

MLC at HECKTOR 2022: The Effect and Importance of Training Data when Analyzing Cases of Head and Neck Tumors using Machine Learning

Vajira Thambawita¹, Andrea M. Storås^{1,2}, Steven A. Hicks¹,
Pål Halvorsen^{1,2}, and Michael A. Riegler^{1,3}

¹ SimulaMet, Norway

² Oslo Metropolitan University, Norway

³ UiT The Arctic University of Norway, Norway

Abstract. Head and neck cancers are the fifth most common cancer worldwide, and recently, analysis of Positron Emission Tomography (PET) and Computed Tomography (CT) images has been proposed to identify patients with a prognosis. Even though the results look promising, more research is needed to further validate and improve the results. This paper presents the work done by team MLC for the 2022 version of the HECKTOR grand challenge held at MICCAI 2022. For Task 1, the automatic segmentation task, our approach was, in contrast to earlier solutions using 3D segmentation, to keep it as simple as possible using a 2D model, analyzing every slice as a standalone image. In addition, we were interested in understanding how different modalities influence the results. We proposed two approaches; one using only the CT scans to make predictions and another using a combination of the CT and PET scans. For Task 2, the prediction of recurrence-free survival, we first proposed two approaches, one where we only use patient data and one where we combined the patient data with segmentations from the image model. For the prediction of the first two approaches, we used Random Forest. In our third approach, we combined patient data and image data using XGBoost. Low kidney function might worsen cancer prognosis. In this approach, we therefore estimated the kidney function of the patients and included it as a feature. Overall, we conclude that our simple methods were not able to compete with the highest-ranking submissions, but we still obtained reasonably good scores. We also got interesting insights into how the combination of different modalities can influence the segmentation and predictions.

1 Introduction

Head and neck cancers are among the most common cancer types worldwide. Early detection is critical as the tumor’s size on diagnosis will dictate the patient’s quality of life and chances of survival [10]. Medical image analysis and

radiomics have shown promising results in detecting different diseases and cancers [16,9,2], including those found in the head and neck [20,25,18]. In this paper, we describe our approaches for the HEad and neCK TumOR (HECKTOR) grand challenge held at MICCAI 2022 [1,17]. Of the two tasks presented at the challenge, we participated in both. In Task 1, the aim was to segment tumors from Computed Tomography (CT) and Positron Emission Tomography (PET) scans of the head and neck (examples shown in Figure 1). Task 2 asked for the prediction of Recurrence-Free Survival (RFS) based on clinical information about the patients presented in a tabular format, which also could be combined with the outputs from Task 1.

As the provided dataset contained different types of data, our strategy to tackle the HECKTOR challenge was to explore how the inclusion and combination of different modalities change the prediction outcome. In this respect, we investigated how CT and PET scans can be used individually or combined for tumor segmentation in Task 1 and how RFS can be predicted using the meta-data with or without tumor information for Task 2. The main contributions of this paper are as follows:

1. A comparison of simple segmentation methods using CT or PET slices individually or combined.
2. Understanding the effect of combining different data modalities on the analysis results.
3. Analysis of what features were most relevant for predicting RFS using patient-related data and image features.

2 Methods

In this section, we describe the methods we applied to solve Task 1 and 2, respectively.

2.1 Task 1: Segmentation of CT and PET scans

For Task 1, we used the provided development dataset consisting of CT scans, PET scans, and corresponding segmentation masks. we experimented with three different approaches as follows (see Figure 2):

Approach 1: Only using individual slices of the CT scans to predict tumors with a UNet++-based model [28].

Approach 2: Combining the CT and PET scans by stacking CT, PET, and the mean of CT and PET images channel-wise and passing them through a UNet++-based model.

Approach 3: Analyzing CT and PET slices separately in an ensemble-like setup using a TriUnet-based model [24].

These three approaches utilize the data provided in the HECKTOR competition differently, from simple to more complex. The following sections describe all the steps of data pre-processing, sampling, augmentation, implementation of the models, and post-processing.

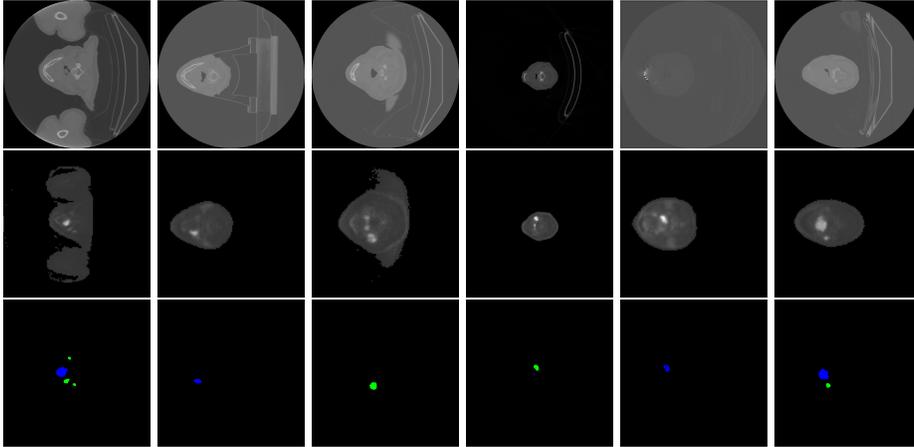


Fig. 1. Six example data points from the development dataset provided by the HECKTOR organizers. The first row contains a slice from one of the CT images, the second row contains a slice from one of the PET images, and the third row contains the corresponding segmentation mask. Please note that the brightness of the PET images has been adjusted so that the contents are more easily visible.

Image Data pre-processing: We divided the development dataset into a training and a validation dataset containing 90% and 10% samples, respectively. For Approach 1, we extracted the slices from the CT and ground truth as *.png* images without applying re-sampling because the shape of the CT and provided ground truth were the same. However, for Approaches 2 and 3, we performed slice extraction after re-sampling (using SimpleITK [26]). We used the same re-sampling as provided by the task organizers⁴ with default spacing (2, 2, 2). In addition, we normalized all CT and PET images into the range between 0 and 255, but not ground truth images that contain pixel values of [0, 1, 2]. After the extraction process, we noticed that the training dataset contained large number of true negative samples. Therefore, to avoid bias, we re-balanced the training dataset by extracting only slices with true positive pixels for H&N Primary tumors (GTVp) and H&N nodal Gross Tumor (GTVn). The class rebalancing was done by combining an equal number of true positive slices with the true negative slices extracted from the initial training dataset. To make a challenging validation dataset, we extracted only slices with GTVp and GTVn from the validation images.

We applied similar image augmentation for all three approaches. The Albu-mentations [6] library provides a set of augmentation options for image segmentation tasks. More information about the input parameters of the augmentation methods can be found in our GitHub repository⁵.

⁴ https://github.com/voreille/hector/blob/master/src/resampling/resample_2022.py

⁵ https://github.com/vlbthambawita/hector_2022_MLC

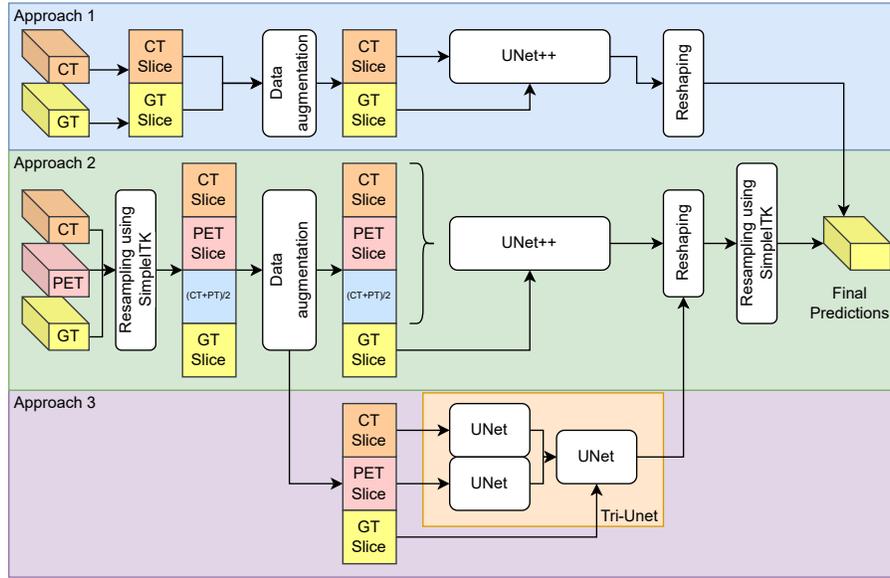


Fig. 2. Three approaches used for Task 1. Approach 1: uses only CT images and the corresponding ground truth (GT). Approach 2: input stack of CT, PET, mean of CT and PET. Approach 3: use two separate UNet models for CT and PET and another UNet for final predictions. Reshaping sizes used in Approach 1 is different from the sizes used for Approach 2 and 3.

Model architectures, hyperparameters and inputs: The models for Task 1 were implemented in Pytorch [19] using the Segmentation Models library [12]. All models were trained for 100 epochs on hardware consisting of two Nvidia RTX 3080 Graphic Processing Units (GPUs) with 10 GB of memory each, an AMD Ryzen 9 3950X 16-core processor, and 64 GB memory. Submissions were made with the best-performing checkpoints, which were selected based on the performance on the validation dataset. For the first 50 epochs, the learning rate was set to 0.0001, then reduced to 0.00001 for the remaining 50 epochs. The Adam optimizer [14] with default parameters except the learning rate was used for all the experiments. Furthermore, we have used DiceLoss with skipped channel 0 as the main loss function in the training process and the Intersection over Union (IoU) as a metric to evaluate our models.

In Approach 1, we have used a UNet++-based model with *se_resnext50_32x4d* as the encoder. The model was trained using only single channel CT input images and the corresponding ground truth masks after resizing them into 256×256 in the augmentation step.

For Approach 2, we re-sampled the CT and PET slices and trained a UNet++ model. For this approach, we stack a CT slice, a PET slice, and the mean of the CT and PET slice in the color channel and use these as input to the model. The main objective of the second approach is to gain more information about using

a UNet++-based architecture without making any major changes from the first approach.

Approach 3 used a different architecture, TriUnet, which we introduced in our previous study [24]. In this model, we input re-sampled single channel CT slices into one UNet [21] and PET slices into another UNet. Then, the output of the two networks was passed through another UNet model, which accepts six input channels (3 channels output from the first and second UNet model for representing three classes of the ground truth). We used the same hyperparameters and trained the network as a single model using a single back-propagation step. The reason for not using UNet++ for this approach was mainly due to the memory limitations imposed by our GPU.

Post-processing and submission preparations: For all approaches, we re-shaped the test images into 256×256 , which is the size of training data. Then, we re-shaped the predicted segments back to the original shape of CT images using re-sampling. However, we had to re-shape the predictions back to the shape of re-sampled input data before re-sizing them into the original shape. In both re-shaping methods, INTER_CUBIC interpolation introduced in OpenCV [13] library was used.

For all approaches in Task 1, we used the academic version of Weights and Biases [3] for tracking and analyzing experiments and the corresponding performance. All the experiments with the corresponding best checkpoints are available on GitHub⁶.

2.2 Task 2: Prediction of Recurrence-Free Survival

Estimation of kidney function We include the estimated kidney function as a feature for the XGBoost model from the third approach of Task 2 as this might improve the predictions of RFS. Prior research indicates that chronic kidney disease can worsen the prognosis of cancer patients and that monitoring the kidney function of cancer patients is crucial [22].

The feature is created using the Cockcroft-Gault formula, which is among the most widely used formulas for estimating the kidney function [8]. This formula requires the gender, age, body weight and serum creatinine concentration. Because serum creatinine is not available in the dataset, the average values for men and women are used instead [23]. Indeed, when plotting the correlation matrix for the training data, we observe that there is a positive correlation between the estimated kidney function and the RFS (correlation = 0.26), indicating that higher kidney function is associated with a longer time to recurrence. The entire correlation matrix for the training dataset is shown in Figure 5.

Description of the approaches For Task 2, we proposed three different approaches. The first approach used only the patient data, while the second and

⁶ https://github.com/vlbthambawita/hecktor_2022_MLC

third approach also included features based on the image data. The image features arrived from the segmentation masks from the best approach of Task 1. Specifically, we calculated the number of pixels per class from the predicted masks in addition to the number of slices of the CT images in the z-dimension, resulting in four additional features. Moreover, the third approach used the estimated kidney function as a feature. For all three approaches, we used 10-fold cross-validation on the development data to determine the best hyperparameters and model. The hyperparameters are selected based on the root mean squared error (RMSE) of the model, which should be as low as possible. The final models are trained on the entire training dataset using the identified set of hyperparameters. The resulting models are then used on the test dataset to make the predictions for the challenge evaluation. All experiments are performed using the scikit-learn library [5].

Approach 1: The first approach used the Random Forest [4] algorithm to predict RFS using only the patient data. Random Forest was chosen because it is known to work well on tabular data and is often used as a baseline for medical-related machine learning problems. All features provided in the training data were used besides the patient ID. Based on the cross-validation results (RMSE of 988.47), the hyperparameters for the Random Forest were set as the following; max features as the number of features divided by three, and the number of trees was set to 100. All other hyperparameters used the default value set by scikit-learn.

Approach 2: For the second approach, we used the same algorithm as the first approach, but with additional image features as described in the beginning of the subsection. The RMSE from the cross-validation of the training data was 962.83, which was an improvement compared to the first approach showing that the inclusion of image data has a positive effect on the results. The hyperparameters used for the Random Forest in the second approach are as follows; max features as the number of features and the number of trees 200. All other hyperparameters used the default value as set by scikit-learn.

Approach 3: Regarding the third approach, an XGBoost [7] regression model was trained to predict RFS using the available patient data and the image features with one additional feature representing the estimated kidney function. The feature representing alcohol consumption was removed because the majority of the patients in both training and test dataset do not have any registered value for this feature. The patient ID was not included in the training dataset. The RMSE from the cross-validation on the training data was 909.09. The hyperparameters for the XGBoost model are: ‘n_estimators’ = 120, ‘learning_rate’ = 0.05, ‘max_depth’ = 4, ‘subsample’ = 0.7, ‘colsample_bytree’ = 0.6, ‘colsample_bynode’ = 1 and ‘colsample_bylevel’ = 0.8. The other hyperparameters used the default value.

Investigating feature importance After training the XGBoost model, feature importance is explored using Shapley additive explanations (SHAP) [15].

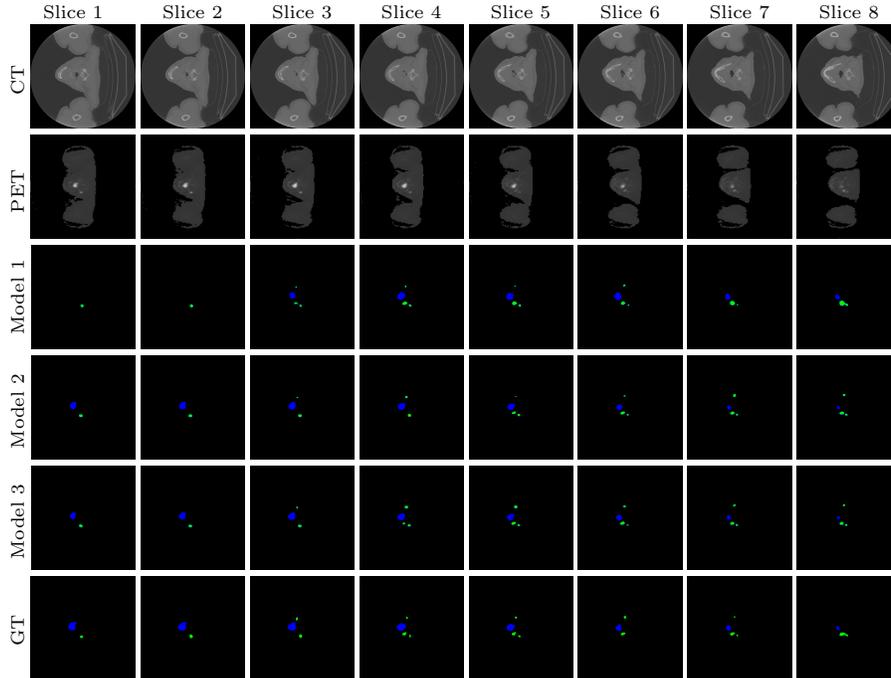


Fig. 3. Example predictions taken from each model at different slices and the corresponding ground truth.

Table 1. Official HECKTOR Challenge 2022 results for Task 1 using the mean aggregated Dice metric for the ranking. The best values are marked using **bold text**.

ID	Model	Dataset	Validation			Test		
			GTVp	GTVn	Mean	GTVp	GTVn	Mean
1	UNet++	CT	0.560	0.721	0.641	0.466	0.536	0.501
2	UNet++	CT + PET	0.601	0.674	0.638	0.607	0.604	0.605
3	TriUnet	CT + PET	0.671	0.722	0.696	0.659	0.654	0.657

SHAP approximates Shapley values, which origin from game theory and assigns values to the features based on how much they contribute to the prediction [15,27]. Consequently, it is possible to investigate which features the model regards as most important.

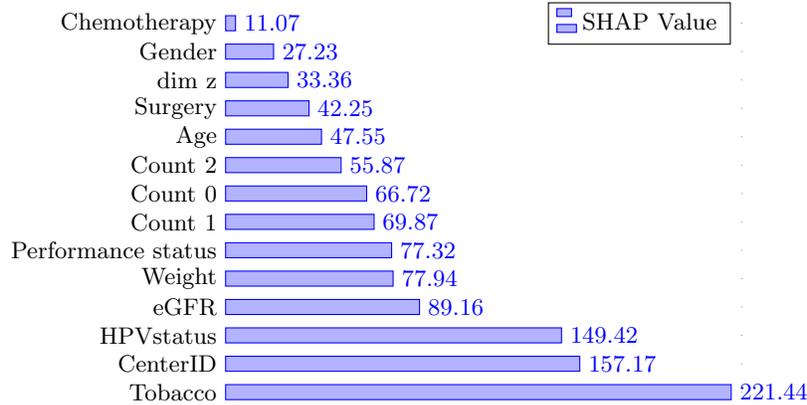
3 Discussion and Results

3.1 Task 1

Looking at the results for Task 1 in Table 1, we see that the first approach struggles to detect GTVp in both the validation and the test datasets. This is

Table 2. Official HECKTOR Challenge 2022 results for Task 2 using C-index for the ranking. The best values are marked using **bold text**.

ID	Model	Data	C-index
1	Random Forest	Patient data	0.585
2	Random Forest	Patient and image data	0.589
3	XGBoost	Patient and image data	0.656

**Fig. 4.** Feature importance for the XGBoost model from Task 2, trained on image features and patient information. Count 0, Count 1, Count 2 and dim z are image features.

also shown in Figure 3, where the first model is unable to detect the presence of GTV_p until the third slice. Despite not being able to detect GTV_p very well, the first approach performs well on segmenting GTV_n on the validation dataset, but not on the test dataset. This can most likely be attributed to the differences between the validation and test datasets, as the other approaches show similar results. Adding information from the PET scans for the latter two approaches seems to help in detecting GTV_p, as evident by the improved scores in Table 1, and they are both able to detect GTV_p in all eight slices from the example in Figure 3. The differences between Approaches 2 and 3 indicate that extracting features from the CT and PET scans independently seems to be the most suitable technique.

3.2 Task 2

For Task 2, the results can be seen in Table 2. From the results, we can observe several interesting insights. Firstly, adding additional data to the patient data gives better predictions. This can be observed in the difference between approach 1 and 2 when image data was added as additional features. We can also observe that XGBoost outperforms Random Forest by a large margin. This correlates

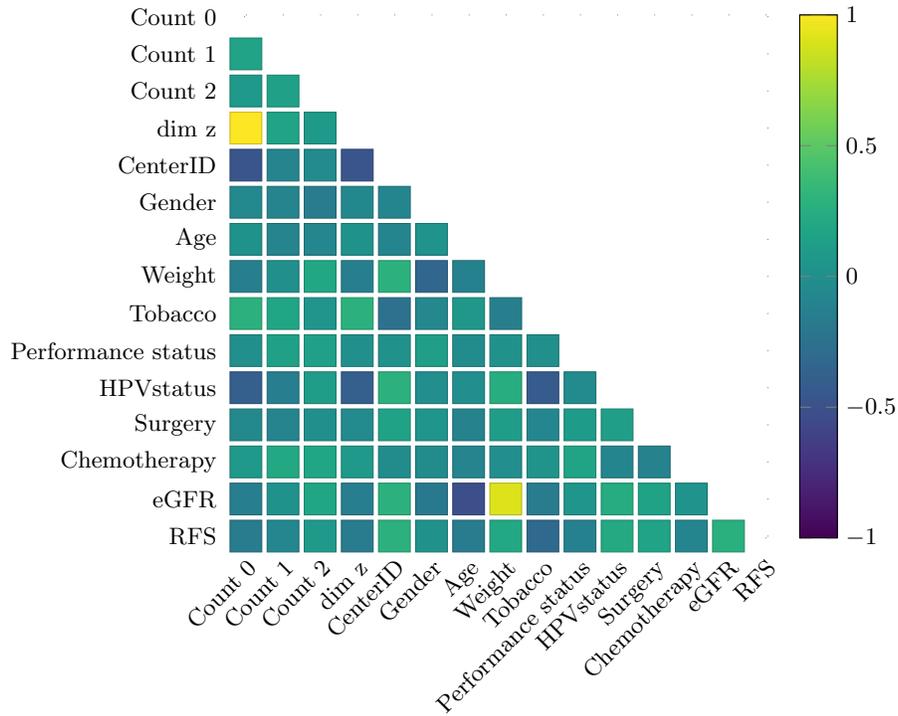


Fig. 5. Correlation matrix for the training data applied in Task 2 for the third approach. Count 0, Count 1, Count 2, and dim z are image features.

with general findings in the literature that XGBoost is one of the best working methods. This questions also the general concept of using Random Forest as a baseline and suggest that in general it should be replaced with XGBoost instead.

The SHAP values estimated for the third approach are plotted in Figure 4 and give us a better understanding of which features are most relevant to the model. We observe that the top five features are Tobacco, CenterID, HPVstatus, estimated glomerular filtration rate (eGFR) which represents the kidney function, and Weight (most important first). Tobacco consumption is the most important feature. This is not surprising as it is well-known that tobacco increases the risk of developing head and neck cancer, see for example [11]. The kidney function is ranked as number four and seems to be an important indicator for RFS. This finding is in line with earlier research, where a relationship between kidney function and prognoses for cancer patients has been identified [22]. An important limitation is that the true serum creatinine values were not available in the provided dataset. The creatinine concentration will to a large extent affect the estimated kidney function, and only applying the average gender values is not enough for getting accurate individual estimations. Despite this, we believe that serum creatinine should be included in future datasets to see if the model

performance can be further improved. Interestingly, CenterID is ranked as the second most important feature, which is also confirmed by a positive correlation between the CenterID and RFS in the correlation plot (Figure 5). These observations might be due to different patient populations at different centers, e.g., there might be more severely ill patients treated at one center while less serious cases are treated at another center. Another possibility is that the medical doctors choose different strategies for treating the patients or that the surgical skills differ between the centers. However, neither Surgery or Chemotherapy are among the highest-ranked features. The CenterID was not encoded as a categorical feature. If this had been done, the results from the SHAP analysis might change. The five least important features are Chemotherapy, Gender, dim z, surgery, and Age (least important first). The image features rank in the middle range regarding feature importance showing that they can be an important factor in predicting RFS. Considering that our imaging method and the image features are simple, we assume that with more advanced image analysis methods and the corresponding resulting features, the importance of the image-related features might increase even further.

4 Conclusion

In conclusion, our simple methods were not able to perform at the same levels as the highest ranked submissions despite achieving reasonable results. We got some interesting insights regarding combinations of different data modalities showing that the combination of different sources improves the results even when simple methods are used. For Task 2, we also had a closer look at feature importance, revealing some interesting features such as the usefulness of eGFR. For future work, it would be interesting to apply the feature importance analysis to other solutions of the competition to investigate if they are leading to similar findings. Furthermore, we would also like to investigate the CenterID correlation to explore if a hospital-specific treatment or country-related factor is influencing it.

References

1. Andrearczyk, V., Oreiller, V., Jreige, M., Vallières, M., Castelli, J., Elhalawani, H., Boughdad, S., Prior, J.O., Depeursinge, A.: Overview of the hecktor challenge at miccai 2022: automatic head and neck tumor segmentation in PET/CT. In: Head and Neck Tumor Segmentation and Outcome Prediction (2022)
2. Bandyk, M.G., Gopireddy, D.R., Lall, C., Balaji, K., Dolz, J.: Mri and ct bladder segmentation from classical to deep learning based approaches: Current limitations and lessons. *Computers in Biology and Medicine* **134**, 104472 (2021)
3. Biewald, L.: Experiment tracking with weights and biases (2020), <https://www.wandb.com/>, software available from wandb.com
4. Breiman, L.: Random forests. *Machine learning* **45**(1), 5–32 (2001)

5. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. pp. 108–122 (2013)
6. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. *Information* **11**(2) (2020). <https://doi.org/10.3390/info11020125>, <https://www.mdpi.com/2078-2489/11/2/125>
7. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). pp. 785–794 (2016). <https://doi.org/10.1145/2939672.2939785>
8. Cockcroft, D.W., Gault, H.: Prediction of creatinine clearance from serum creatinine. *Nephron* **16**, 31–41 (1976). <https://doi.org/10.1159/000180580>
9. Duran, A., Dussert, G., Rouvière, O., Jaouen, T., Jodoin, P.M., Lartizien, C.: Prostatention-net: A deep attention model for prostate cancer segmentation by aggressiveness in mri scans. *Medical Image Analysis* **77**, 102347 (2022)
10. Gerstner, A.: Early detection in head and neck cancer – current state and future perspectives. *GMS current topics in otorhinolaryngology, head and neck surgery* **7**, Doc06 (01 2008)
11. Hashibe, M., Brennan, P., Chuang, S.c., Boccia, S., Castellsague, X., Chen, C., Curado, M.P., Dal Maso, L., Daudt, A.W., Fabianova, E., Fernandez, L., Wünsch-Filho, V., Franceschi, S., Hayes, R.B., Herrero, R., Kelsey, K., Koifman, S., La Vecchia, C., Lazarus, P., Levi, F., Lence, J.J., Mates, D., Matos, E., Menezes, A., McClean, M.D., Muscat, J., Eluf-Neto, J., Olshan, A.F., Purdue, M., Rudnai, P., Schwartz, S.M., Smith, E., Sturgis, E.M., Szeszenia-Dabrowska, N., Talamini, R., Wei, Q., Winn, D.M., Shangina, O., Pilarska, A., Zhang, Z.F., Ferro, G., Berthiller, J., Boffetta, P.: Interaction between Tobacco and Alcohol Use and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiology, Biomarkers & Prevention* **18**(2), 541–550 (2009). <https://doi.org/10.1158/1055-9965.EPI-08-0347>
12. Iakubovskii, P.: Segmentation models pytorch. https://github.com/qubvel/segmentation_models.pytorch (2019)
13. Itseez: Open source computer vision library. <https://github.com/itseez/opencv> (2015)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
15. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I.: Explainable ai for trees: From local explanations to global understanding (2019). <https://doi.org/10.48550/arXiv.1905.04610>
16. Mittmann, B.J., Braun, M., Runck, F., Schmitz, B., Tran, T.N., Yamlahi, A., Maier-Hein, L., Franz, A.M.: Deep learning-based classification of dsa image sequences of patients with acute ischemic stroke. *International Journal of Computer Assisted Radiology and Surgery* pp. 1–9 (2022)
17. Oreiller, V., Andrearczyk, V., Jreige, M., Boughdad, S., Elhalawani, H., Castelli, J., Vallières, M., Zhu, S., Xie, J., Peng, Y., Iantsen, A., Hatt, M., Yuan, Y., Ma, J., Yang, X., Rao, C., Pai, S., Ghimire, K., Feng, X., Naser, M.A., Fuller, C.D., Yousefrizi, F., Rahmim, A., Chen, H., Wang, L., Prior, J.O., Depeursinge, A.: Head and neck tumor segmentation in PET/CT: The HECKTOR challenge. *Medical Image Analysis* **77**, 102336 (2022). <https://doi.org/10.1016/j.media.2021.102336>

18. Outeiral, R.R., Bos, P., van der Hulst, H.J., Al-Mamgani, A., Jasperse, B., Simões, R., van der Heide, U.A.: Strategies for tackling the class imbalance problem of oropharyngeal primary tumor segmentation on magnetic resonance images. *Physics and Imaging in Radiation Oncology* (2022)
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) *Advances in Neural Information Processing Systems* 32, pp. 8024–8035. Curran Associates, Inc. (2019), <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
20. Ren, J., Eriksen, J.G., Nijkamp, J., Korreman, S.S.: Comparing different ct, pet and mri multi-modality image combinations for deep learning-based head and neck tumor segmentation. *Acta Oncologica* **60**(11), 1399–1406 (2021)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
22. e Silva, V.T.d.C., Costalonga, E.C., Coelho, F.O., Caires, R.A., Burdmann, E.A.: Assessment of kidney function in patients with cancer. *Advances in chronic kidney disease* **25**(1), 49–56 (2018)
23. Staff, M.C.: Creatinine tests. <https://www.mayoclinic.org/tests-procedures/creatinine-test/about/pac-20384646> (2021)
24. Thambawita, V.L.B., Hicks, S., Halvorsen, P., Riegler, M.: Divergentnets: Medical image segmentation by network ensemble. In: *Proceedings of the International Workshop and Challenge on Computer Vision in Endoscopy (EndoCV)*. pp. 27–38 (2021)
25. Wahid, K.A., Ahmed, S., He, R., van Dijk, L.V., Teuwen, J., McDonald, B.A., Salama, V., Mohamed, A.S., Salzillo, T., Dede, C., et al.: Evaluation of deep learning-based multiparametric mri oropharyngeal primary tumor auto-segmentation and investigation of input channel effects: Results from a prospective imaging registry. *Clinical and translational radiation oncology* **32**, 6–14 (2022)
26. Yaniv, Z., Lowekamp, B.C., Johnson, H.J., Beare, R.: Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research. *Journal of digital imaging* **31**(3), 290–303 (2018)
27. Young, H.P.: Monotonic solutions of cooperative games. *International Journal of Game Theory* **14**, 65–72 (1985). <https://doi.org/10.1007/BF01769885>
28. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging* **39**(6), 1856–1867 (2019)