

University of Groningen

Self-supervised Multi-modality Image Feature Extraction for the Progression Free Survival Prediction in Head and Neck Cancer

Ma, Baoqiang; Guo, Jiapan; De Biase, Alessia; Sourlos, Nikos; Tang, Wei; van Ooijen, Peter; Both, Stefan; Sijtsema, Nanna Maria

Published in:

Head and Neck Tumor Segmentation and Outcome Prediction - 2nd Challenge, HECKTOR 2021, Held in Conjunction with MICCAI 2021, Proceedings

DOI:

[10.1007/978-3-030-98253-9_29](https://doi.org/10.1007/978-3-030-98253-9_29)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Ma, B., Guo, J., De Biase, A., Sourlos, N., Tang, W., van Ooijen, P., Both, S., & Sijtsema, N. M. (2022). Self-supervised Multi-modality Image Feature Extraction for the Progression Free Survival Prediction in Head and Neck Cancer. In V. Andrearczyk, V. Oreiller, M. Hatt, & A. Depeursinge (Eds.), *Head and Neck Tumor Segmentation and Outcome Prediction - 2nd Challenge, HECKTOR 2021, Held in Conjunction with MICCAI 2021, Proceedings* (pp. 308-317). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 13209 LNCS). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-98253-9_29

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.








Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Self-supervised Multi-modality Image Feature Extraction for the Progression Free Survival Prediction in Head and Neck Cancer

Baoqiang Ma^(✉) , Jiapan Guo^(✉) , Alessia De Biase^(✉) ,
Nikos Sourlos^(✉) , Wei Tang , Peter van Ooijen , Stefan Both,
and Nanna Maria Sijtsema 

University Medical Center Groningen (UMCG),
Groningen 9700, RB, Netherlands

{b.ma, j.guo, a.de.biase, n.sourlos, w.tang, p.m.a.van.ooijen,
s.both, n.m.sijtsema}@umcg.nl

Abstract. Long-term survival of oropharyngeal squamous cell carcinoma patients (OPSCC) is quite poor. Accurate prediction of Progression Free Survival (PFS) before treatment could make identification of high-risk patients before treatment feasible which makes it possible to intensify or de-intensify treatments for high- or low-risk patients. In this work, we proposed a deep learning based pipeline for PFS prediction. The proposed pipeline consists of three parts. Firstly, we utilize the pyramid autoencoder for image feature extraction from both CT and PET scans. Secondly, the feed forward feature selection method is used to remove the redundant features from the extracted image features as well as clinical features. Finally, we feed all selected features to a DeepSurv model for survival analysis that outputs the risk score on PFS on individual patients. The whole pipeline was trained on 224 OPSCC patients. We have achieved a average C-index of 0.7806 and 0.7967 on the independent validation set for task 2 and task 3. The C-indices achieved on the test set are 0.6445 and 0.6373, respectively. It is demonstrated that our proposed approach has the potential for PFS prediction and possibly for other survival endpoints.

Keywords: Progression free survival prediction · OPSCC · DeepSurv · Autoencoder · PET-scans and CT-scans

1 Introduction

Almost 60,000 US patients are diagnosed with head and neck (H&N) cancer every year, causing 13,000 deaths annually [1]. The treatment strategies of H&N cancer

Aicrowd Group Name: “umcg”

© Springer Nature Switzerland AG 2022

V. Andrearczyk et al. (Eds.): HECKTOR 2021, LNCS 13209, pp. 308–317, 2022.

https://doi.org/10.1007/978-3-030-98253-9_29

such as Oropharyngeal squamous cell carcinoma (OPSCC) are usually nonsurgical such as chemotherapy, radiotherapy, and combinations of these. Although loco-regional control of most OPSCC is good, five-year OS for OPSCC have ranged from 46% to 85% including all stages, and 40–85% in advanced stage cohorts [2]. It would be beneficial to be able to identify patients with an expected worse treatment response before start of treatment. When prediction models for tumor related endpoints and complications would be available it would become possible to select the most optimal treatment method (with the optimal balance between predicted tumor control and complications) for individual patients. E.g. a more intensive treatment regimen could be considered for patients with a predicted high-risk for tumor recurrence, whereas a de-intensified treatment regimen could be an option for patients with a low risk for tumor recurrence, to limit the risk of complications like swallowing problems and xerostomia [3]. Therefore, we have developed a PFS prediction mode using clinical data and image data.

Radiomics [4] - quantitative imaging features from high throughput extraction - has been successfully applied to outcome prediction of H&N cancers [5–8]. However, its clinical application is restricted due to its dependence on manual segmentation and handcrafted features [9]. Deep learning-based methods includes algorithms and techniques that identify more complex patterns than radiomics in large image data sets without handcrafted feature extraction, and they have been employed in various medical image fields [10–12] as well as H&N cancer outcome prediction [13–16]. In our method, we select Autoencoders as the basic architecture for image feature extraction.

Features significantly relating to PFS prediction can be obtained through features selection process. The obtained features can be used to create a survival analysis model, such as Cox proportional hazard model (CPHM) [17], random survival forests (RFS) [18] and DeepSurv [19] (a Cox proportional hazards deep neural network). We chose DeepSurv as our PFS prediction model because it can successfully model increasingly complex relationships between a patient’s covariates and their risk of failure.

Our aim is building a DeepSurv model with the capability of predicting PFS prior to treatment using available clinical data and image features of CT and PET extracted by the Autoencoder. The work described in this paper was used to participate in the task 2 and task 3 of HECKOR 2021 challenge [20,21].

2 Materials

2.1 Dataset Description

The training set includes 224 head and neck cancer patients from 5 hospitals (CHGJ, CHUS, CHMR, CHUM and CHUP). There are co-registered 3D CT and FDG-PET images and GTVt images (primary Gross Tumor Volume label) for each patient. The voxels sizes are nearly 1.0 mm in the x and y directions and vary between 1.5 to 3.0 mm along the z direction. A bounding box is provided for each patient around the oropharyngeal primary tumors. Clinical data of all patients can be found in a csv file. The testing set consists of 101 patients treated

in 2 hospitals (CHUP and CHUV). Co-registered CT- and PET-scans, bounding box and clinical data are available for patients in the test set, but not the GTV contour.

2.2 Data Preparation

We selected Gender, Age, T-stage, N-stage, TNM group, HPV status and Chemotherapy as the potential predictive data. Age is normalized by dividing by 100, and other clinical data are used as categorical variables. Through KM-survival analysis, some categories values with similar survival curves were combined. A detailed description of the definition of the categorical values is summarized in Table 1.

Table 1. The summary of classification of values in each category variables of clinical data.

Category variable	Value classification
Gender	(Male = 0), (Female = 1)
T-stage	(T1 T2 T3 = 0), (T4 T4a T4b = 1)
N-stage	(N0 N1 = 0), (N2 N2a N2b N2c = 1), (N3 = 2)
TNM group	(I = 0), (II III IV = 1), (IVA IVB IVC = 2)
HPV status	(negative = 0), (positive = 1), (unknown: 2)
Chemotherapy	(not = 0), (yes = 1)

The bounding box region image of CT, PET and GTVt (the mask image of gross tumor volume of the primary tumor) of each patient in training set and testing set are first cropped and extracted. Then, these CT and PET 3D images were resampled to $1 \times 1 \times 1 \text{ mm}^3$ pixel spacing with trilinear interpolation. The GTVt masks were resampled to the same resolution with $1 \times 1 \times 1 \text{ mm}^3$ CT but using nearest interpolation. CT region image pixel values are truncated to $[-200, 200]$ and then normalized to $[0, 1]$ by the max-min value method. The pixel values of the PET region image smaller than 0 are set to 0. PET region images are normalized by first z-score and then the max-min value method. Finally, the normalized CT and PET region images of each patient are summed up to form a new combined image named CT/PET image. The pre-processed CT, PET and CT/PET with the size of $1 \times 144 \times 144 \times 144$ are the input of the autoencoder in task 2 (not using GTVt). To get the input of autoencoder in task 3 (using GTVt), we first dilated GTVt with size of 5 voxels, and use the dilated GTVt to multiply with CT and PET, then extracting the GTVt-region CT, PET and CT/PET images by two methods. The first method is GTVt center cropping (to a size of $64 \times 64 \times 64$). The second method is first cropping a sub-cube of GTVt according to the border positions in three directions of the tumor, then resampling the sub-cube to a size of $64 \times 64 \times 64$ voxels. Method 1 gives GTVt images with the same pixel spacing across cases whereas method 2 results in GTVt images with varying pixel spacing across cases depending on the tumor

size. By combining both GTVt images information about the tumor size as well as tumor images with an optimal resolution, tumor information are effectively used in the training process. Images from the two methods are concatenated together to a size of $2 \times 64 \times 64 \times 64$, and they are the input of the task 3 autoencoder.

We adopt two different strategies to divide the provided dataset (224 patients) into a training and validation set. The first one is to use leave-one-center-out, in which 4 centers are used as the training set while one as the validation set. The second one is that we randomly selected 179 patients as the training while 45 as the validation set. Thus, we could perform 6-fold cross validation and ensemble results of different models on the final test set (101 patients).

3 Methods

The success of deep learning methods in computer vision tasks has brought its wide applications to the medical image analysis. They, however, require a large amount of labeled samples. The model performances are also biased by the manually provided labels. In this work, we proposed a deep learning based pipeline that adopts unsupervised learning approach for image feature extraction in the prediction of progression free survival (PFS) for head and neck cancer. We utilized a self-supervised deep learning approach for the extraction of tumor characteristics from both CT and PET scans. The extracted image features and clinical parameters were then used to train a DeepSurv model for time-to-event prediction on the PFS. Figure 1 illustrates the proposed pipeline.

For each image set of CT, PET, CT/PET, CT-GTVt, PET-GTVt and CT/PET-GTVt, we trained 6 models using 6-fold-cross-validation. The train/validation set ratio of each fold is different, because we performed leave one center out cross validation. For each fold, feature selection is performed in all image features that were identified by the autoencoders. The selected image features (around 2–6 features) and selected clinical data (Age, T-stage and HPV status) are used to train a DeepSurv model. We trained 30 DeepSurv models using the training set in each fold, and finally selected 3 models with the highest validation set C-index. In total 18 DeepSurv models (each fold has 3 DeepSurv models) are obtained, and their predicted risk scores on the test set are averaged to obtain the final result in the test set.

3.1 Autoencoder

Autoencoders are used to extract high-level features through reconstructing the input. In our method, we used CT, PET and CT/PET images to train three autoencoders, separately. An autoencoder consists of an encoder and a decoder. The encoder compresses the input image to high-level features, then the decoder reconstructs input image from these features. Those high-level features from the last layer of the encoder are chosen for the feature selection process.

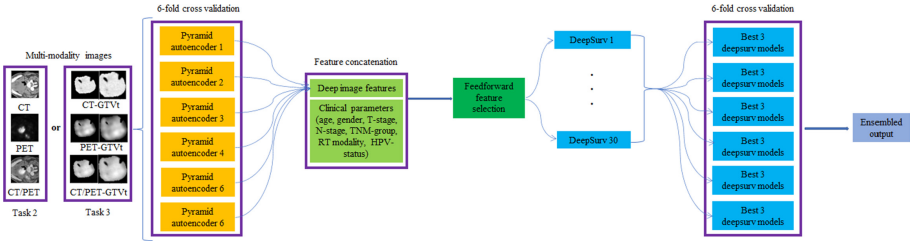


Fig. 1. The whole pipeline of the proposed method. 6-fold cross validation is applied in training autoencoder and DeepSurv models. Each modality image (CT, PET and CT/PET of task 2 or CT-GTVt, PET-GTVt and CT/PET-GTVt of task 3) is used to train one autoencoder to extract image features. All extracted image features and clinical data are selected using feedforward selection. The selected features are applied to train 30 DeepSurv models for each fold. Finally, 18 models (3 models for each fold) with highest validation C-index are used for testing, and their output on the test set are ensembled.

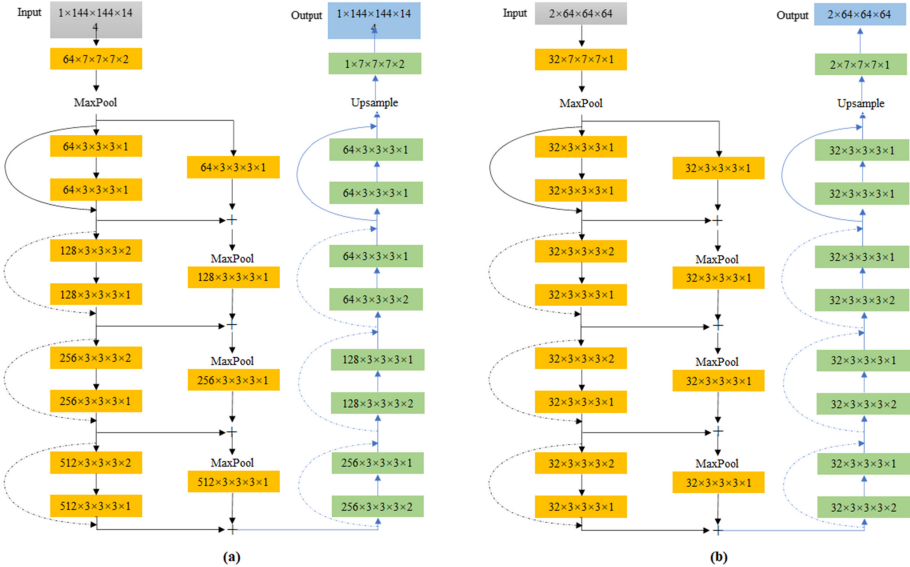


Fig. 2. The 3D ResNet-like architecture of autoencoders in task 2 (a) and task 3 (b). In task 2, the input is CT, PET or CT/PET patches in boxing regions with size $144 \times 144 \times 144$. In task 3, the input is CT, PET or CT/PET patches in primary tumor region only with size $2 \times 64 \times 64 \times 64$. Identity (solid arrows) and projection (dashed arrows) shortcuts are shown in residual blocks. Yellow and green rectangle stand for convolution and deconvolution, separately (Color figure online)

The architecture of autoencoders in task 2 and task 3 are displayed in Fig. 2. Our autoencoder is built upon on 3D ResNet [22] with the use of a pyramid architecture between convolution blocks. We used $N \times H \times W \times D \times S$ to describe the

convolution or transpose convolution kernel number (N), size in three directions (H, W, D) and the stride (S) in Fig. 2. The encoder consists of one convolution layer, a maxpooling layer and 4 convolution residual blocks and a pyramid architecture to combine different level images features. The stride of all maxpooling layers is 2. At the end of the encoder, 512 high-level feature maps with size of $5 \times 5 \times 5$ (task 2) or 32 high-level feature maps with size of $4 \times 4 \times 4$ (task 3) are obtained. Four continuous transposed convolution residual blocks, an upsampling layer with sampling factor 2 and a final transposed convolution layer form the decoder. A Relu function and Batch Normalization layer follow all convolution and transposed convolution layers of the autoencoder.

A combined loss function including L1 loss, mean square error (MSE) and Structural Similarity (SSIM) is employed in the autoencoder training process. The L1 loss for one training example can be written as:

$$L_{L1} = \|A_{(x)} - x\|_1 \quad (1)$$

the MSE is defined as:

$$L_{MSE} = \|A_{(x)} - x\|_2 \quad (2)$$

For SSIM the loss function L_{SSIM} as described in [23] was used: SSIM is designed by modeling any image distortion as a combination of three factors that are loss of correlation, luminance distortion and contrast distortion. The combined loss function is:

$$L_{combined} = L_{L1} + L_{MSE} + 0.5 * L_{SSIM} \quad (3)$$

3.2 Feature Selection

We first selected clinical data which are known to be related to PFS prediction. Gender, Age, T-stage, N-stage, TNM group, HPV status and Chemotherapy are kept for selection. We used the SequentialFeatureSelector of scikit-learn as feature selector (set direction as forward). The estimator of the selector is set as CPHM model. We ran the feature selector 1000 times using a random subset of the training set every time. Finally, the most frequently selected 3 features (Age, T-stage, HPV status) are reserved to perform PFS prediction.

We use the same feature selection method for image features selection. First each 3D feature map extracted from autoencoders is changed to single value feature by maxpooling. Then these 512 CT, 512 PET, 512 CT/PET, 32 CT-gtv, 32 PET-gtv and 32 CT/PET-gtv features are input to feature selector. All those image features are ranked according to their selected frequency. The 2–6 features with highest rankings are chosen.

3.3 DeepSurv

DeepSurv [19] is a deep learning based survival analysis model. We do not elaborate on it here and refer the interested readers to [19]. We set the DeepSurv

architecture as two fully connected layers with 50 nodes, Relu and Batch Normalization. The DeepSurv outputs the risk score of PFS. The loss function is the average negative log partial likelihood.

4 Experiment

4.1 Training Details

The Autoencoders were trained using the Adam optimizer with the initial learning rate 0.001 in Tesla V100 GPU. The total number of training epochs is set to 80. The learning rate will decrease by multiplying by 0.1 if the training loss doesn't reduce in 10 consecutive epochs. Flipping and random rotation are used for data augmentation.

The official DeepSurv (<https://github.com/jaredleekatzman/DeepSurv>) code setting is applied to train our DeepSurv models. The total training steps are 5000, the validation set is used to select the best C-index model.

4.2 Results

This section shows the reconstructed images of the autoencoder and the C-index on the training set, validation set and test set.

The input images and the reconstructed images by autoencoders from one patient in the test set are displayed in Fig. 3. The reconstructed PET image is very similar to the input one. And we can recognize the highlighted tumor region in the reconstructed CT/PET image. Autoencoders successfully restored the shape of the tumor from the high-level features when we compare the input and output of CT-GTVt, PET-GTVt and CT/PET-GTVt. These results show that the high-level features extracted from autoencoders are representative and relevant.

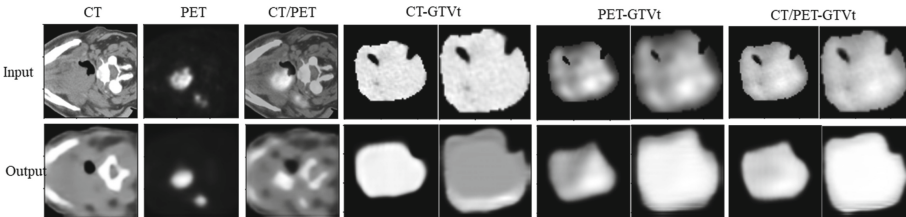


Fig. 3. The input and output of autoencoders. Tumor region images are successfully reconstructed in the output PET and CT/PET images. Tumor shape information are recovered on the output of CT-GTVt, PET-GTVt and CT/PET-GTVt images.

We summarized the training set and validation set C-index values of the DeepSurv model with highest validation C-index of each fold in Table 2. In task 2,

we used selected clinical features and images features of CT, PET and CT/PET to train the DeepSurv models. In task 3, in addition to the features used in task 2, we added image features from CT-GTVt, PET-GTVt and CT/PET-GTVt. However, compared with C-index values on the validation set in fold 1, 4 and 6 of task2, we did not obtain a higher C-index value in task 3 when adding image features from CT-GTVt, PET-GTVt and CT/PET-GTVt. Therefore, we used the image features from task 2 also in task 3 in these three folds in where task 2 and task3 had the same C-indexes results. In fold 2, 3 and 5, task 3 obtained higher C-index values on both training and validation sets than task 2 because of adding features from CT-GTVt, PET-GTVt and CT/PET-GTVt.

From the Table 3, we can see that our method achieved good C-index values in the independent test set (0.6445 in task2 and 0.6373 in task 3). Although the C-index values of the validation set in task 3 were little higher than that in task 2, the C-index values of the test set were a little lower in task 3. This is possible due to the noise when training. The experimental results showed that our method does not need GTVt to locate the tumor. Our autoencoder can automatically extract image features in the tumor region, which is demonstrated by the reconstructed PET and CT/PET images in Fig. 3, in which the highlighted tumor image was constructed successfully from high-level features extracted by the encoder.

Table 2. The C-index values of training set and validation set of task 2 and task 3, using the best DeepSurv model with highest validation C-index.

Task name	Set name	fold1	fold2	fold3	fold4	fold5	fold6
Task2	Training	0.5939	0.6418	0.4389	0.7233	0.5723	0.7009
Task2	Validation	0.8073	0.7052	0.7644	0.8360	0.7383	0.8324
Task3	Training	0.5939	0.8025	0.4852	0.7233	0.5723	0.7009
Task3	Validation	0.8073	0.7063	0.8506	0.8360	0.7477	0.8324

Table 3. The C-index values on test set of task 2 and task 3.

Task name	C-index
Task2	0.6445
Task3	0.6373

5 Discussion and Conclusion

We have shown that our method was able to predict PFS with relative high average C-indexes of 0.7806 and 0.7967 for task 2 and 3 respectively in the

validation set of all folds. However, the C-index of the test set is much lower than that of the validation sets (>0.7), which shows that our DeepSurv models overfit on the validation set. For example in fold 3, the C-index on the training set were very low (0.4389 in task2 and 0.4852 in task 3) but much higher (0.7644 in task2 and 0.8506 in task 3) on the validation set. The reason may be that the validation set (only 18 patients) has a very different feature distribution from the training set (206 patients). In our experiment, we selected the DeepSurv models with highest C-index values in the validation set for testing purpose, but they might perform worse in both the training and testing set. Thus, these models performing worse in the training set will decrease the final test set C-index value.

In order to improve the result on the test set in the future, we plan to change the methods in three aspects. Firstly, we will only select a part of DeepSurv models have good C-index in both the training and validation set for using on the test set, such as only using models in fold 2,4 and 6. Secondly, splitting the training set and validation set in a another way to make them have similar feature distribution. Finally when we retrain DeepSurv models of each fold in the future, we should save the model with high validation C-index on the condition of a high training C-index value instead of selecting models with only highest validation C-index.

We proposed a method that used a 3D pyramid autoencoder to extract high-level image features for PFS prediction. Obtained images features and clinical data are selected to acquire PFS-prediction related features. These selected features are applied to train a DeepSurv model for PFS prediction. Experimental results demonstrated that whether using GTVt or not, we could obtain a good C-index value on the test set. The proposed method has the potential for PFS prediction and possibly for other survival endpoints.

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2016. *CA Cancer J. Clin.* **66**, 7–30 (2016)
2. Clark, J.M., et al.: Long-term survival and swallowing outcomes in advanced stage oropharyngeal squamous cell carcinomas. *Papillomavirus Res.* **7**, 1–10 (2019)
3. Tolentino, E.S., et al.: Oral adverse effects of head and neck radiotherapy: literature review and suggestion of a clinical oral care guideline for irradiated patients. *J. Appl. Oral Sci. Revista FOB* **19**, 448–54 (2011)
4. Kumar, V., et al.: Radiomics: the process and the challenges. *Magn. Reson. Imaging* **30**(9), 1234–1248 (2012)
5. Cheng, N.M., Fang, Y.D., Tsan, D.L., Lee, L.Y., Chang, J.T., Wang, H.M., et al.: Heterogeneity and irregularity of pretreatment (18)F-fluorodeoxyglucose positron emission tomography improved prognostic stratification of p16-negative high-risk squamous cell carcinoma of the oropharynx. *Oral Oncol.* **78**, 156–62 (2018)
6. Haider SP., Zeevi T., Baumeister P.: Potential added value of PET/CT radiomics for survival prognostication beyond AJCC 8th edition staging in oropharyngeal squamous cell carcinoma. *Cancers (Basel)* **12**(7) (2020). <https://doi.org/10.3390/cancers12071778>

7. Leijenaar, R.T., Carvalho, S., Hoebbers, F.J., Aerts, H.J., van Elmpt, W.J., Huang, S.H., et al.: External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta Oncol.* **54**(9), 1423–9 (2015). <https://doi.org/10.3109/0284186x.2015.1061214>
8. Wu, J., et al.: Tumor subregion evolution based imaging features to assess early response and predict prognosis in oropharyngeal cancer. *J. Nucl. Med.* **61**(3), 327–36 (2020). <https://doi.org/10.2967/jnumed.119.230037>
9. Bi, W.L., et al.: Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J. Clin.* **69**(2), 127–57 (2019). <https://doi.org/10.3322/caac.21552>
10. Ma, B., Zhao, Y., Yang, Y., et al.: MRI image synthesis with dual discriminator adversarial learning and difficulty-aware attention mechanism for hippocampal subfields segmentation. *Comput. Med. Imaging Graph.* **86**, 101800 (2020)
11. Zhao, Y., Ma, B., Jiang, P., Zeng, D., Wang, X., Li, S.: Prediction of Alzheimer’s disease progression with multi-information generative adversarial network. *IEEE J. Biomed. Health Inform.* **25**(3), 711–719 (2020)
12. Zeng, D., Li, Q., Ma, B., Li, S.: Hippocampus segmentation for preterm and aging brains using 3D densely connected fully convolutional networks. *IEEE Access* **8**, 97032–97044 (2020)
13. Diamant, A., Chatterjee, A., Vallières, M., Shenouda, G., Seuntjens, J.: Deep learning in head & neck cancer outcome prediction. *Sci. Rep.* **9**(1), 1–10 (2019)
14. Kann, B.H., et al.: Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. *Sci. Rep.* **8**(1), 1–11 (2018)
15. Fujima, N., et al.: Prediction of the local treatment outcome in patients with oropharyngeal squamous cell carcinoma using deep learning analysis of pretreatment FDG-PET images. *BMC Cancer* **21**(1), 1–13 (2021)
16. Cheng, N.M., et al.: Deep learning for fully automated prediction of overall survival in patients with oropharyngeal cancer using FDG-PET imaging. *Clin. Cancer Res.* **27**, 3948–3959 (2021)
17. Cox, D.R.: Regression models and life-tables. In: Kotz, S., Johnson, N.L. (eds.) *Breakthroughs in Statistics*. SSS, pp. 527–541. Springer, New York (1992). https://doi.org/10.1007/978-1-4612-4380-9_37
18. Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S.: Random survival forests. *Ann. Appl. Stat.* **2**(3), 841–860 (2008)
19. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **18**(1), 1–12 (2018)
20. Andrearczyk, V., et al.: Overview of the HECKTOR challenge at MICCAI 2021: automatic head and neck tumor segmentation and outcome prediction in PET/CT images. In: Andrearczyk, V., Oreiller, V., Hatt, M., Depeursinge, A. (eds.) *HECKTOR 2021*. LNCS, vol. 13209, pp. 1–37. Springer, Cham (2022)
21. Oreiller, V., et al.: Head and neck tumor segmentation in PET/CT: the HECKTOR challenge. *Med. Image Anal.* **77**, 102336 (2022)
22. Hara, K., Kataoka, H., Satoh, Y.: Learning spatio-temporal features with 3D residual networks for action recognition. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 3154–3160 (2017)
23. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)