed tomography (CT), X-rays are used to create cross-sectional images of bones and tissue inside the body (Xi et al., 2020). To accomplish this, a patient is placed on a table inside a tunnel that spins around the

patient's body and releases X-ray beams. The radiation that is absorbed by the patient's body is measured. A CT scan has the advantage of providing accurate anatomical localization of organs and lesions (Townsend, 2008).

This thesis is done in collaboration with the Elisabeth-TweeSteden Hospital (ETZ) in Tilburg, the Netherlands. To be specific, in this thesis, a deep learning-based medical image segmentation procedure on full-body PET-CT scans will be developed to detect and segment lung cancer on PET and/or CT modalities. The inclusion of both PET and CT modalities, and specifically the effects of the incorporation of PET data, will be comprehensively examined and investigated. Additionally, two commonly used deep learning methods in medical image segmentation - U-Net and CNN - will be deployed and compared. The models developed in this thesis will be adapted for automatic image segmentation of bone lesions, which will be incorporated into the ETZ workflow.

## 2.2 *Societal Motivation*

The American National Cancer Institute has estimated that the number of new annual cancer cases will increase from 18 million in 2018 to nearly 30 million in 2040, with the number of deaths increasing from 9.5 million to 16.4 million (Thapar et al., 2022). Globally, lung cancer is the most common cause of cancer-related death for males and the second most common cause of cancer-related death for women (Schabath & Cote, 2019). To increase the survival risk of lung cancer patients, early staging and identification of the disease is highly important (Bomanji et al., 2001). Hence, deep learning-based lesion segmentation approaches can be of life-saving importance. To be specific, deep-learning based approaches offer many benefits in the detection of lung cancer, such as:

- They can aid clinicians in concentrating on a specific area of the disease and help select comprehensive information about the disease for a more precise conclusion (Jha et al., 2020). Moreover, a deep learning approach with both PET and CT data as model input can swiftly and efficiently identify the presence and spread of lung cancer in the thoracic area (Hochhegger et al., 2015)
- Deep learning-based methodologies reduce the issues of intra-reader and inter-reader variability (Osadebey et al., 2021) by providing a more consistent diagnosis of cancer and lesions than subjective judgments of radiologists (Chen et al., 2021)
- An automated deep learning approach can assist in early detection of lung cancer tumors, increasing the survival rate for lung cancer patients from 14% to 49% (Kalaivani et al., 2020)

## 2.3 *Scientific Motivation*

Medical image segmentation is the process of pixel-wise labeling of an image to determine whether a single pixel belongs to a certain target structure such as an organ or tumor, or should be treated as background (Jha et al., 2020). This process is highly important in the biomedical domain, since malignant data should be properly localized (Ronneberger et al., 2015) in order to quickly determine an effective treatment. PET and CT play an important role in tumor diagnosis and staging, monitoring of treatment effectiveness, and radiotherapeutic planning (Saif et al., 2010). Compared to single modality PET and CT scanners, multimodal PET/CT scanners have a higher accuracy in detecting diseases such as lung cancer, melanoma, and lymphoma (Townsend, 2008). Hence, combining PET and CT as input to a deep learning model can result in a computationally efficient segmentation procedure that is able to use both the sensitivity of CT towards a target structure's anatomical location and the sensitivity of PET towards a region's metabolic activity (Wang et al., 2022).

## 2.4 *Research Questions*

Despite the promising abilities of deep learning, it also has its flaws. Some of the most common issues related to these approaches are limited annotated data, imbalance between foreground and background classes, overfitting, training time, and a vague organ or lesion appearance (Hesamian et al., 2019). Various possible strategies to overcome these obstacles have been applied, ranging from very effective to having a negligible effect. In this thesis, the extent to which deep learning-based tumor segmentation approaches can be utilized in PET/CT scans will be addressed and answered. To be exact, the main research question of this work will be:

> **To what extent can a deep learning-based approach segment lung cancer tumors in PET-CT scans?**

Moreover, both PET and CT imaging do not come without their flaws. When compared to CT images, PET images have poorer resolution (Fu et al., 2021). Often, false positives are encountered because of a similar level of F-FDG accumulation from organs and tumors, meaning that a high F-FDG acquisition is not exclusive to tumors (Wang et al., 2022). PET images also suffer from a poor ability to visually represent anatomical regions compared to CT (Townsend, 2008). PET is more time consuming than CT and the minimum infrastructure necessary is considerably higher (Bomanji et al., 2001). Consequently, some hospitals, including ETZ, only have facilities in place for CT scanners. CT images, on the other hand,

suffer from a possible issue of irregularities: in COVID-19 patients, for example, the appearances of different lung lesions in CT images vary a lot (Chen et al., 2021). Additionally, unlike PET, CT imaging is not able to detect lesions' metabolic activity.

Since both modalities have their own strengths an weaknesses, this thesis proposes a multimodality approach using both PET and CT data. A multimodality approach using and combining PET and CT imaging can combine the accurate detection of the activity of a lesion of PET imaging with the precise anatomical localization of lesions and organs in CT imaging (Wang et al., 2022). For this reason, since the introduction of the Hawkeye, the first multimodal PET/CT scanner to be deployed in a medical setting, in 1999, PET/CT scans have been widely used in the medical domain (Townsend, 2008). In this thesis, the effect of the inclusion of PET data on a deep learning model's performance will be examined. To be exact, the first subquestion will be:

> **What is the effect of a multimodality approach incorporating both PET and CT modalities on model performance compared to single modality approaches?**

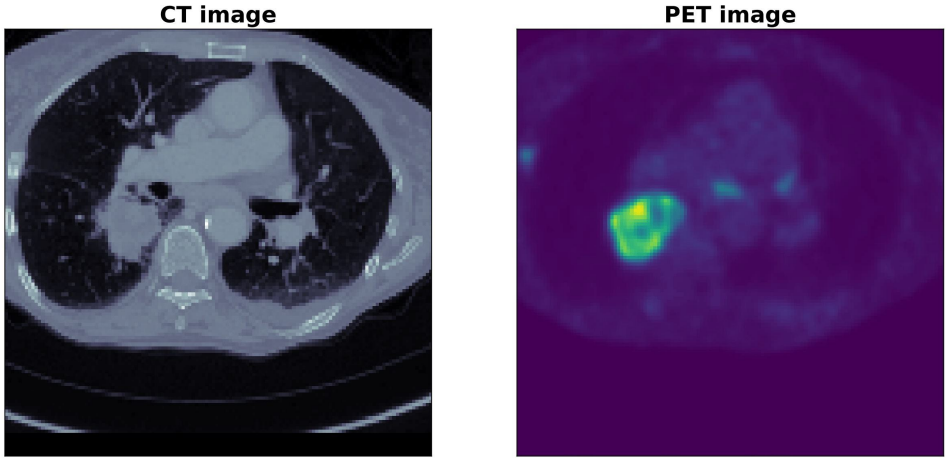



Figure 1: Comparison of a PET and CT patch of the same patient scan slice generated and used in this thesis. Both the benefits and shortcomings of CT and PET scans can be observed in this figure. Specifically, the figure shows that the CT patch clearly visualizes the lung region but poorly depicts the metabolic activity within this region. For the PET patch, the reverse can be observed

Secondly, in medical image segmentation, two commonly applied deep learning methods are the U-Net model and the CNN model. The CNN architecture is a powerful architecture which has been applied in many areas, including image classification, object detection, and scene labeling

(Gu et al., 2018). In this thesis, the CNN model will serve as the baseline model. In the U-Net architecture, skip connections enable the model to both localize target objects and use contextual information, thereby overcoming the trade-off between localization and use of context (Hesamian et al., 2019). Due to the fact that models will be trained on both single modality and multimodal input, using a pre-trained model via transfer learning is not possible. The second subquestion in this thesis will be:

> **What is the difference in performance between a CNN and a U-Net architecture on automatic lung tumor segmentation?**

### 2.5 *Summary of Findings*

The main findings obtained in this thesis demonstrate that using PET as an additional input, or even as a single input, results in a remarkably higher segmentation and detection performance compared to simply using CT data. The differences in tumor segmentation and detection between the PET and PET-CT models are minimal. This means that, in the multimodal PET-CT models, the PET modality is the main driver behind the lung tumor segmentation while the CT modalitiy has a marginal to negligible effect. Regarding the performance of the U-Net and CNN models, the CNN models obtained marginally higher detection rates than the U-Net models, while the U-Net models obtained slightly higher Dice coefficients than the CNN models. Some limitations of this work are the high false positive rates, the occurrence of metastases outside the lung regions, and the usage of 2D instead of 3D data.

## 3 RELATED WORK

In this section, some related work on the application of PET/CT data and deep learning methodologies using PET and CT data will be discussed. The first subsection will briefly focus on the application of PET and CT imaging for lung nodule detection. In the second subsection, deep learning techniques for the detection of lungs and lung nodules using PET and/or CT data will be discussed. Then, applications of deep learning techniques using PET and/or CT data for other detection and segmentation purposes will be addressed. Lastly, the societal and unique contributions of this thesis will be briefly mentioned.

## 3.1 *Application of PET and CT imaging for Lung Cancer*

PET and CT have been widely applied in the detection and evaluation of lung cancer tumors. Verschakelen et al. (2004) evaluated the role of CT in lung cancer staging: while CT is widely available and it can provide accurate insights into the presence and spread of lung cancer in the thoracic area, it performs poorly in determining whether a lung cancer lesion is benign or malignant. Hence, other modality techniques need be used to address and combat the limitations of CT in lung cancer detection. PET imaging, providing information about the metabolic activity within the lung regions, can be used to more thoroughly differentiate between and detect benign and malignant lung cancer tumors (Vansteenkiste, 2003). Visual correlation of PET and CT - finding similar lung cancer locations in a separate PET and CT scan - is a more powerful technique than simply using PET or CT without regarding the contributions of the other modality (De Wever et al., 2007).

In recent decades, the application of integrated PET/CT imaging has increased (Hochhegger et al., 2015). PET/CT scanning has the benefit of providing anatomical information from the CT modality and metabolic information from the PET modality at the same time using a single device (Bruzzi & Munden, 2006). Moreover, using PET/CT, lung cancer lesions, nearby lymph nodes, and distant metastases can be more easily detected than when using PET or CT as a single modality (Hochhegger et al., 2015), thereby effectively identifying a lung tumor's infiltration into nearby and more distant structures.

## 3.2 *Deep Learning Models Applied to CT and PET for Lung and Lung Nodule Detection*

Fu et al. (2021) introduced a deep learning-based multimodal PET-CT lung tumor image segmentation approach. Their approach consists of two components: a multimodal spatial attention module (MSAM) and an encoder-decoder CNN. The MSAM was used to exploit the high sensitivity of PET to produce an attention map. The CNN was used to segment and extract tumors from the CT data, with the CNN focusing on areas that are more likely to be a tumor based on the attention map produced by the MSAM. The spatial attention maps generated by the MSAM and the CNN feature maps are multiplied element-wise to focus on areas with the strongest spatial attention, producing a final segmentation output. The main findings demonstrate that the utilization of an MSAM improves the model performance. Moreover, the PET-CT segmentation proved to be more successful than using a single modality or a concatenation of PET

and CT data. The MSAM outperformed other existing attention-based methods in terms of precision, sensitivity, specificity, and Dice coefficient.

Secondly, Chen et al. (2021) introduced a new deep learning-based automatic lung lesion segmentation method from COVID-19 CT images. Patches were extracted to only include relevant information (i.e., Region of Interest (ROI)) and exclude redundant background information. Then, seven different data augmentation techniques were applied to reduce overfitting. Afterwards, a 3D Attention U-Net was developed, which is able to increase the attention to the target object. A combination loss function was used, which results in a faster network convergence and a greater Dice coefficient. The results show that data augmentation resulted in a higher accuracy in the testing dataset by preserving the main characteristics of the original data and having a stronger generalization capability. Moreover, the network with the combination loss function achieved a higher Dice coefficient than networks trained with binary cross-entropy or dice loss. Overall, the method proposed in this network outperformed other types of networks.

Lastly, Osadebey et al. (2021) used a deep learning-based approach to segment lung regions from CT images based on the traditional three-stage segmentation approach (i.e., pre-processing, processing, and post-processing). For the pre-processing stage, a CNN was used to classify all input images into lung and none-lung regions, aiming at preventing false positives from occurring too frequently. In the processing stage, two U-Nets were used to segment lung regions and refine lung contours, respectively. In the post-processing stage, another CNN was used to filter out none-lung regions that the first U-Net model segmented as a lung region. The main advantages of the methodology are that it is fast, simple, computationally efficient, and that it produces reproducible segmentation results. The model using the three stage-stage system obtained higher Dice scores than the model without the CNNs or contrast enhancement. In future work, explainable AI and intensive evaluation of a radiologist should be used to further investigate the explainability and efficacy of the proposed method.

## 3.3 *Deep Learning Models Applied to CT and PET in Other Contexts*

Wang et al. (2022) developed a deep learning-based method for automatic heart and bladder detection and segmentation by using FDG PET/CT data. Frequently, in tumor detection, there are many false positives because the level of uptake of fluorodeoxyglucose (FDG) in the heart and bladder is similar to the accumulation of FDG in tumors (Wang et al., 2022). Hence, in this study, a deep learning network was created to overcome this burden

of physiological noise. To be exact, two 3D U-Nets were created, separately segmenting the heart and the bladder by using the PET and CT data as multimodal input. The main results demonstrate that this model achieved Dice coeffcients of 0.96 and 0.95 for heart and bladder segmentation, respectively. Moreover, the methodology managed to greatly resolve the problem of physiological noise, having fewer false positive instances.

Secondly, Naser et al. (2021) implemented a deep learning model the purpose of automatic head and neck cancer (HNC) segmentation. They trained and evaluated both a 2D U-Net model and a 3D U-Net model, and used multimodal PET/CT images as input for the model to capture both metabolic activity and anatomical information. The intensities in the PET/CT images were normalized and data augmentation was performed in order to prevent overfitting. 5-fold cross validation was used to evaluate the model. For the 3D model, the mean and median Dice similarity coefficient (DSC) scores were 0.79 and 0.67, respectively. For the 2D U-Net, these scores equaled 0.79 and 0.67. The difference between the Dice coefficients of the 3D and 2D U-Nets were significant. These results demonstrate that, although further improvement is needed, using 3D data, combined with multimodal PET/CT input and data augmentation, results in a relatively strong model performance.

Lastly, Xu et al. (2018) used two deep learning models for full-body bone lesion detection for multiple myeloma (MM) with PET/CT imaging. To be specific, two networks, a V-Net and a W-Net, were used to develop an automatic MM bone lesion detection and segmentation procedure. The V-Net model used in this work is based on the U-Net architecture, using 3D convolutions instead of 2D convolutions. V-Net consists of a compression path in which the volumetric size is decreased and features are extracted, and a decompression path, in which the feature maps obtained in the compression path are expanded to produce a final output. In the W-Net, two V-Nets are cascaded. The first V-Net produces a binary mask for the skeleton, and the second V-Net utilizes both this binary skeleton mask and PET and/or CT data as input. The main results demonstrated that a V-Net trained on multimodal PET-CT data obtained a higher Dice score (69.49%) than a V-Net trained on CT data (26.37%) or PET data (28.51%). The W-Net with PET/CT input only slightly outperformed the V-Net with PET/CT input, obtaining a Dice coefficient of 72.98%. These findings highlight the promising capabilities of using multimodal PET-CT data as model input compared to only using PET or CT data.

### 3.4 *Literature Gaps and Contributions*

In this thesis, the contribution of the PET modality on a deep learning-based lung tumor segmentation procedure will be explored, using a dataset that has not been widely used. Since the original dataset used in this work also comprises lymphoma and melanoma patients, the contributions of PET scans to medical image segmentation can be extended for other lesion types as well. In previous research on lung cancer segmentation, CT has mostly been used as model input, or CT and PET data were used to extract different features in a stacked model architecture. In this thesis, however, different models will be trained using either CT, PET, or combined PET-CT data, thereby thoroughly and explicitly investigating the inclusion of PET data and the exclusion of CT data. Moreover, the current state-of-the-art model for biomedical image segmentation – U-Net – will be compared to a CNN model. Thus, the effect of skip connections and concatenation layers on automatic lung tumor segmentation will be investigated as well.

## 4 METHODS

Figure 2 depicts the process conducted throughout the research. In this section, all the steps will be explained and discussed.

### 4.1 *Dataset Description*

The dataset used in this thesis comprises 900 patients who were examined in two medical centers: the University Hospital Tübingen, Germany, and the University Hospital of the Ludwig Maximilian University (LMU) in Munich, Germany. The set of patients can be divided into melanoma, lung cancer, lymphoma, and negative control patients. The CT data, PET data, and segmentation data are stored as `nifti` files. A `nifti` file consists of 2D images stacked along a third dimension, thereby effectively representing a 3D scan. The full-body PET/CT images were obtained using a Biograph mCT PET/CT scanner.

For all patients, there is a 3D volume representing the PET volume, a volume representing the CT volume, and a volume representing a binary mask (i.e., the segmented data). For all patients, the CT volumes are resampled - to account for a possible shift in a patient's physical position - so that the CT and PET imaging resolution are identical. The slice thickness used for the CT scans was 3 mm. In total, there were 411 negative control patients and 489 patients having at least one scan containing lung cancer, lymphoma, or melanoma. Table 1 shows the exact breakdown of the count of patients per lesion type.

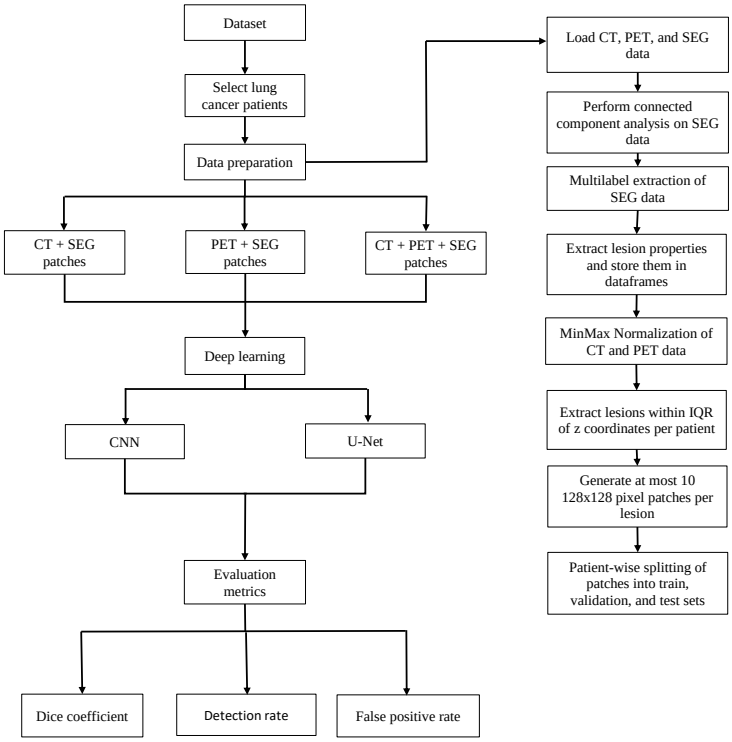



Figure 2: Workflow diagram depicting the process conducted. First, only lung cancer patients were extracted from the dataset. Then, step-wise data preparation was performed to generate appropriate input patches. The CT, PET, and PET-CT patches were each used as input for a different CNN and U-Net model, resulting in six different models. During model training, the models were trained with binary cross-entropy as a loss function and the Dice coefficient as an evaluation metric. The detection rate and false positive rate for the validation and test data were computed after the model training procedure. Negative patches (i.e., patches exclusively containing healthy tissue) were used to compute the false positive rates of the models

| **Patient label** | **No. patients** |
|---|---|
| Negative (Control) | 411 |
| Lung cancer | 168 |
| Lymphoma | 144 |
| Melanoma | 177 |

Table 1: The number of patients with at least one lesion in one of their scans

## 4.2 *Data Preparation*

For the purpose of this research, only patients containing lung cancer were used for analysis and model training. First, the files containing the PET, CT, and segmentation data were loaded. Connected component analysis, using the Python package `cc3d`, was performed to combine and connect different, consecutive 2D annotations into one coherent 3D structure. Then, to label each individual component as a separate lung cancer lesion, the `metrics.regionprops_3D` function from the Python package `porespy` is used. The centroid x, y, and z coordinates, and the minimum and maximum x, y, and z coordinates are stored in a separate dataframe.

After extracting and storing the data, min-max normalization was performed on the CT and PET data. The normalization was performed on the entire scan. Subsequently, per patient, at most 10 random patches, having a size of 128x128, were created, using the x, y, and z coordinates to precisely extract the tumor's location in the PET, CT, and segmentation data. Moreover, two requirements needed to be met in order for random patches to be created for the lesion:

- In order for the model to be trained and improve model convergence, patches with larger annotation size were required. Hence, annotations within the patches needed to be at least 1% of the total patch size
- For each patient, only patches will be generated where the minimum and maximum z coordinates are within the interquartile range of all lesion z coordinates. By only selecting patches within an acceptable range, the amount of metastases outside the lung regions will be reduced. Hence, the interquartile range of the population of lung cancer z coordinates was used to filter out metastases as much as possible while largely retaining tumors in the lung area. In Figure 3, the locations of different tumors - based on 25 lung cancer patients - are visualized using a body heatmap. The probability distribution of all lung cancer z coordinates is shown in Figure 4. Figure 5 shows examples of CT and PET patches of the same patient scan slices which meet this and the above requirement

Lastly, in order to properly train, validate, and test the deep learning models, patient-wise splitting was performed to assign each set of patches per patient to a respective training, validation, and test set. 70% of patches were assigned to the training set, whereas 15% were reserved for the validation and test set each. Data augmentation in the form of horizontal mirroring, vertical mirroring, and both horizontal and vertical mirroring, was performed on the training set in order for the models to become more

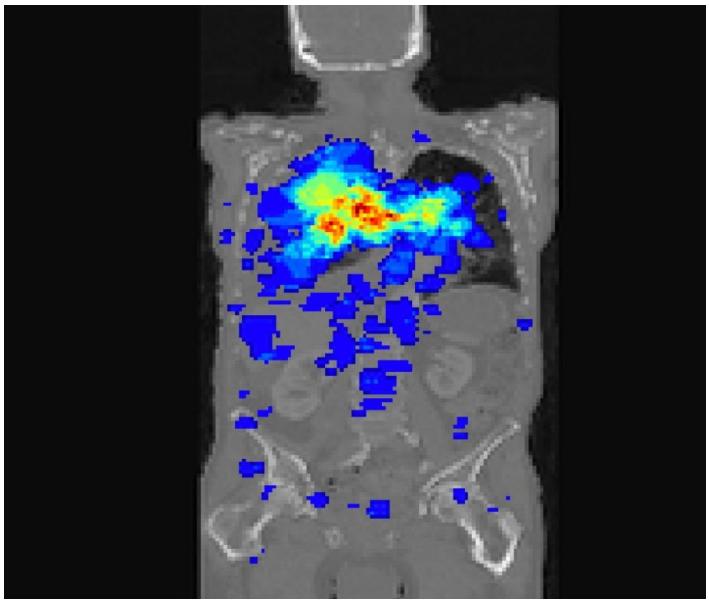


Figure 3: Body heatmap visualizing the location of tumors in the set of lung cancer patients. Rigid registration was used to align the bodies and anatomical structures of different patients. A sample of 25 patients was used to create this visualization. While not all patients are included, the figure shows that the lung cancer data comprise both actual lung cancer tumors and metastases outside the lung region. Reprinted with permission from L. Popper (Popper, 2023)

robust to invariance. This means that, for each patch in the training set, three additional, augmented patches were created. In total, 5,568 patches were generated from all the positive scans corresponding to lung cancer patients. Of these, 3,878 were assigned to the training set. After data augmentation, the size of the training set increased by a factor of 4 to 15,512. The validation set consisted of 839 patches, whereas the test set comprised 851 patches.

### 4.3 *Negative patches*

Besides positive (i.e., lung cancer) patches, 2,349 negative patches were generated from the CT, PET, and segmentation data from the negative control patients. Similarly to the positive patches, only negative patches within a pre-defined location were extracted in order to exclude none-lung regions as much as possible. However, instead of only looking at the range of z-coordinates, the x- and y-coordinates were also taken into account. To be exact, all lung cancer tumor x-, y-, and z-coordinates from all lung cancer patients were taken. Consecutively, a negative patch was only considered to be valid if, for all three dimensions, the minimum and

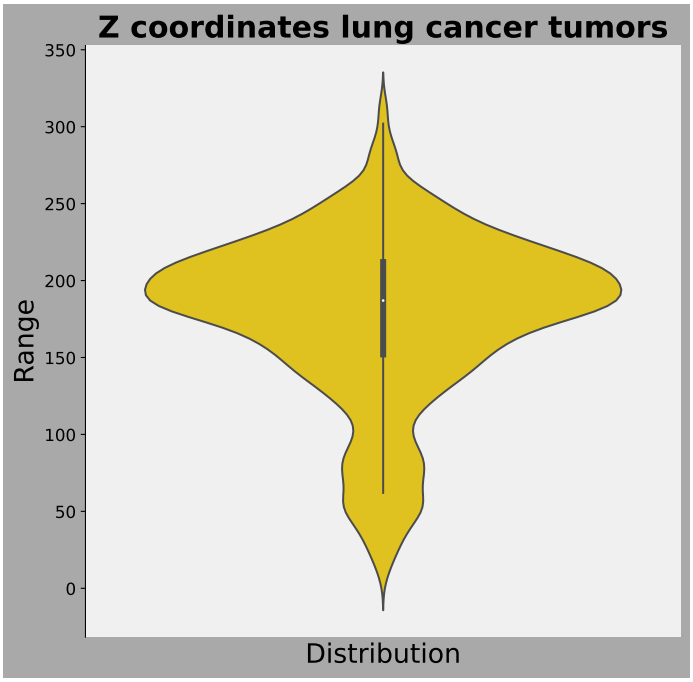



Figure 4: Violin plot visualizing the distribution of z coordinates of all lung cancer tumors. The width represents the frequency of a particular z coordinates, the dots represent the interquartile range and the vertical line represents the interquartile range multiplied by 1.5. The white dot in the middle represents the median. The figure demonstrates that tumors can be found at almost any location along the longitudinal dimension, highlighting the presence of metastases outside the lung regions

maximum value were between the 30th and 70th quantile of all three sets of tumor coordinates.

## 4.4 *Deep Learning Models*

After the data preparation and generation phase, two deep learning models - a 2D U-Net and a 2D CNN model - were trained. To be exact, each model was trained on three different kinds of input data:

1. A single modality input exclusively containing the CT data
2. A single modality input only comprising the PET data
3. A multimodal input comprising both the CT data (as the first channel) and the PET data (as the second channel)

This means that, in total, 6 different models were trained and validated. In this work, the CNN architecture is identical to the U-Net structure,

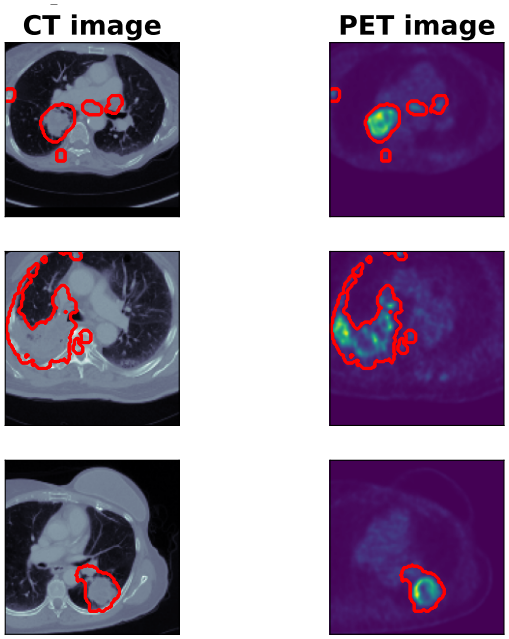



Figure 5: Examples of PET and CT patches of the same patient scan slices generated in the data preparation phase. The left column depicts CT patches, the right column visualizes PET patches, and the tumor annotations are depicted using red contour lines.

with the only difference being the absence of the skip connections and concatenation layers.

## 4.5 *CNN*

Probably one of the best-known deep learning networks, a Convolutional Neural Network (CNN) is composed of intermediate layers which each perform a particular action, such as convolution and pooling (Hesamian et al., 2019). The first layer of a CNN is the input layer, where the inputs are inserted into the network. The next layers are the convolutional layers which perform certain operations on the inputs so that features are extracted. After these layers, non-linear functions are applied by the activation layers. Lastly, the fully connected layers extract high-level features (Hesamian et al., 2019). A CNN is based on an Artificial Neural Network (ANN) (Thapar et al., 2022), and since the introduction of the CNN, various variants have been introduced, including the 2.5D CNN, 3D CNN, and the U-Net, which will be discussed below. The CNN architecture used in this thesis is visualized in Figure 6.

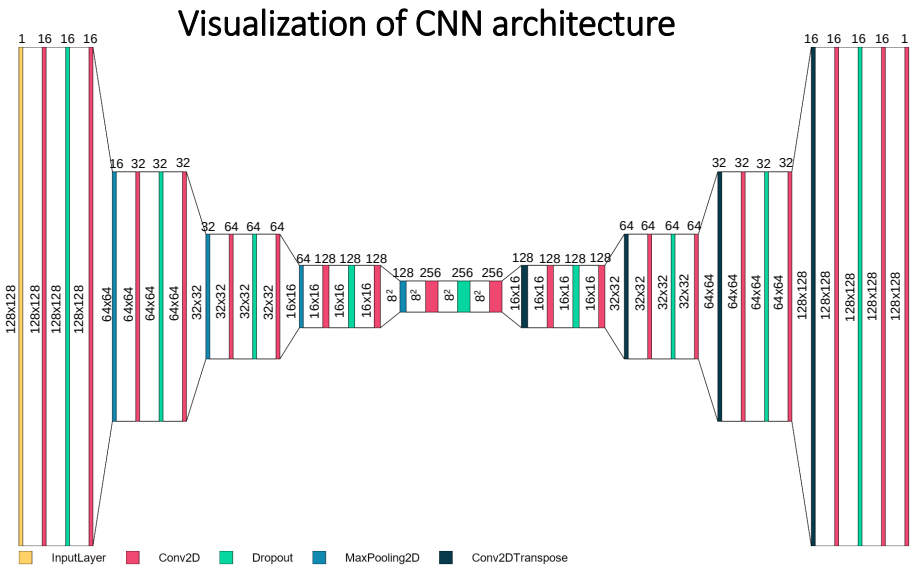



Figure 6: Visualization of the CNN architecture used in this thesis. The annotations on the vertical side of the layer indicate the size of the first and second dimension of the layer, whereas the annotations on the horizontal side indicate the size of the third dimension. The down-sampling phase is composed of four blocks, with each block containing a 2D max-pooling layer, a 2D convolutional layer, a dropout layer, and another 2D convolutional layer. The upsampling phase is also composed of four blocks, with each block comprising a 2D transpose convolutional layer, a 2D convolutional layer, a dropout layer, and a second 2D convolutional layer. For the multimodal PET-CT model, the only difference is the size (i.e., the third dimension) of the input layer, which is 2 instead of 1. `Visualkeras` was used to create the CNN architecture visualization (Gavrikov, 2020)

## 4.6 *U-Net*

First introduced by Ronneberger et al. (2015), the U-Net model has grown to be one of the most well-known deep learning architectures in medical image segmentation (Hesamian et al., 2019). The model is derived from a Fully Convolutional Network (FCN) (Xi et al., 2020), and it is composed of two main components: a contracting path and an expansive path (Ronneberger et al., 2015). In the contracting path, also known as the analysis path (Jha et al., 2020), the input data is repeatedly down-sampled in each successive layer, thereby capturing context and extracting specific deep features. In the expansive path, also known as the synthesis path (Jha et al., 2020), the features extracted in the contracting path are used to perform the segmentation. Skip connections are used to concatenate equally sized layers in the contracting and expansive path (Xi et al., 2020). Since the introduction of the U-Net, various adaptions of the model have

been introduced, including 3D U-Net and V-Net (Hesamian et al., 2019). The U-Net architecture used in this thesis is visualized in Figure 7.

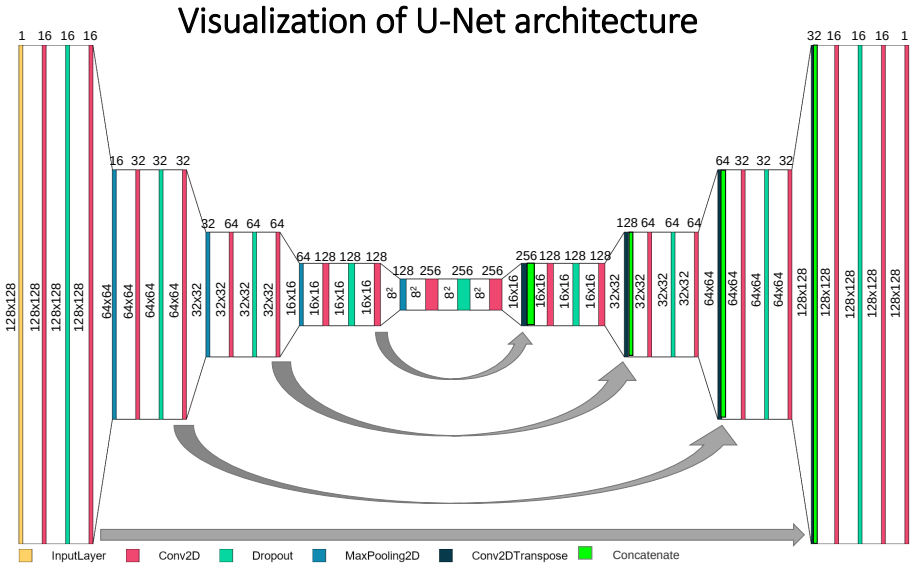



Figure 7: Visualization of the U-Net architecture used in this thesis. The grey arrows indicate the skip connections which are used to concatenate identically sized layer in the contracting and expanding phase. The lime-colored areas indicate the part of the layer which is added to the previous layer by a concatenation layer. Except for these skip connections and concatenation layers, the U-Net structure is identical to that of the CNN. Similar to the CNN architecture, for the multimodal PET-CT input, the only difference is that the input size of the U-Net model is 2 instead of 1. `Visualkeras` was used to create the U-Net architecture visualization (Gavrikov, 2020)

## 4.7 *Hyperparameter Tuning*

Hyperparameter tuning was performed in a manual way to improve model performance. The only hyperparameter values that were tuned in this stage were the dropout rates for each individual dropout layer. The dropout rates were the same for the U-Net and CNN models.

## 4.8 *Evaluation Metrics*

To evaluate the performance of the trained models on the validation and test set, three different evaluation metrics were used: Dice coefficient, detection rate, and the false positive rate. During model training, the Dice coefficient was used as an evaluation metric. The Dice coefficient indicates

how much overlap there is between the ground truth $X$ and the model's prediction $Y$, as in:

$$Dice(X) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

After model training and validation, the detection rate and the false positive rate were computed. The detection rate computes the ratio of - at least partially - overlapping annotations in the ground truth and the predictions over all annotated tumor components. A value closer to 1 indicates that a higher percentage of the predictions within the 2D patches at least partially overlap with an actual annotation. The false positive rate computes the ratio of all negative patches where at least one pixel was predicted to be cancerous over all negative patches. Hence, a value of 0 indicates that the model is able to perfectly distinguish healthy tissue from cancerous tumors, whereas a value of 1 indicates that the model always at least partly classifies some healthy tissue to be cancerous.

### 4.9 *Model Implementation*

The CNN and U-Net were implemented in Python using the `Tensorflow` package. An Early Stopping Rate of 50 epochs was used during model training, and binary cross-entropy was used as a loss function. The sigmoid function was used as a final activation function, which results in each pixel being assigned a value between 0 and 1. This value indicates the probability of a pixel being (part of) a lesion or not. Thresholding was applied to produce a final set of predictions, with pixels having a value of at least 0.5 being classified as cancerous. Pixels with values lower than 0.5 were classified as healthy tissue.

## 5 RESULTS

The results obtained in this thesis are stored in Table 2, 3, and 4. In the following subsections, the results will be discussed more thoroughly.

### 5.1 *Segmentation Results*

In Table 2 and 3, the evaluation metric scores for the validation and test data are depicted, respectively. Based on these results, it can be observed that for both the 2D U-Net and CNN models, the model purely trained on CT data obtains a Dice Coefficient and detection rate that differs from the PET and PET-CT models by at most nearly 40% (for the validation data). A similar effect can be observed in the test data, although the CT,

PET, and PET-CT data obtain higher scores for all evaluation metrics. The multimodal models having both PET and CT data as input obtain slightly higher scores than the single modality models purely trained on PET data. However, this difference is demonstrably smaller than the margin between the single modality CT and multimodality PET-CT models. Moreover, for the test data, the multimodality U-Net model has a Dice Coefficient 1% above the Dice Coefficient of the multimodal CNN model. Nonetheless, the multimodal CNN model performs somewhat better than the multimodal U-Net model by a narrow margin of 1% as well in terms of detection rate.

Additionally, in Appendix C (page 36), the development of the binary cross-entropy loss and the Dice coefficient score for both the training and validation data are plotted for the first 50 epochs for the CNN and U-Net models. Upon visually inspecting the evolution of the loss and Dice Coefficient scores over the number of epochs, it can be noticed that already after 10 epochs, the models start to converge to a constant value. Furthermore, for the training data, the binary cross-entropy value decreases over time for all modalities while the loss value actually increases for the validation data. Regarding the development of the Dice Coefficient, for all modalities, the training data converge to a higher value than the validation data. Hence, overfitting occurs, with the validation performance lagging behind the training performance.

| Model | Modality | Val dice [-] | Val detection rate[-] |
|---|---|---|---|
| **U-Net** | CT | 0.22 | 0.24 |
| | PET | 0.55 | 0.55 |
| | PET-CT | 0.57 | 0.55 |
| **CNN** | CT | 0.18 | 0.21 |
| | PET | 0.56 | 0.57 |
| | PET-CT | **0.59** | **0.58** |

Table 2: Table depicting the Dice coefficients and detection rates for the validation data

| Model | Modality | Test dice [-] | Test detection rate [-] |
|---|---|---|---|
| **U-Net** | CT | 0.28 | 0.28 |
| | PET | 0.66 | 0.72 |
| | PET-CT | **0.67** | 0.73 |
| **CNN** | CT | 0.23 | 0.27 |
| | PET | 0.60 | 0.73 |
| | PET-CT | 0.66 | **0.74** |

Table 3: Table depicting the Dice coefficients and detection rates for the test data

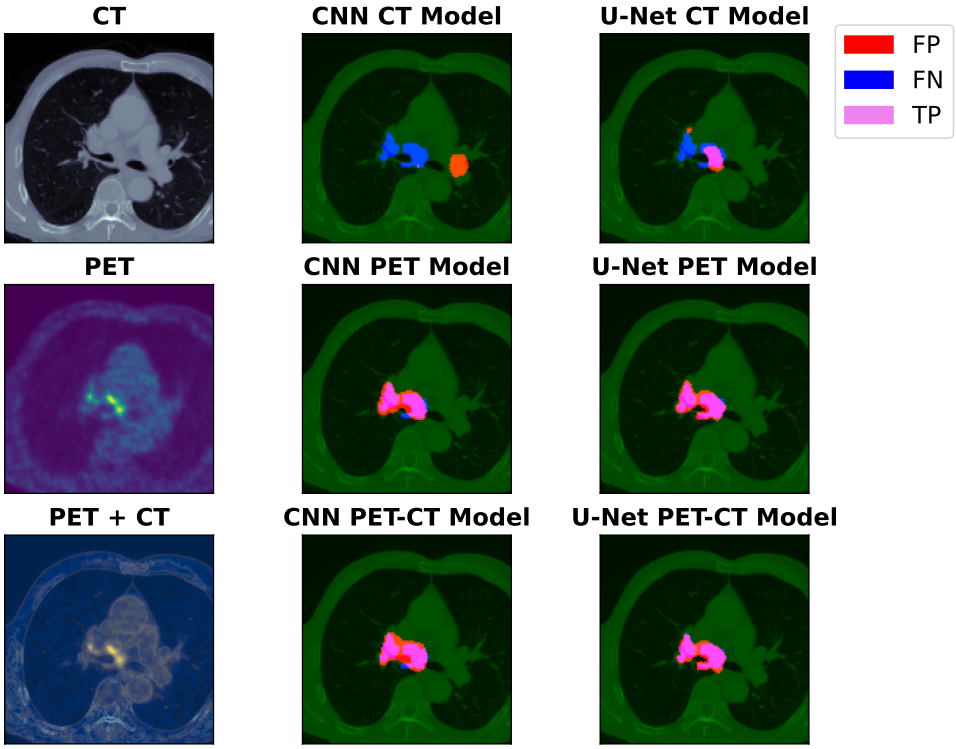



Figure 8: Example of the segmentation results of the CNN and U-Net models on the test data. The left column depicts the raw scan, the middle column visualizes the annotations and the CNN segmentations using an overlay plot, and the right columns contains an overlay plot of the annotations and U-Net predictions. The CT scan is depicted in green in the overlay plots for better contrast and an indication of the tumors' location. In the overlay plots, the false positive (FP) pixels are colored red, whereas the false negative (FN) pixels are colored blue and the true positive (TP) pixels are colored violet. In other words, violet pixels indicate the overlap between the ground truth and predictions

## 5.2 *Comparison of annotations and predictions inside lung region*

Figure 8 depicts the segmentation results of the models trained on CT data, PET data, and PET and CT data on the first, second, and third row, respectively, for both the CNN model and the U-Net model. The figure shows that, for the CT modality, the U-Net model partially detects the tumor within the lung region, whereas the CNN model predicts an entirely different area within the lungs - which in reality is healthy tissue - to be a cancerous tumor. On the other hand, the models trained on PET and combined PET-CT data appear to perform better in both detecting and segmenting tumor areas and in distinguishing malignant regions from

background areas. For more segmentation results within lung regions, see Appendix A (page 33).

## 5.3 *Comparison of annotations and predictions outside lung region*

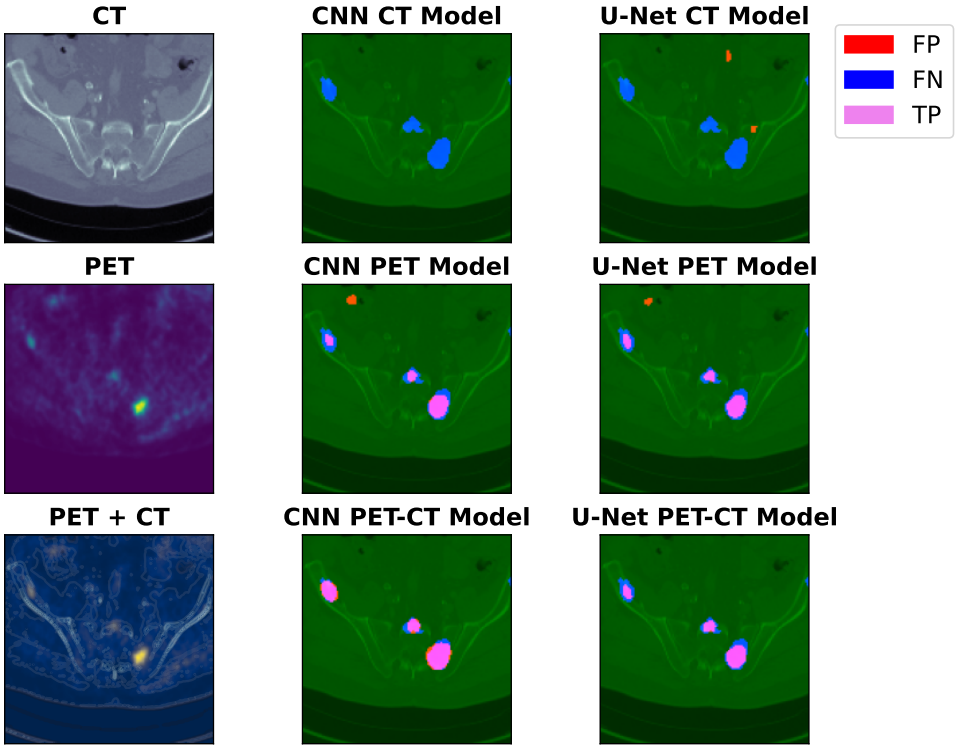



Figure 9: Example of the segmentation results of the CNN model on a patch outside the lung region. This figure shows the respective predictions of a CT, PET, and PET-CT CNN and U-NET models for a metastasis. It demonstrates that, while the CT models are unable to detect and segment this metastasis (in the case of the CNN) or only partially able to do this (in the case of the U-Net), the PET and PET-CT models still perform reasonably well in terms of detection and segmentation

As previously mentioned, during the pre-processing step, only patches with minimum and maximum z coordinates within a patient's interquartile range of z coordinates were used for training. However, this method did not prevent metastases to be fully excluded from the training, validation, and test data. As such, occasionally, patches containing metastases outside the lung region were included in the train validate and evalutate the models. An example of such a patch and the corresponding model predictions can be seen in Figure 9. These visualizations show that, although the models

trained on CT data perform poorly in detecting and segmenting lesions outside the lung regions, the models trained on PET and PET-CT data still are able to detect metastases reasonably well. Further segmentation results on metastases found outside the lung region can be seen in Appendix B (page 36).

## 5.4 *False positive rate*

| Model | Modality | False positive rate [-] |
|---|---|---|
| **U-Net** | CT | **0.50** |
| | PET | 0.79 |
| | PET-CT | 0.86 |
| **CNN** | CT | 0.65 |
| | PET | 0.87 |
| | PET-CT | 0.80 |

Table 4: Table depicting the false positive rate scores for the held-out negative test set. The table shows that the false positive rates for all models are high, with the single modality PET and multimodal PET-CT models obtaining the highest false positive rates

Lastly, as an additional evaluation metric, the false positive rate was computed for all models. In order for this metric to be computed, 2,349 negative patches were generated from the set of negative control patients. Consequently, a patch was considered to be a false positive if at least one pixel was predicted to be a tumor. In Table 4, the false positive rates for all models are stored. Contrary to the Dice coefficient and detection rate, the models purely trained on CT data obtained the lowest false positive rate for both the U-Net and CNN architecture. Moreover, with a margin of 15%, the U-Net CT model has a considerably smaller false positive rate than the CNN CT model. The false positive rates for the models trained on PET and PET-CT are remarkably high, exceeding 0.80 in all but one circumstance. Examples of negative patches and their respective model segmentation outputs are shown in Figure 10. It can be observed that, while the CT-trained model accurately segments no pixel (in the CNN architecture) or almost no pixel (in the U-Net architecture), the single modality PET models and the multimodality PET-CT models predict peculiarly high amounts of regions to be malignant where in fact they are healthy tissue.

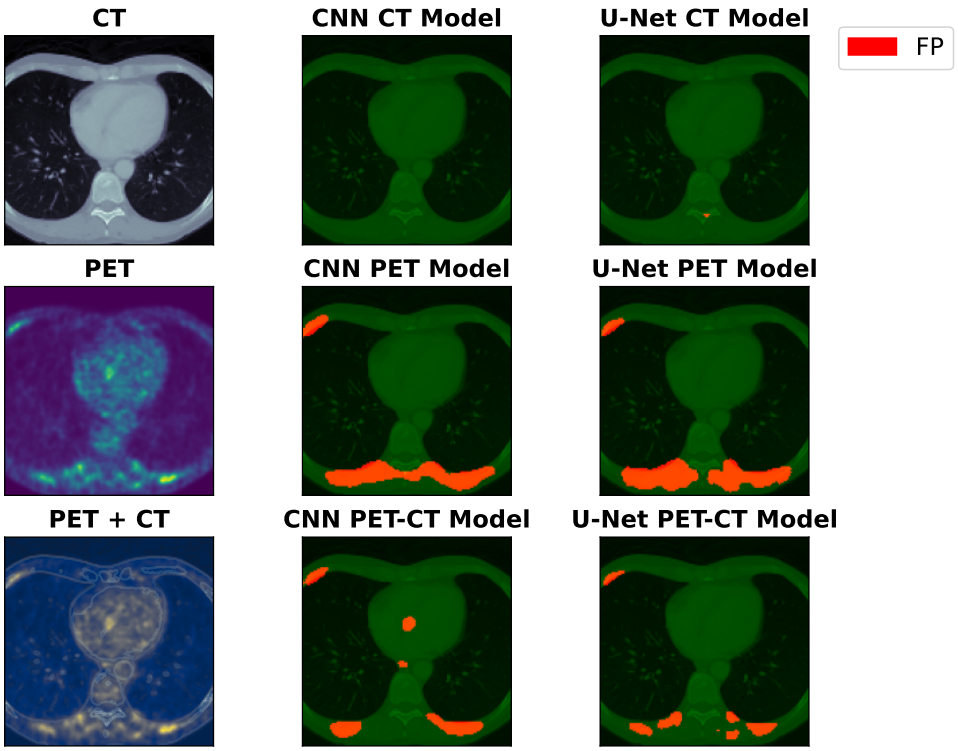



Figure 10: Example of the segmentation results of the CNN and U-Net models on a negative patch. The figure shows the predictions of the CT, PET, and PET-CT CNN and U-Net models on patches that do not contain any lung tumors. It shows that the CT CNN model correctly does not classify any pixel to be cancerous, whereas the PET and PET-CT CNN model falsely segment a large area to be a tumor. In the U-Net CT model, a very small area is incorrectly segmented as a tumor whereas in the U-Net PET and CT models, large areas are segmented as cancerous

# 6 DISCUSSION

In this section, the results will be analyzed and discussed. First, an interpretation of the results will be outlined. Secondly, limitations of this thesis, and how they can be improved in future work, will be discussed. Lastly, the positive contributions of these results to society will be briefly mentioned.

## 6.1 *Interpretation of results*

The main findings demonstrate that both the CNN and U-Net models trained on single modality PET data and multimodality PET-CT data

outperform the models trained on only CT data. This entails that, at least for this dataset, the PET modality is the main driver behind the lung tumor segmentation process. In positive patches, the models trained on CT data often do not segment regions corresponding to lung cancer tumors or metastases. Moreover, while missing these regions, the CT models may incorrectly classify random healthy tissue to be a cancerous tumor. The PET and PET-CT models, on the other hand, frequently correctly segment and detect these malignant regions. Upon visual inspection, it can regularly be seen that the correctly segmented regions in the PET and PET-CT patches correspond to regions with a high metabolic activity in the PET scans. Accordingly, the PET modality proves to be a strong component in terms of lung cancer segmentation and detection, whereas the CT modality appears to have a marginal to negligible effect. Hence, for this dataset, a region's metabolic activity proves to be a more important factor than its anatomical location when it comes to lesion detection and segmentation.

However, the disparity in Dice coefficient and detection rate between the single modality PET models and the PET-CT models was relatively minimal. This is in accordance with the results obtained by Fu et al. (2021): their main findings demonstrate that a model with merely concatenated PET-CT input places disproportionate emphasis on the PET modality while almost neglecting the CT modality, resulting in equal performance with a model trained on PET input alone. On the other hand, the main findings obtained in this thesis are in stark contrast with the results obtained by (Xu et al., 2018), whose model trained on PET-CT input outperformed the models trained on either PET or CT data by more than 40%. Nonetheless, in their work, 3D models instead of 2D models were used, and the total amount of available data was smaller than the amount of data used in this thesis, making it harder to establish the exact cause of the discrepancies between the different findings.

Additionally, the U-Net model does not perform better than the CNN architecture in all circumstances. As a matter of fact, for the test data, the CNN models obtained higher detection rates than the U-Net models, with the U-Net models obtaining higher Dice coefficients. This means that the U-Net architecture used in this thesis results in a somewhat better spatial overlap between ground truth and segmentation mask, while the CNN architecture results in a slightly better identification of the target - i.e., tumor - location. The U-Net architecture is identical to the CNN architecture, with the only difference being the presence of skip connections and concatenation layers between equally sized layers in the contracting and expanding path. Hence, the usage of skip connections and concatenation layers to combine local information with contextual information does not always provide additional benefits to the models trained in this thesis. This stands

in contrast to the results of Ronneberger et al. (2015), who introduced the U-Net architecture and found that it performed significantly better than a CNN without skip connections.

## 6.2 *Limitations and Future Work*

The results obtained in this thesis should be taken with caution, since this research comes with a number of limitations. In this section, some limitations of the methodology will be analyzed and discussed.

### 6.2.1 *Representativeness*

To begin with, the high difference in performance in terms of both Dice coefficient and detection rate between the models on the validation and test data suggest that, given higher metric scores, the test data is more representative of and comparable to the input data than the validation data. This might give a skewed impression of the model performing worse (when inspecting the validation data results) or performing better (upon inspection of the test data results) than it actually performs in reality.

### 6.2.2 *Metastases*

The number of metastases itself proves to be an additional limitation in this study as well. In order to prevent metastases from occurring in the data, the patches were restricted to be within a certain range of z-coordinates for each patient (namely, the interquartile range of z-coordinates). While reducing the number of lesions outside the lung region, this approach did not fully prevent metastases from occuring in the train, validation, and test sets. Consequently, since the lung cancer segmentation models are partially trained on patches that represent anatomical structures outside the lung region, the models might be skewed in segmenting structures with certain characteristics. Particularly, since metastases outside the lung region often correspond to areas with a high uptake of F-FDG in PET scans, the models too often improperly consider regions with a high metabolic activity to be cancerous.

### 6.2.3 *False positives*

Thirdly, the high false positive rates for all three modalities suggest that the models fail to – at least partly – differentiate some healthy tissue from cancerous tumors in a majority of circumstances. Particularly, the models trained on PET and PET-CT data fail to distinguish some healthy tissue from malignant tumors or metastases in more than 80% of all negative

patches. The high false positive rates for the PET and PET-CT models, combined with a comparatively low false positive rate for the CT data, suggest that there is a trade-off between segmentation performance and the occurrence of false positives: the CT models, being significantly worse in lesion segmentation and detection, more often correctly do not classify healthy tissue as being malignant because of their poor ability to segment cancerous data in a majority of circumstances. The reverse can be said for the PET and PET-CT models: because of their relatively strong segmentation performance in positive patches, too frequently, these models incorrectly assume regions with a high metabolic activity in negative data to be cancerous. The occurrence of a large amount of false positives can be attributed to the fact that, during the generation of patches, only patches with an annotation size comprising at least 1% of the total patch size were generated for model training. These patches comprised 50% of the total number of patches generated, with the other 50% - having an annotation size smaller than 1% of the patch size – being discarded. For this reason, only relatively large tumors were used as input for the models.

### 6.2.4 *Marginal Difference between PET and PET-CT and Overfitting*

Additionally, although the multimodal PET-CT models and the single modality PET models outperform the CT models by a large margin, the differences in performance between the former two is very marginal, and can be attributed to chance. Furthermore, overfitting proved to be a significant problem, since the training data obtained a much lower loss value and a greater Dice Coefficient for all modalities and both the CNN and U-Net architecture.

### 6.2.5 *Usage of 2D data*

Lastly, in this thesis, 2D patches were provided as input for both the CNN and U-Net models. This was done because 3D data have a higher computational cost and consumes more GPU memory. In 2D data, the contextual and spatial information within the scans is not exploited as much compared to using 2D data.

## 6.3 *Future Work*

In this section, possible solutions to the above mentioned limitations will be outlined.

### 6.3.1 *Representativeness*

In order to mitigate the discrepancy between the validation and test sets in terms of patches' appearances and more thoroughly investigate the models' actual performance, different techniques using multiple subsets to evaluate the segmentation and detection scores such as *k*-fold cross validation and leave-one-patient-out (LOPO) cross validation can be applied – as done by Naser et al. (2021) and Grossiord et al. (2017), respectively. Using multiple subsets to validate and evaluate the model performance, the number of unrepresentative data (i.e., the metastases outside the lung region) might be spread out more evenly across these subsets. This results in equally representative subsets to evaluate the model, thereby obtaining a better reflection of model performance. This strategy also eliminates the discrepancy in evaluation metric scores between different subsets of data (as is the case between the validation and test data in this work).

### 6.3.2 *Metastases*

In order to properly identify and extract lung cancer patches and exclude metastases outside the lung regions, a more advanced architecture incorporating different or multiple models might be used. For example, Xi et al. (2020) used a cascade U-ResNet which is composed of two individual U-ResNet models: the first model segmented liver regions within the input data, and the second model used these segmented liver regions as input to detect and segment lesions inside the liver. Similarly, a structure in which the first model extracts the lung regions in all patches and another model then segments the tumors within these regions can be a better and more powerful alternative to prevent metastases from occurring.

Furthermore, complementary models can be used to directly segment and detect lesions or tumors without the need of initial organ segmentation. For instance, Fu et al. (2021) used one model - a multimodal spatial attention module (MSAM) - to extract activation maps containing useful features obtained from the PET data, and used a second model - an encoder-decoder CNN - to segment lung tumors based on CT data. The outputs of both models were multiplied elementwise to produce a final segmentation. Thus, the CT and PET data were used in a complementary manner. These segmentations proved to be more accurate than segmentations produced by a merely concatenated PET-CT input, which, similarly to the results in this thesis, placed disproportionate emphasis on the PET data while effectively neglecting the CT data. Hence, in future work, instead of simply providing concatenated PET-CT data as model input, informative features produced by a model trained on PET data can be used to guide tumor segmentation in CT scans. In this way, the benefits of both modalities are

more effectively exploited and no modality is unreasonably prioritized over the other.

### 6.3.3 *False positives*

In this thesis, including extremely small annotations resulted in the models not being able to segment anything during and after training. However, in future work, using a larger dataset and adjusting the model parameters, patches with smaller tumor annotations can be included as well to make the model more robust to patches with a larger amount of background pixels.

### 6.3.4 *Marginal Difference between PET and PET-CT and Overfitting*

In order to mitigate the problem of overfitting, a larger dataset, containing more patches, could be used. This could be achieved by performing more data augmentation techniques, simply extracting more patches per individual lesion per patient, or both. By using a larger dataset, the effect of the inclusion of PET data can be more effectively examined, and the differences between the U-Net and CNN architecture in terms of model performance can be better established.

### 6.3.5 *Usage of 2D data*

Using 3D patches - i.e., a sequence of three neighboring 2D slices, contextual and spatial information might be more effectively explored because of a more efficient volumetric representation (Hesamian et al., 2019). Hence, in future work, 3D patches can be used instead of 2D patches to improve the model's segmentation and detection performance.

## 6.4 *Contributions to Society*

This thesis is a contribution to the existing research investigating the usage of PET and CT data in a deep learning-based medical image segmentation procedure. The results demonstrate that the inclusion of PET data greatly improves a deep learning model's lung tumor segmentation and detection performance. Manually annotating an entire sequence of PET and CT scans is very time consuming, and an effective, automatic image segmentation procedure is therefore highly important and desirable. This work is the first investigation into developing a bone lesion segmentation model using multimodal PET-CT input which can be incorporated into the ETZ workflow. Using a multimodal deep learning model, the detection and

segmentation of bone lesions could be improved. This will aid physicians in saving time and doing other necessary work for patients.

## 7 CONCLUSION

In this thesis, the main research question revolved around the extent to which a deep learning-based approach can segment lung cancer tumors in PET-CT scans. This issue was addressed in two ways: first, the effect of a multimodality approach incorporating both PET and CT modalities was compared to single modality approaches using PET and CT only to efficiently compare the incorporation and addition of PET data. Secondly, the difference in model performance between a CNN and a U-Net architecture on automatic lung tumor segmentation was investigated. The main findings demonstrate that using PET data as a single modality input or as part of a multimodal input along with CT data results in a remarkably higher segmentation performance and detection rate. Visual inspection of the segmentation results demonstrated that the PET and PET-CT models were able to correctly segment malignant regions with a high metabolic activity, even in metastases outside the lung region. The CT models, however, often failed to detect and segment (part of) a tumor or incorrectly segmented an entirely different region. However, the differences in segmentation and detection performance between the single modality PET models and the multimodal PET-CT models, as well as the differences between CNN and U-Net, are marginal. Some additional limitations include the high difference in model performance between the validation and test data, the inclusion of metastases within the datasets, the high false positive rates, overfitting, and the usage of 2D data. Additional methods and strategies should be used in future work to combat these issues.

## REFERENCES

Bomanji, J., Costa, D., & Ell, P. (2001). Clinical role of positron emission tomography in oncology. *The lancet oncology*, *2*(3), 157–164.

Bruzzi, J. F., & Munden, R. F. (2006). Pet/ct imaging of lung cancer. *Journal of thoracic imaging*, *21*(2), 123–136.

Chen, C., Zhou, K., Zha, M., Qu, X., Guo, X., Chen, H., ... Xiao, R. (2021). An effective deep neural network for lung lesions segmentation from covid-19 ct images. *IEEE Transactions on Industrial Informatics*, *17*(9), 6528–6538.

De Wever, W., Ceyssens, S., Mortelmans, L., Stroobants, S., Marchal, G., Bogaert, J., & Verschakelen, J. (2007). Additional value of pet-ct in the staging of lung cancer: comparison with ct alone, pet alone and visual correlation of pet and ct. *European radiology*, *17*, 23–32.

Fu, X., Bi, L., Kumar, A., Fulham, M., & Kim, J. (2021). Multimodal spatial attention module for targeting multimodal pet-ct lung tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), 3507–3516.

Gavrikov, P. (2020). *visualkeras.* https://github.com/paulgavrikov/visualkeras. GitHub.

Grossiord, E., Talbot, H., Passat, N., Meignan, M., & Najman, L. (2017). Automated 3d lymphoma lesion segmentation from pet/ct characteristics. In *2017 ieee 14th international symposium on biomedical imaging (isbi 2017)* (pp. 174–178).

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... others (2018). Recent advances in convolutional neural networks. *Pattern recognition*, *77*, 354–377.

Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, *32*, 582–596.

Hochhegger, B., Alves, G. R. T., Irion, K. L., Fritscher, C. C., Fritscher, L. G., Concatto, N. H., & Marchiori, E. (2015). Pet/ct imaging in lung cancer: indications and findings. *Jornal Brasileiro de Pneumologia*, *41*, 264–274.

Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., & Johansen, H. D. (2020). Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 ieee 33rd international symposium on computer-based medical systems (cbms)* (pp. 558–564).

Kalaivani, N., Manimaran, N., Sophia, S., & Devi, D. (2020). Deep learning based lung cancer detection and classification. In *Iop conference series: materials science and engineering* (Vol. 994, p. 012026).

Naser, M. A., van Dijk, L. V., He, R., Wahid, K. A., & Fuller, C. D. (2021).

Tumor segmentation in patients with head and neck cancers using deep learning based-on multi-modality pet/ct images. In *Head and neck tumor segmentation: First challenge, hecktor 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, proceedings 1* (pp. 85–98).

Osadebey, M., Andersen, H. K., Waaler, D., Fossaa, K., Martinsen, A. C., & Pedersen, M. (2021). Three-stage segmentation of lung region from ct images using deep neural networks. *BMC Medical Imaging*, *21*(1), 1–19.

Popper, L. (2023). *Research in action - research report* (Research Report). Jheronimus Academy of Data Science.

Raichle, M. E. (1983). Positron emission tomography. *Annual review of neuroscience*, *6*(1), 249–267.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii 18* (pp. 234–241).

Saif, M. W., Tzannou, I., Makrilia, N., & Syrigos, K. (2010). Role and cost effectiveness of pet/ct in management of patients with cancer. *The Yale journal of biology and medicine*, *83*(2), 53.

Schabath, M. B., & Cote, M. L. (2019). Cancer progress and priorities: lung cancer. *Cancer epidemiology, biomarkers & prevention*, *28*(10), 1563–1579.

Thapar, P., Rakhra, M., Cazzato, G., & Hossain, M. S. (2022). A novel hybrid deep learning approach for skin lesion segmentation and classification. *Journal of Healthcare Engineering*, *2022*.

Townsend, D. W. (2008). Positron emission tomography/computed tomography. In *Seminars in nuclear medicine* (Vol. 38, pp. 152–166).

Valanarasu, J. M. J., Sindagi, V. A., Hacihaliloglu, I., & Patel, V. M. (2020). Kiu-net: Towards accurate segmentation of biomedical images using over-complete representations. In *Medical image computing and computer assisted intervention–miccai 2020: 23rd international conference, lima, peru, october 4–8, 2020, proceedings, part iv 23* (pp. 363–373).

Vansteenkiste, J. F. (2003). Pet scan in the staging of non-small cell lung cancer. *Lung cancer*, *42*(2), 27–37.

Verschakelen, J. A., De Wever, W., & Bogaert, J. (2004). Role of computed tomography in lung cancer staging. *Current opinion in pulmonary medicine*, *10*(4), 248–255.

Wang, X., Jemaa, S., Fredrickson, J., Coimbra, A. F., Nielsen, T., De Crespigny, A., ... Carano, R. A. (2022). Heart and bladder detection and segmentation on fdg pet/ct by deep learning. *BMC Medical Imaging*, *22*(1), 58.

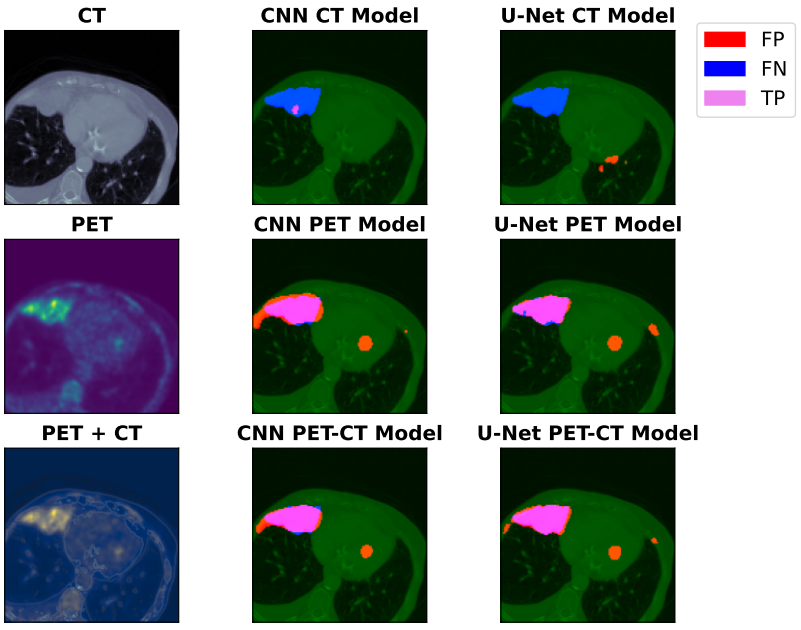
Weinberg, R. A. (1996). How cancer arises. *Scientific American*, *275*(3),

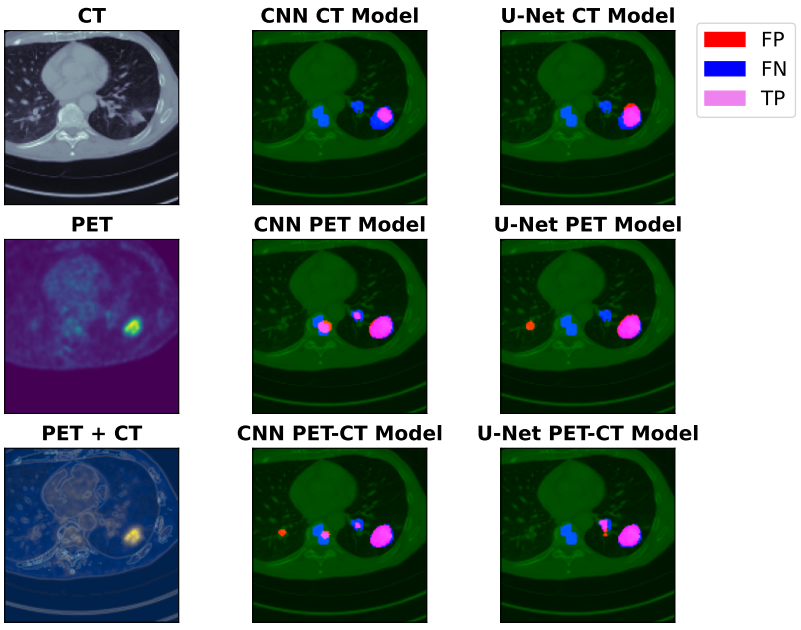
62–70.

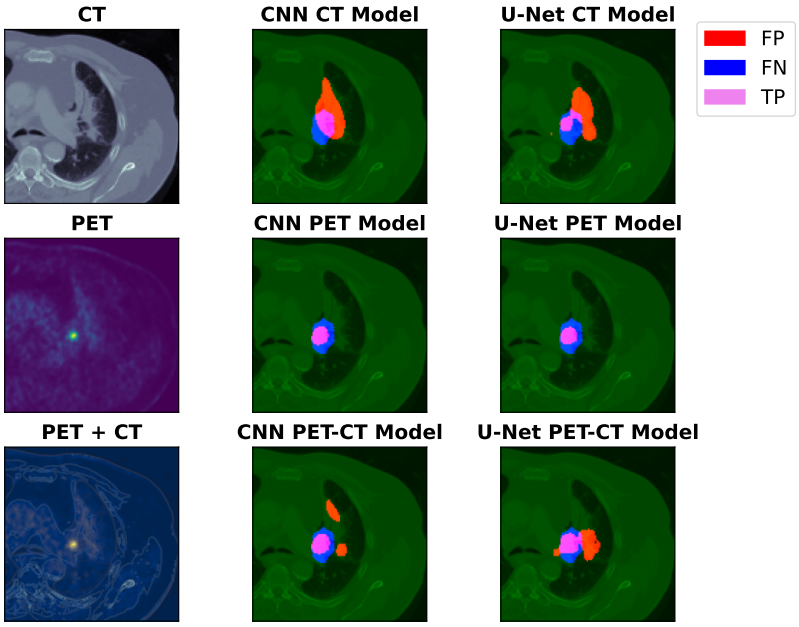
Xi, X.-F., Wang, L., Sheng, V. S., Cui, Z., Fu, B., & Hu, F. (2020). Cascade u-resnets for simultaneous liver and lesion segmentation. *IEEE Access*, *8*, 68944–68952.

Xu, L., Tetteh, G., Lipkova, J., Zhao, Y., Li, H., Christ, P., ... others (2018). Automated whole-body bone lesion detection for multiple myeloma on 68 ga-pentixafor pet/ct imaging using deep learning methods. *Contrast media & molecular imaging*, *2018*.

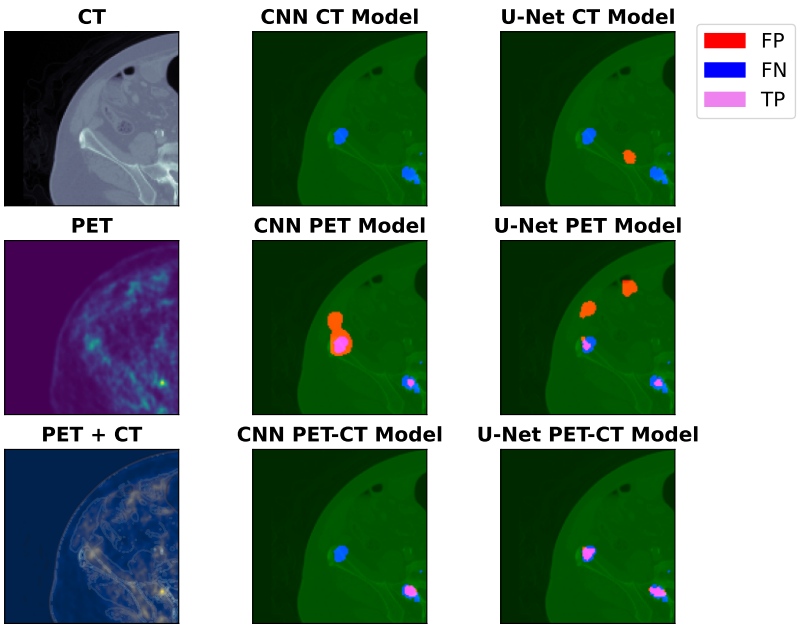
## Appendix A: Visualizations of CNN and U-Net segmentations in lung regions

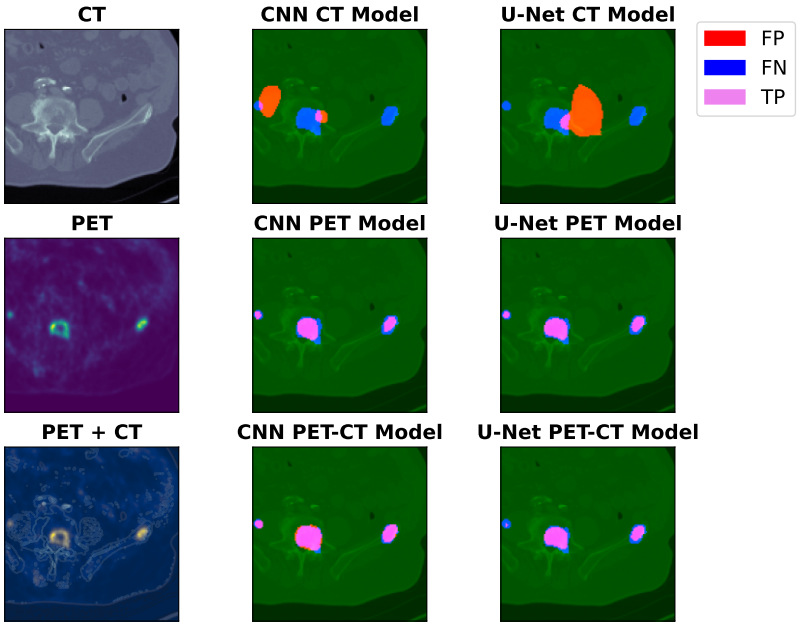


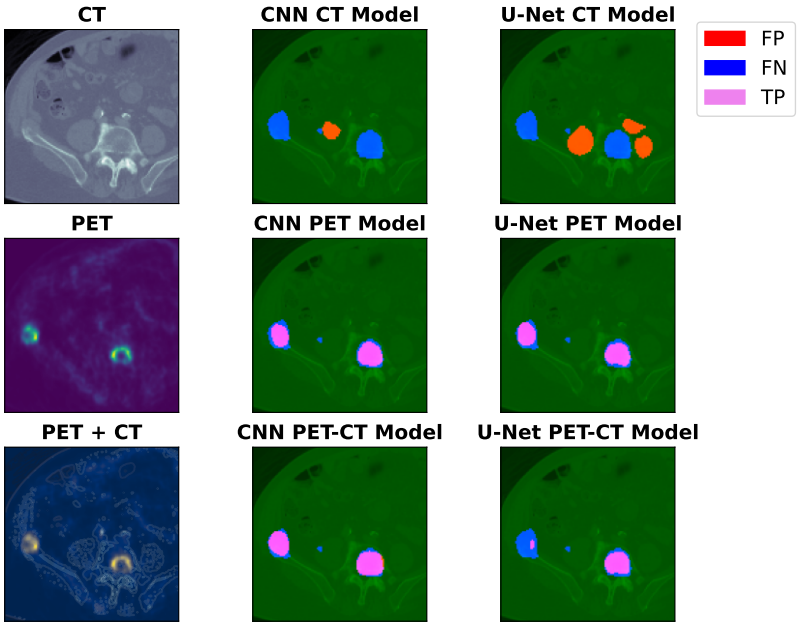





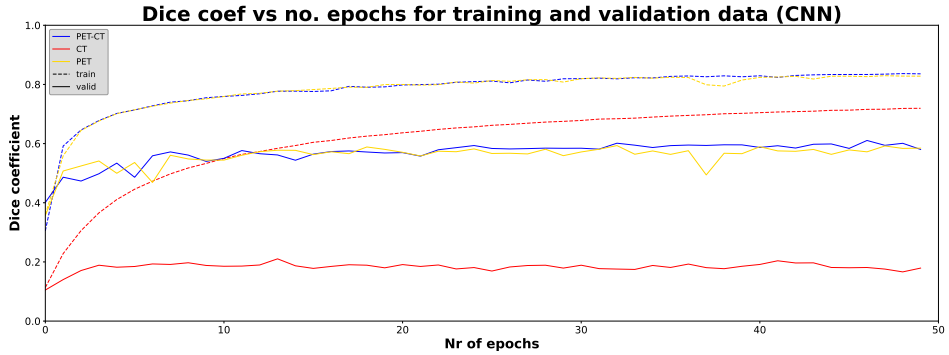

## APPENDIX B: VISUALIZATIONS OF CNN AND U-NET SEGMENTATIONS OUTSIDE LUNG REGIONS

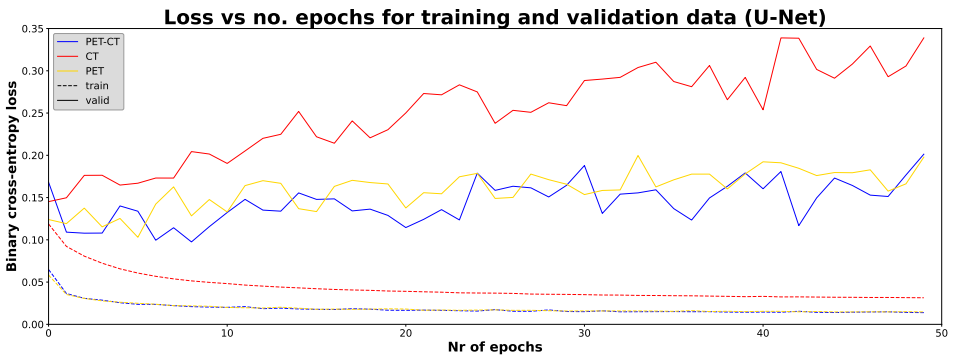


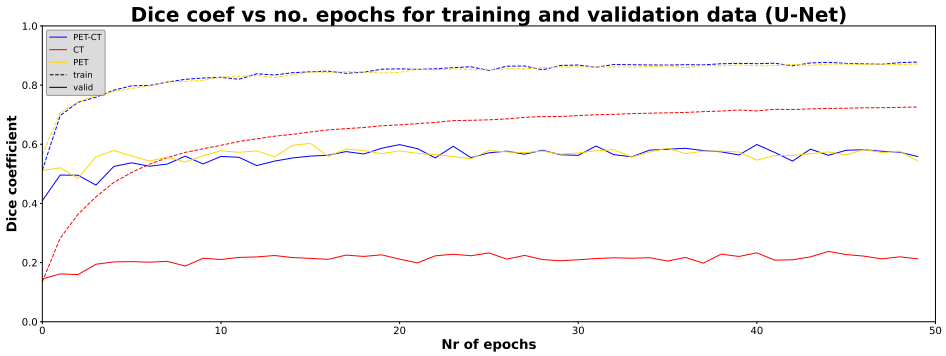

CT
CNN CT Model
U-Net CT Model
FP
FN
TP
PET
CNN PET Model
U-Net PET Model
PET + CT
CNN PET-CT Model
U-Net PET-CT Model


CT
CNN CT Model
U-Net CT Model
FP
FN
TP
PET
CNN PET Model
U-Net PET Model
PET + CT
CNN PET-CT Model
U-Net PET-CT Model

## Appendix C: Development of binary cross-entropy loss and dice coefficient during model training


Dice coef vs no. epochs for training and validation data (U-Net)
PET-CT
CT
PET
train
valid
Dice coefficient
Nr of epochs
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50