Novel Data Adaptation Techniques for Enhanced Lung Cancer Detection in CT Scans

Cătălin–Mihail Chiru[*,1], Mihai Nan[*,2], Mihai Trăscău[*,3], Adina Magda Florea[*,4]

[*] *Computer Science Department, National University of Science and Technology POLITEHNICA Bucharest*, Romania

[1] catalin.chiru@stud.acs.upb.ro, [2] mihai.nan@upb.ro, [3] mihai.trascau@upb.ro, [4] adina.florea@upb.ro

Abstract—Lung cancer persists as a global leader in cancer-related deaths, highlighting the critical need for precise and efficient detection methods. This paper investigates the use of the Medical Segmentation Decathlon dataset to train neural networks for lung cancer segmentation in CT scans via semantic segmentation. We propose and evaluate four new data adaptation techniques specifically designed for this dataset, with each technique being assessed using U-Net-based architectures. Our approach incorporates a thorough exploratory data analysis to uncover the dataset's strengths and weaknesses, which in turn guided our data preprocessing and augmentation strategies.

Index Terms—lung cancer, cancer segmentation, U-NET, 3D volume analysis, repeatability

I. INTRODUCTION

Based on Siegel's 2023 Cancer Statistics [1] over the US population, lung cancer was the 3rd in regard to new cases in 2023 and the deadliest among all the forms. Of the 127,070 lung cancer deaths registered in 2023, approximately 103,000 are caused directly by smoking and 3,560 by second-hand smoke.

The study also shows that with early detection and treatment, the cancer death rate declines, and the lifespan of individuals suffering from it increases, highlighting the importance of automating early detection tasks.

Furthermore, Tan's oncology [2], shows the importance of monitoring Non-Small Cell Lung Cancer (NSCLC), as it represents 85% of all lung cancers in the world. The importance of this aspect comes when analysing the datasets, as there are automated identification tasks developed over NSCLC [3].

Automatic lung tumor segmentation is crucial for both medical professionals and machine learning experts. From a medical point of view, interactive segmentation algorithms can reduce the effort of a radiologist by highlighting cancerous areas. From a technical point of view, the task of obtaining coherent segmentation masks using machine learning techniques is challenging because of at least two factors: the data imbalance between healthy and diseased patients, and the limited number of 2D tumor images and cancers' percentage inside the sections where they appear, in Computed Tomography (CT) analysis.

This paper presents several original contributions to the field of lung cancer detection through automatic CT scan segmentation using Deep Learning:

- We propose and evaluate four innovative data adaptation techniques designed to enhance the training of UNets for lung tumour segmentation tasks. These techniques are specifically tailored for the Medical Segmentation Decathlon dataset and aim to address the inherent variability and challenges within this dataset.
- An in-depth exploratory data analysis to discover the strengths and weaknesses of the Decathlon dataset. This analysis guides our preprocessing and augmentation strategies, ensuring that the dataset's variability is effectively managed during model training.
- Our proposed techniques make us of the 3D aspect of CT scans in order to derrive semantic aspects about tumorous regions, thus sequentially improving the segmentation performance of 3D Convolution based UNets, as evidenced by our experimental results. This leads to more reliable and accurate lung cancer detection.

II. RELATED WORK

Navarrete et al. [3] employ an encoder-decoder U-Net with attention, built on the MONAI framework [4], for lung nodule segmentation and classification. Their method uses NSCLC CT segmentation as a feature extractor for an Epidermal Growth Factor Receptor (EGFR) classification block, where EGFR mutations can indicate cancer by promoting excessive cell growth.

This approach is explored and validated on three Datasets:

- **"NSCLC-Radiomics" (RAD)** [5] (Lung1 dataset from the reference) contains 422 patients with available CT and segmentation masks, 42 of those are removed due to the quality of the segmentations
- **"Medical Decathlon Lung" (MSD)** [6], [7] which contains CTs from 63 patients, 96 3D volumes (64 Training + 32 Testing).
- **"NSCLC-Radiogenomics" (RADGEN)** [8] dataset, used as a test set. It has 211 patients with available CT images. 144 patients have tumor segmentation masks, from which EGFR mutation status is known for 117 (94 wildtypes and 23 mutant) - this aspect is used for the classification task of the pipeline

Another perspective on lung cancer is presented in Pohl et al. [9]: The authors try to predict the motion of the chest during radiotherapy in order to minimise the levels of radiation delivered to healthy tissue around the cancerous zone. They use optical flow to perform Deformable Image Registration (DIR) of CT scans from four patients with lung cancer. Afterwards, a classical RNN is trained through real-time recurrent learning

(RTRL) with gradient clipping to predict and keep track of the position of three internal points close to the tumour.

Another relevant dataset for lung cancer is presented in **"Data from 4D lung imaging of NSCLC"** [10]. It consists of four-dimensional (4D) fan beam (4D-FBCT) and 4D cone beam CT (4D-CBCT) from 20 locally-advanced, non-small cell lung cancer patients undergoing chemoradiotherapy. Besides PapersWithCode, this dataset is part of the Cancer Imaging Archive collection [11], another important source for developing our work.

In Table I we have gathered the main datasets used in the literature for lung abnormality segmentation or detection.

The Medical Segmentation Decathlon (MSD) - Lung dataset stands out due to several key factors that make it an optimal choice for our study. First, its comprehensive 3D volumes, comprising 51 training, 12 development, and 33 test patients, provide a rich and diverse dataset essential for developing robust segmentation algorithms. The high resolution and detailed annotations ensure that models can learn from precise examples, improving their ability to generalize to new data.

The NSCLC-Radiomics (RAD) dataset has 422 labelled patients, but 42 poor quality segmentations limit its effectiveness. The NSCLC-Radiogenomics (RADGEN) dataset has only 144 labelled patients, which may not be enough for training generalized models. LIDC-IDRI dataset is useful for nodule detection and segmentation, but it has limitations in terms of low-dose CT focus and smaller scan numbers, which affect their applicability for 3D volume analysis.

The ELCAP dataset has 50 CT scans and is useful for specific applications, but cannot match the breadth and depth offered by the MSD dataset. Similarly, the NIH Chest X-rays and VinDr-CXR datasets consist of 2D X-ray images, lacking the 3D volumetric data required for detailed tumor segmentation task.

Consequently, the MSD-lung dataset, with its detailed and high-quality 3D volumes, and a well-structured train/dev/test split, is an excellent choice for developing and validating lung cancer segmentation models.

III. DATASET

The Medical Segmentation Decathlon Dataset (MSD) [6] is a comprehensive benchmark in the medical imaging community, encompassing various medical imaging modalities and segmentation tasks. Specifically, for lung cancer, the dataset provides high-resolution CT scans annotated with meticulous labels for lung tumours, facilitating the development and evaluation of advanced segmentation algorithms. Leveraging the rich and varied data in the Medical Decathlon, researchers can train robust models that generalize well across different patient populations, tasks and imaging conditions, addressing the variability inherent in clinical practice.

Our analysis of the Medical Segmentation Decathlon for Lung Cancer involved rigorous exploratory data analyses. This process revealed insights into dataset characteristics, guiding decisions on data preprocessing strategies and model architecture adjustments. Our findings underscore the critical importance of thorough dataset exploration in identifying optimal parameters and enhancing algorithm performance for accurate lung cancer detection.

This Exploratory Data Analysis process included the following steps:

- **Analysis and Findings**: We computed quantitative plots and derived insights from them.
- **Hyperparameter Analysis on Stack-Size**: We analyzed the optimal number of slices per batch and performed oversampling and undersampling techniques accordingly.
- **Signal Removal and Normalization**: We investigated options for removing non-task-specific components and normalizing images when augmenting the dataset.

We have started by analysing the number of CT slices per patient for train and validation (Figures 1 and 4)

A qualitative analysis illustrates that the patients were subjected to different methodologies of obtaining the CTs. For some patients the sampling procedure starts in the lower abdominal area and continues up to their cervical backbone (Figure 2, when limiting to only the lung volume each patient, the plot still exhibits high variance 3 suggesting aspect as large differences in patients' heights or disparity in sampling frequencies between the patients.

We wanted to determine lungs' position in the CTs based on the cancerous regions, as well as the general position of tumours inside them from a 2D perspective and the possible overlap with other organ structures.

At first we split the 2D sections in patches, such that we would get 32×32 patches (Figure 5) and aggregated the tumorous in those patches[1], we also determined the tumorous region at the 512×512 pixel-level (Figure 6) owing to the squared compression factor between the two scales.

Another reasoning for which we computed the full resolution visualisation was for the above-mentioned automated lung-filtering of the train set. We have enlarged the tumorous bounding box heuristically based on random training patients samples and filtered using the Hounsfield value for air inside the box. Later, we manually excluded stomach related frames, as they also passed those trimming conditions.

Starting from the visual difference between the 2D sections, we built an intuition regarding the scale up to which we can condense the 2D feature space. We calculate the average tumour area for all the 2D slices (Figure 7) to find the mean approximately at 110 pixels for a 512×512 image. This suggests that the minimum size of a latent space would be 128×128, as the the scaling factor affects the number of tumorous pixels in a quadratic dependency, on both axes. As a result, in a 128×128 latent image space, the mean of target pixels is $\frac{109.13}{16} \approx 7$ pixels of tumours, which can lead to models sensitive to noise and may influence transposed convolution's ability to correctly decompress information.

The last aspect investigated in our in-depth exploratory data analysis (EDA) regarded tumour positioning on depth.

---

[1] Each patch is 16×16, to downsize from 512×512 to 32×32, thus respecting the idea from Dosovitskiy et al. [16]

TABLE I: Lung related State-of-the-art Datasets

DATASET NAME	MODALITY	DATA QUANTIFICATION	HAS SPLIT
Medical Segmentation Decathlon (MSD) - Lung [6]	CT	51/12/33 patients (3D Volumes) → 26880 CTs	Train/Dev/Test
NSCLC-Radiomics (RAD) [5]	CT	422 labeled patients (- 42 poor quality segmentations)	-
NSCLC-Radiogenomics (RADGEN) [8]	CTs/PETs	211 patients (144 with labels)	-
LIDC-IDRI [12]	low-dose CTs	1010 patients → 1018 CTs	-
ELCAP [13]	low-dose CTs	50 CTs	-
NIH Chest X-rays [14]	XRays	30000 patients → 112000 X-rays	-
VinDr-CXR [15]	XRays	15000/3000 labeled Xrays (14 types of thoracic abnormalities)	Train/Test
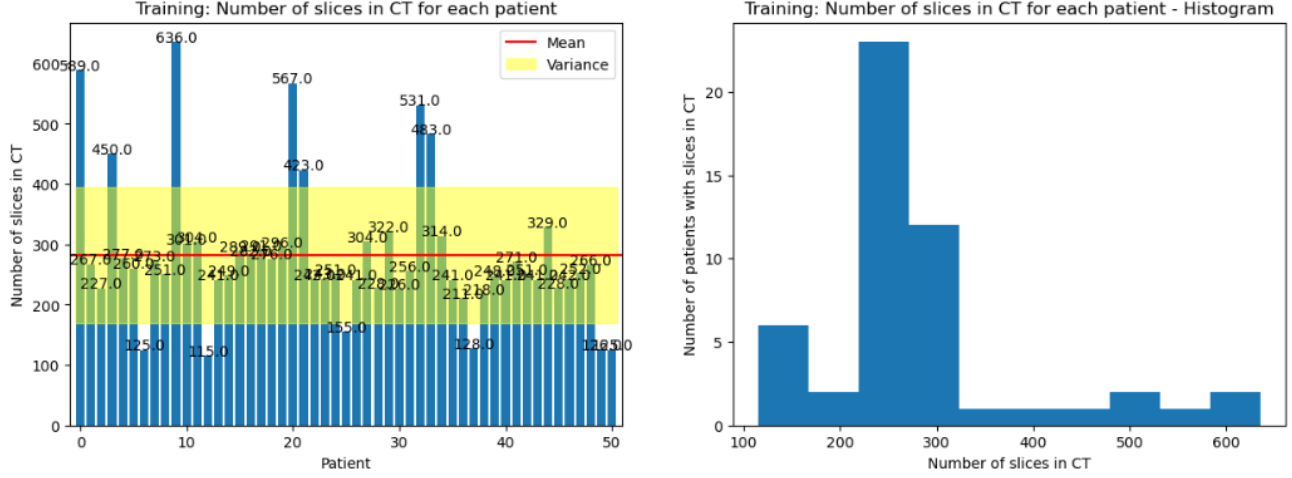
Fig. 1: Bar plots and histograms for the original train slices
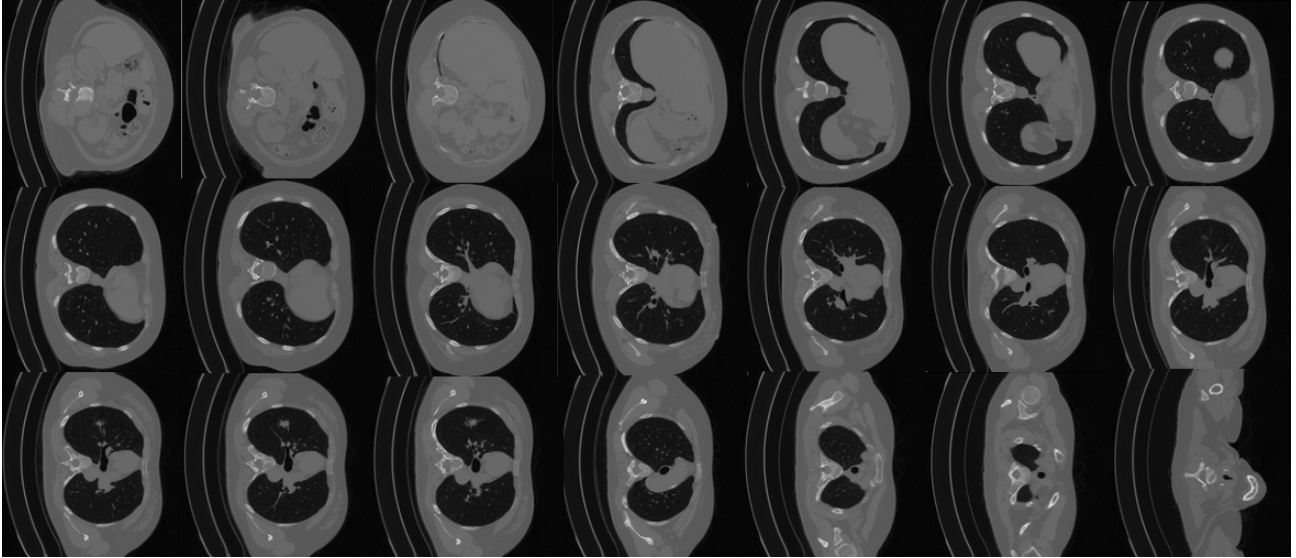
Fig. 2: CT volume from abdomen to upper body - Train Patient 32

On account of the sampling aspects mentioned above, the number of CT slices per patient could not be standardized. We decided to split in bins corresponding to current possible position percentage in the total volume the original dataset and count the number of tumour pixels in each bin (Figure 8).

We observe that the first half of patients' CT is non-tumorous[2], besides Validation's Patient 11. And it can be seen that the tumours appear in the region corresponding to 50-70% of the volume, region that corresponds to highest lungs' surface covering. Finally, and most importantly, patients in

[2]Numerous patients are examined from the lower abdominal area and no lung cancer can be located there.
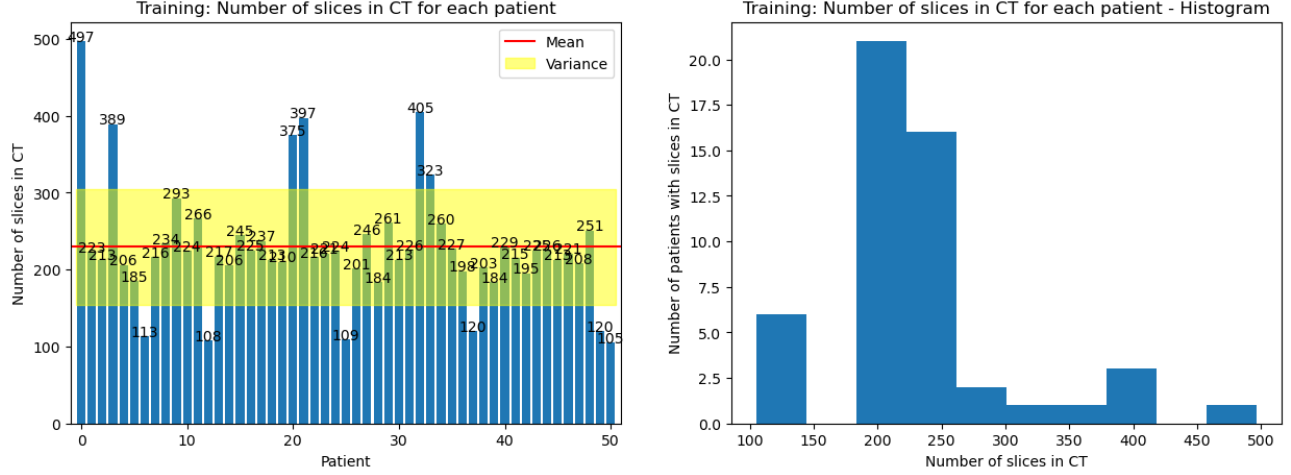
184

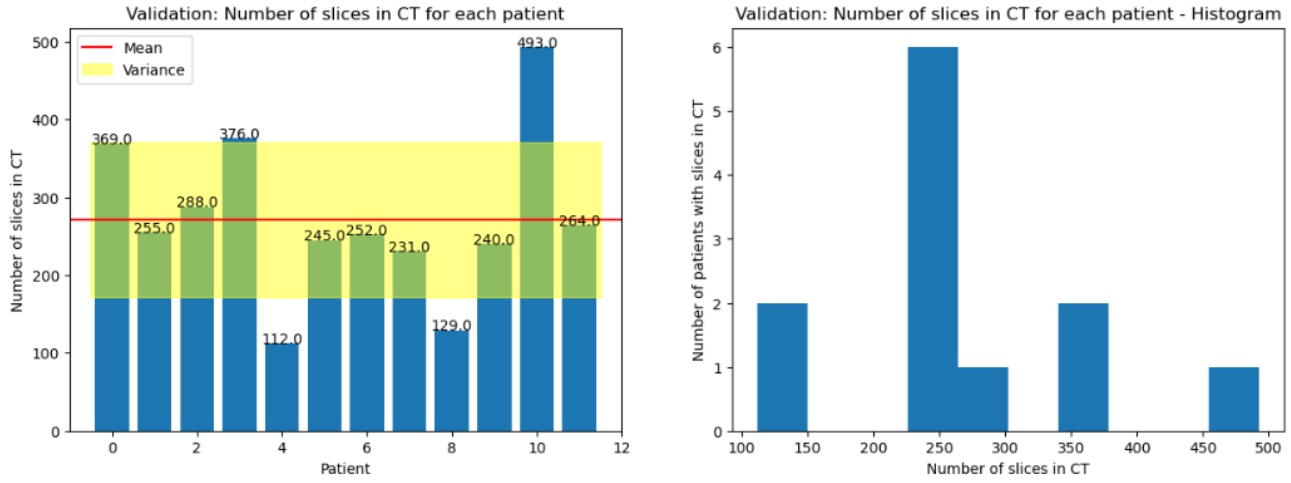Fig. 3: Bar plots and histograms for lung-filtered training patients

Fig. 4: Bar plots and histograms for the validation slices

MSD have continuous tumours, in the sense that no tumour spreads are disconnected from the target volume's main body.

This last observation was essential in our experiments as it permitted us to train from scratch or finetune U-Net models strictly on the padded region with tumorous slices without omitting local context out or imbalancing in favour of healthy regions.

IV. PROPOSED SOLUTION

We proposed and analyzed four potential techniques for training a neural model capable of performing semantic segmentation of CT scans for lung cancer detection. All these techniques are tailored for the Medical Decathlon dataset and were evaluated based on the performance of architectures rooted in the U-Net neural network concept. In-depth implementation details can be found on the Github of this project.

A. Technique I: Baseline

The technique I began with the original dataset, applying standard preprocessing operations for data standardization and normalization. An important aspect highlighted in the graphs presented in Section III is the variability in the number of slices across the available CT scans in the dataset. To facilitate training using batches containing multiple CT scans simultaneously, we introduced a transformation to convert each dataset sample into one with a fixed number, M, of slices.

B. Technique II: Tumour Padded Training

In Technique II, we create a modified dataset based on the original. We analyzed the CT scans of 51 patients in the training set and confirmed the presence of tumours in all of them. These scans are 3D volumes with dimensions for height and width (each 512 pixels) and depth, representing the number of slices in the CT scan. The depth varies among
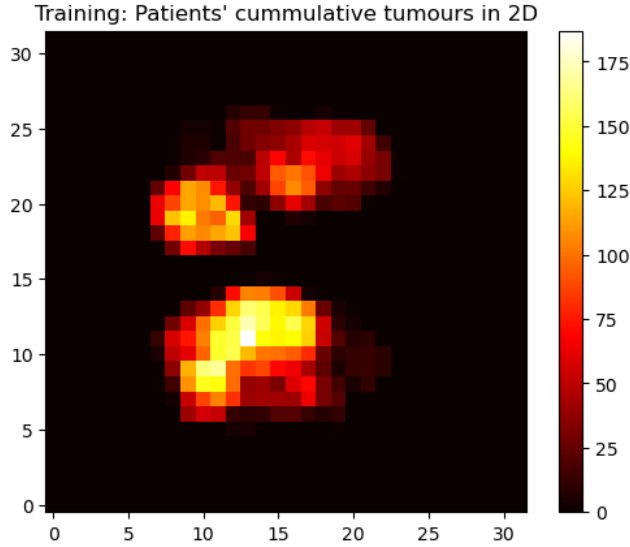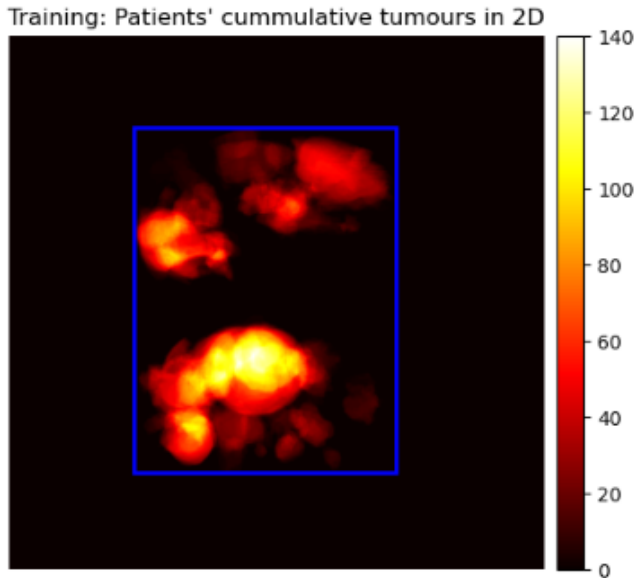
185

Fig. 5: 32x32 2D patches with tumour

Fig. 6: 512x512 2D image with tumour and blue bbox

Fig. 7: Average tumour area in 2D

patients, with only a small proportion indicating the presence of a tumour. In Technique I, we utilized a transformation from MONAI to resize the number of slices. This transformation randomly extracts M slices using MONAI's Resize operation. However, a challenge arises during training when the neural network encounters only a few examples of slices containing tumourous tissue. As a result, it might not be penalized sufficiently for failing to accurately classify non-background pixels.

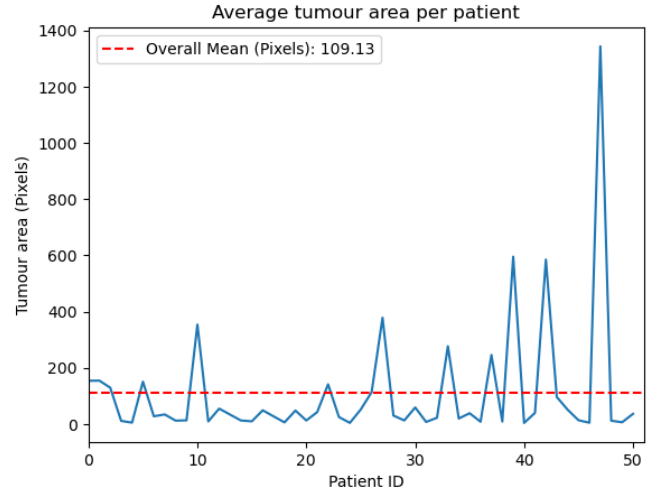Therefore, for this technique, we aimed to retain as many tumour-containing sections as possible for each CT scan by applying a modified resizing approach. Depending on the target value for the third dimension, slices from the initial volume containing tumour tissue were selected, along with adjacent slices that do not contain tumours, until the desired number of slices for the depth of each training sample was reached. This method ensures that resizing is performed while preserving the desired sections, thereby minimizing the risk of losing tumour-containing image examples.

C. Technique III: Training on Lung, Finetuning on Cancer

To develop a highly generalizable model that can function effectively with data from patients who might be healthy, we proposed a training technique that combines the properties of the first two techniques. Thus, Technique III employs both samples from the original dataset and samples adjusted according to Technique II in the training process.

This technique involves training the model in two stages. In the first stage, the model is trained using samples derived from standard preprocessing applied to the original dataset. The model obtained from this initial training stage is then used in a subsequent training process with samples from the modified dataset proposed in Technique II. This process aims to refine the model, enhancing its ability to learn the structures present in tumour tissue areas, thereby improving the overall quality of segmentation. The second training phase will maintain the same configuration but will use a lower learning rate.

D. Technique IV: Balanced normalised minivolume batches

Experiments in previous techniques showed significant drops in performance or stagnation of the loss function during training. In Technique IV we tried a different approach that would be easily replicable and would ameliorate the limitations of the previous variants: We split all the volumes on depth in mini-samples, performed hyper-parameter analysis on the number of CT slices[3] to be taken in a mini-batch,

---

[3]In this work we refer to the terms CT slices and CT stacks interchangeably.

186

(a) Train Patients' Cancer Heatmap on Depth

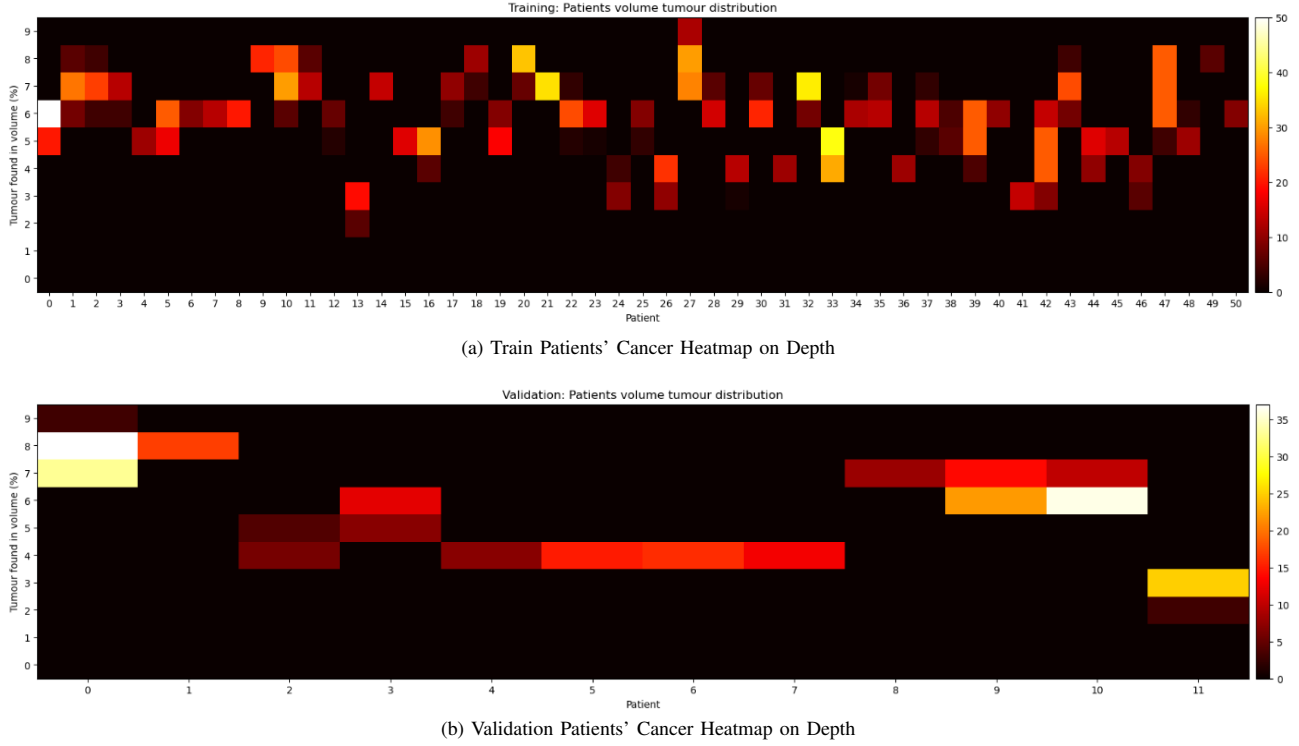
(b) Validation Patients' Cancer Heatmap on Depth

Fig. 8: Tumour localization on Depth

limited the number of affine augmentations and reduced the stochasticity in the experiments.

We identified that the smallest tumour mask in the entire dataset extends over 6 frames, whereas the ideal mini-samples would not contain the tumour boarded by healthy regions, as that would hinder models' ability of learning the structure of cancers. Therefore, we set the depth for each sample to 6.

Further, we applied an overlapping sliding window with unitary stride to the training split, thus obtaining 14148 mini-samples. Out of those 12517 contain just healthy stacks, 744 have tumour volume percentages smaller than $10^{-5}\%$ and 1493 have the tumour volume greater or equal to 0.125%. We have considered as cancerous only the last portion, meaning that the class imbalance is 89.45-10.55 between healthy samples and ones with various volumes of tumour.

When establishing how to select samples to group them in batches, we have tackled two strategies: taking the entirety of the healthy portion and oversampling by repetition the tumorous samples or oversampling them with a factor and undersampling the healthy ones to respect a given final distribution. The experiments that involved undersampling produced better results for the same resources[4] and running times, as models saw each sample more during the training, whereas the oversampling procedure would need times bigger by an order of magnitude for the same number of epochs and would give significantly worse results for the same setup.

[4]A DGX station equipped with 4 Tesla V100 DGX 32GB GPUs

We explored both a Curriculum learning setup, in which we fed the model in an epoch batches from high tumour percentages to healthy, and a random shuffling with the same samples. Experiments showed that the Curriculum approach was not suited for this task, as the network first learnt to focus on the tumour and then forgot it to the point to which weights vanished.

This insight was later supported by applying Technique III with batches formed with random shuffling, as experiments showed that the models stabilized and produced consistent results when first training on the entire lung section and then finetuning on the cancerous region, rather then the inverse setup.

To avoid overfitting on the tumorous area and cancer over-prediction on unknown structures, we apply identical small rotation to individual samples at each epoch, in order to simulate patients' possible movements on the CT bed.

We mask the 2D surfaces outside the bounding box area mentioned in Section III to force the model to not concentrate on the bed or other surrounding aspects. At last, we apply Min-Max Normalization to move the images from the Hounsfield interval to $[0, 1]$ and Gradient Clipping to control models' gradient vanishing/explosion tendency.

V. EXPERIMENTAL RESULTS

All the experimental results reported in Table II were obtained on an augmentation of the standard validation split given by MONAI DecathlonDataset for the Lung Challenge

of MSD. In this setting, we took images sizes 128×128 on a region of 64 slices, centered and padded around the tumour, with the limitation of truncating patient's 0 cancerous zone, which was bigger. This validation subset was obtained after applying the processing steps proposed in Technique II and selecting the most relevant 64 slices for each patient for the Tumour Segmentation Task.

Architecture	No. of layers	Technique	Dice score
UNet	3	I	0.2009
UNet	4	I	0.1689
UNet	3	II	0.5668
UNet	4	II	0.5393
UNet	3	III	0.6513
UNet	4	III	0.5901

TABLE II: The results obtained applying the Techniques I, II and III

The Technique I involves training on the original dataset with standard preprocessing applied. Technique II involves training only on the modified dataset. Technique III is a two-stage approach: initially training on the original dataset, followed by fine-tuning on the modified dataset. In the second training stage, we maintain the same configuration but use a lower learning rate.

The results summarized in Table II demonstrate the performance differences across the three techniques applied to the UNet architecture with varying depths. Technique I resulted in relatively low Dice scores of 0.2009 and 0.1689 for three and four layers, respectively. This indicates that standard preprocessing alone is insufficient for effectively segmenting lung tumours due to the high variability and imbalance in the dataset.

Technique II significantly improved the Dice scores to 0.5668 and 0.5393 for three and four layers, respectively. This enhancement underscores the importance of retaining relevant tumour information during training, as it provides the model with more examples of critical features necessary for accurate segmentation.

Technique III yielded the highest Dice scores of 0.6513 and 0.5901 for three and four layers, respectively. This two-stage approach demonstrates the benefits of initial broad learning followed by focused refinement, allowing the model to generalize well across diverse patient data while also learning specific tumour characteristics more effectively.

Overall, these results validate our approach, showing that combining standard preprocessing with targeted data augmentation and fine-tuning can substantially enhance model performance in medical image segmentation tasks. This method ensures a balanced learning process, capturing both general patterns and specific features critical for accurate tumour detection.

Those results were obtained on an augmentation of the validation split from the MSD, with images sizes 128x128 on a region padded around the tumour.

Compared with other models submitted for the development phase of the Medical Segmentation Decathlon, based on the

L2 (WD)	Data augments	Dice score (OG)	Dice score (Aug)	Training Time (h)	BATCHES USED
1e-3	BATCH SPLITS	0.2066	0.5703	17	77
1e-3	**Trim, Norm, Cancer**	**0.23**	**0.56**	**20.84**	**77**
1e-4	Trim, Norm, Lung → Cancer	0.2028	0.5053	14 +21.14	159 +77

TABLE III: Quantitative results UNet Based experiments

Supplementary Information for the competition [6], Technique IV would have positioned in top 16, above the last 4 results submitted in the development phase. However, our score is strongly affected by numerous False Positives, a numerical analysis on the tumorous volume offers a dice coefficient of 0.5776, aspect which would significantly improve the positioning of our approach. As a future contribution, we will look into a possible thresholding mechanism to increase the number of zero masks and we would apply slice-level classification to further reduce the rate of overpredicting cancer.

The results in Table III were obtained on the original validation split from the MSD, with images sizes 512×512 ("Dice score (OG)"). For transparency with the previous Techniques and results that were run on an environment with limited resources[5] we have resized with interpolation and reported Technique IV's results on the augmented dataset as well ("Dice score (Aug)"). We have **bolded** best results and underline second bests, or bests with high variance.

All the experiments presented in Table III were trained on 8M-parameters Residual U-Net, with Dice Loss[6] as a loss function.

Although we have tested other configurations, the 3 experiments highlighted use Nesterov accelerated Adam (NADAM) with a learning rate of $5e^{-3}$ as this optimizer performed the best, furthermore, L2 regularization.

Moreover, to compress the table we have used the following abbreviations for our Table:

"L2 (WD)" denotes the use of L2 Regularization with "weight decay" as a set hyperparameter. "Batch Splits" denotes Technique IV's baseline batch construction in which the minivolumes are undersampled with a proportion, shuffled and stacked. "Trim" indicates the presence of masking non-lung signal, "Norm" indicates the use of "Min-Max" Normalization and "Cancer", "Lung", "Cancer → Lung", "Lung → Cancer" indicate the structure of the dataset from which samples were drawn. "Lung"/"CANCER" contain all the training CT stacks in the specified area with paddings of continuous healthy regions, such that the samples' distribution of tumorous and non-tumorous stacks is even for the latter. The arrow → indicates that we first trained from scratch on one dataset,

saved the best checkpoint and fine-tuned on it using the other dataset.

"Batches used" quantifies the number of samples used in a training epoch, as the chosen batch size was fixed to 16 and we can deduce fast other aspects from this column.

For the evaluation metrics, "Dice Score" is based on MONAI's DiceHelper implementation that produces 0s if False Positives are detected, and 1 if the mask matches healthy regions.

From an analytical perspective, the Table highlights the best 3 experiments obtained using Technique IV. In the first experiment, we undersampled evenly healthy and cancerous minisamples to build our batches. This attempt represented our first major improvement in Dice in this setup and corresponded to the model actually learning the local context of the tumours. The limitation of that experiment was that its gradients still exploded after epoch 68, exhibiting a saw tooth loss behaviour, indication of the model being stuck between local minima.

Running the same setup with min-max normalization, masking of other sources of signal and undersampling in the region of training tumours lead to our best model. Its only limitation resides in the low Dice Score caused by the increased number of False Positive images with small noisy dots.

Our final experiment followed Technique III, by applying those transformations, training from scratch on the under-sampled dataset from the lung region, where the distribution was 60%-40% in favour of healthy samples, and finetuning on the undersampled tumorous region evenly distributed. The evolution of the loss and evaluation metrics in this experiment was the smoothest. By comparison its base get the highest performance at the 16th epoch and plateaued and saturated at a worse results in another 16 epochs. The comparison shows massive improvements in evaluation metrics between the two, as finetuning improved base's Dice by 7%. However, it produces qualitative visualisations with false positives (FP) areas more prominent than the general best.
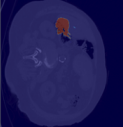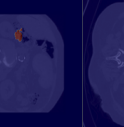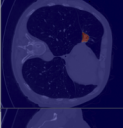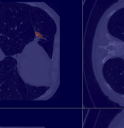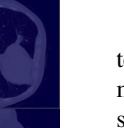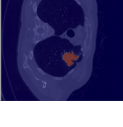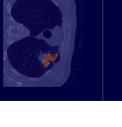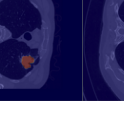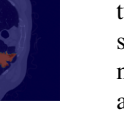

Fig. 9: Qualitative results for Technique IV

In the Figure 9 we rolled out significant slices of the first validation patient obtained through the 3 best models of

Techniques IV. The first line shows overprediction in non-lung areas, the second exhibits the same behaviour with bronchi and the last illustrates the percentage of tumour correctly segmented, along with the ground truth.

VI. CONCLUSIONS

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, necessitating the development of precise and efficient diagnostic tools. Advances in medical imaging and machine learning have shown great promise in improving the early detection and treatment of lung cancer. In this context, the segmentation of CT scans plays a crucial role in identifying and delineating lung tumours, which is essential for accurate diagnosis and subsequent treatment planning.

Our experiments clearly demonstrate that the proposed data adaptation techniques enhance the performance of lung tumour segmentation. Technique II, which focuses on preserving tumor-containing slices, and Technique III, which combines broad learning with focused refinement, have both shown significant improvements over the standard preprocessing approach. Specifically, Technique III's two-stage training process achieved the highest Dice scores, highlighting its effectiveness in leveraging initial broad learning followed by fine-tuning tumour-specific data. Technique IV supports the training in stages, as most robust results are obtained through this procedure. Further, it illustrates that best experiments share a higher number of epochs, with final batches built on samples of volumes padded around the tumours. At last, Technique IV motivates the use of specific augmentations in controlling backpropagation's behaviour.

These findings emphasize the importance of incorporating advanced data adaptation techniques in medical image segmentation tasks. By addressing the inherent variability and imbalance within datasets, our proposed methods significantly improve the reliability and accuracy of lung cancer detection models. The practical utility of these techniques is evidenced by the marked increase in segmentation performance, confirming their value in real-world medical applications. Integrating such innovative approaches into clinical workflows can potentially enhance early detection and treatment outcomes for lung cancer patients, ultimately contributing to improved patient care and survival rates.

VII. FUTURE WORK

Future work can explore the adaptation of the proposed techniques for other pathologies included in the Medical Segmentation Decathlon dataset. By applying our data adaptation strategies to different medical imaging tasks, such as brain tumour segmentation, cardiac segmentation, and liver tumour segmentation, we can assess the generalizability and robustness of our methods across various medical domains. This approach would validate the versatility of our techniques and contribute to developing comprehensive segmentation tools that can address a wide range of medical conditions. Additionally, exploring these techniques in multi-modal imaging contexts, such as combining CT, PET and MRI data, could

further enhance the accuracy and reliability of automated segmentation models.

Further, stopping tumour overprediction should be an essential task in developing this work. To achieve this, we could consider a thresholding hyperparameter for the predicted masks, an auxiliary healthy/cancerous classification task and a clustering method on the healthy samples to decide which are "the centroids" of different groups. By applying those techniques, we assure False Positive filtering methods to alleviate models' shortcommings.

Other forms of normalisation, such as CLAHE, have been applied in literature [17] with success on other medical modalities, which might contribute to our current setup, and an ablation study on the stack hyperparameter would validate our 6-slices choice.

To reduce the training time and the number of overlapping subvolumes, we could simulate "scan time reduction" by skipping slices in the volume. This technique was applied in the Radiology Literature [18], to reduce patients exposure to radiation and to mitigate possible artefacts resulted from respiratory motion.

At last, a GAN-based segmentation architecture on 3D patches would benefit a better volumetric understanding of tumours, as GANs are already used to enhance Computer Tomography in patients with cancer [19].

ACKNOWLEDGMENT

REFERENCES

[1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *Ca Cancer J Clin*, vol. 73, no. 1, pp. 17–48, 2023.

[2] M. Winston W Tan, "Non-small cell lung cancer (nsclc)." https://emedicine.medscape.com/article/279960-overview?form=fpf. Last accessed: 2nd of December 2023.

[3] I. G. Navarrete and M. Yaqub, "A radiogenomics pipeline for lung nodules segmentation and prediction of egfr mutation status from ct scans," *arXiv preprint arXiv:2211.06620*, 2022.

[4] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, *et al.*, "Monai: An open-source framework for deep learning in healthcare," *arXiv preprint arXiv:2211.02701*, 2022.

[5] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, no. 1, p. 4006, 2014.

[6] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, *et al.*, "The medical segmentation decathlon," *Nature communications*, vol. 13, no. 1, p. 4128, 2022.

[7] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.

[8] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, J. A. Benson, W. Zhang, A. N. Leung, *et al.*, "A radiogenomic dataset of non-small cell lung cancer," *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.

[9] M. Pohl, M. Uesaka, K. Demachi, and R. B. Chhatkuli, "Prediction of the motion of chest internal points using a recurrent neural network trained with real-time recurrent learning for latency compensation in lung cancer radiotherapy," *Computerized Medical Imaging and Graphics*, vol. 91, p. 101941, 2021.

[10] G. D. Hugo, E. Weiss, W. C. Sleeman, S. Balik, P. J. Keall, J. Lu, and J. F. Williamson, "Data from 4d lung imaging of nsclc patients," *The Cancer Imaging Archive*, vol. 10, p. K9, 2016.

[11] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, pp. 1045–1057, 2013.

[12] L. Lin, Y. Wu, S. Song, and S. Wu, "Lung image database consortium and image database resource initiative: reviewing progresses toward improved computer-aided diagnosis of pulmonary nodule," *Chinese Medical Equipment Journal*, vol. 39, no. 10, pp. 95–99, 2018.

[13] A. P. Reeves, A. M. Biancardi, D. Yankelevitz, S. Fotin, B. M. Keller, A. Jirapatnakul, and J. Lee, "A public image database to support research in computer aided diagnosis," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3715–3718, IEEE, 2009.

[14] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.

[15] H. Q. Nguyen, K. Lam, L. T. Le, H. H. Pham, D. Q. Tran, D. B. Nguyen, D. D. Le, C. M. Pham, H. T. Tong, D. H. Dinh, *et al.*, "Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations," *Scientific Data*, vol. 9, no. 1, p. 429, 2022.

[16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[17] A. H. Nizamani, Z. Chen, A. A. Nizamani, and K. Shaheed, "Feature-enhanced fusion of u-net-based improved brain tumor images segmentation," *Journal of Cloud Computing*, vol. 12, no. 1, p. 170, 2023.

[18] A. Bryce-Atkinson, T. Marchant, J. Rodgers, G. Budgell, A. McWilliam, C. Faivre-Finn, G. Whitfield, and M. Van Herk, "Quantitative evaluation of 4d cone beam ct scans with reduced scan time in lung cancer patients," *Radiotherapy and Oncology*, vol. 136, pp. 64–70, 2019.

[19] P. Yang, X. Ge, T. Tsui, X. Liang, Y. Xie, Z. Hu, and T. Niu, "Four-dimensional cone beam ct imaging using a single routine scan via deep learning," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1495–1508, 2022.