Multi-stage Multimodal Progressive Learning for Coordinated Segmentation, Diagnosis, and Prognosis in Head and Neck Cancer

Yue Lin[2], Shenghai Wu[2], Hao Wang[1 [0000−0002−6235−8425]], Lei Bi[1 [0000−0001−9759−0200]], and Mingyuan Meng[1 [0000−0002−9562−1613]]✉

[1] School of Computer Science, The University of Sydney, Sydney, Australia
mmen2292@uni.sydney.edu.au
[2] Research Institute, Newland Digital Technology, Fuzhou, China

Abstract. Head and Neck (H&N) cancer is among the most common cancers worldwide, and its related clinical decision-making constitutes a systematic process that requires the integration of multimodal clinical data and the coordination of diverse tasks in the clinical workflow. However, how to effectively coordinate the interrelated clinical tasks to maximize their synergistic potential is still an open question. In this study, we propose a Multi-stage Multimodal Progressive Learning (named MMPL) framework for coordinated modeling of segmentation, diagnosis, and prognosis tasks, in the context of *HECKTOR 2025* challenge at *MICCAI 2025*. Our MMPL progressively learns three clinical tasks that collectively facilitate personalized treatment planning: (i) tumor segmentation, (ii) HPV status classification, and (iii) survival prediction. Specifically, we establish a unified network backbone, consisting of a triple-stream encoder with adaptive PET/CT information fusion and an attention-gated decoder that can be applied to all three tasks. This backbone is successively trained for segmentation, classification, and survival prediction at three learning stages, where the knowledge is progressively learned with the guidance of prior knowledge accumulating from former stages. Further, the intermediate outputs (e.g., segmentation masks, HPV status) are leveraged as guidance on radiomics analysis or as supplementary indicators for the final prediction. Our team (*InterStellar*) attained top-tier performance across all three tasks in the validation phase, while the final testing results have yet to be released.

Keywords: Multi-stage Progressive Learning · Multimodal Learning · Segmentation · Diagnosis · Prognosis

1 Introduction

With over 900,000 cases diagnosed annually worldwide, Head and Neck (H&N) cancer poses a significant and persistent challenge to global public health [1]. Clinical decision-making for H&N cancer patients often integrates heterogeneous multimodal clinical data, such as medical imaging (e.g., MRI, PET/CT) and clinical indicators (e.g., demographics, TNM stage), for comprehensive assessment of

disease severity and progression risk [2,3,4,5]. This assessment process typically involves the coordination of three interrelated tasks. First, the precise segmentation of the primary tumor (GTVp) and metastatic lymph nodes (GTVn) establishes the lesion localization and spatial context. Second, the diagnostic determination of crucial clinical characteristics (e.g., HPV status) identifies key etiologic and biological factors that carry significant prognostic implications and inform the interpretation of cancer stage and treatment intensity. Third, the prognostic assessment of the patients' outcomes synthesizes these upstream findings to estimate disease progression risk, which directly facilitates personalized treatment planning in accordance with clinical guidelines [2]. The above clinical workflow reflects a systematic process that requires the integration of multimodal data and the coordination of three tasks: *segmentation* → *diagnosis* → *prognosis*.

Most related methods are optimized for individual tasks, leaving cross-task synergies underexploited. For example, segmentation-target methods focus on automated delineation of H&N tumors in PET/CT [6,7,8,9], while imaging-based diagnosis/prognosis are typically treated as standalone problems that are developed independently using radiomics or deep-learning models [10,11,12,13,14,15]. Some recent works have explored the joint modeling of segmentation and survival prediction via joint multi-task learning [16,17,18], indicating that transferring knowledge from tumor segmentation can significantly benefit survival prediction [19,20]. However, diagnostic phenotyping is still kept outside the loop, and systems aligned with the clinical workflow of *segmentation* → *diagnosis* → *prognosis* remain scarce, underscoring the need for a unified framework that can effectively coordinate these tasks to leverage their cross-task synergies.

Since MICCAI 2020, the HECKTOR challenge has provided a multi-center, multimodal PET/CT benchmark and has completed three editions by 2022 with broad community engagement [21,22]. In 2025, building on prior focus on segmentation and survival outcome prediction, the HECKTOR 2025 challenge extends to include HPV status classification as a third task, aligning with the clinical workflow of segmentation, diagnosis, and prognosis [23]. Specifically, the 2025 edition (HECKTOR 2025) comprises three tasks: Task 1 — automatic detection and segmentation of H&N primary tumor and lymph nodes in FDG-PET/CT images; Task 2 — prediction of recurrence-free survival (RFS) from FDG-PET/CT images together with available clinical information and radiotherapy planning dose maps; and Task 3 — diagnosis of HPV status from FDG-PET/CT images together with available clinical information [23].

In the context of HECKTOR 2025, we introduce a Multi-stage Multimodal Progressive Learning (MMPL) framework that mirrors the clinical workflow to leverage cross-task synergies among segmentation, diagnosis, and prognosis. Specifically, MMPL adopts a unified network backbone shared across the three tasks, comprising a triple-stream encoder for PET/CT with adaptive cross-modal fusion and an attention-gated decoder together with task-specific prediction heads for (i) primary tumor and lymph node segmentation, (ii) HPV status classification, and (iii) RFS prediction. Learning proceeds in three successive stages that follow the clinical workflow, beginning with segmentation to learn

robust representations of anatomy and disease extent, continuing with diagnosis that uses tumor-anchored features together with PET/CT information and relevant clinical indicators (e.g., demographics, TNM stage, performance status) to classify HPV status, and eventually concluding with RFS prediction that builds on all the prior knowledge from the earlier stages. Intermediate outputs are explicitly reused downstream: segmentation provides lesion localization and enable radiomics extraction from PET/CT within GTVp/GTVn, while HPV predictions serve as complementary predictive indicator for survival estimation. This progressive, clinically aligned learning paradigm allows knowledge to flow from *segmentation* to *diagnosis* and then to *prognosis*, thereby leveraging both task dependencies and PET/CT complementarity.

2 Methods

2.1 Overview

We propose a Multi-stage Multimodal Progressive Learning (MMPL) framework that is aligned with the clinical workflow of *segmentation* → *diagnosis* → *prognosis*. MMPL proceeds through three successive learning stages: (S1) primary tumor and lymph node segmentation, (S2) HPV status classification, and (S3) RFS prediction, using a unified network backbone shared across all stages. The backbone follows the encoder–decoder design validated in the prior work [20]: It uses a *triple-stream encoder* with separate PET and CT streams and an adaptive fusion stream that mixes information across pyramid levels, followed by an *attention-gated decoder* that aggregates multi-scale features via gated skip connections, and produces a task-agnostic representation reused across all three stages. Three *task-specific heads* are individually attached to this shared representation for task-specific prediction, as illustrated in Fig. 1.

Our MMPL is built on a *prior-guided progressive learning strategy* that coordinates stages at two complementary levels while remaining faithful to clinical practice. At the parameter level, the shared backbone undergoes knowledge transfer and refinement: weights learned during segmentation are propagated and further optimized for diagnosis and then for prognosis, allowing prior knowledge to accumulate and be reused across tasks. At the output level, intermediate results serve as explicit guidance: The GTVp/GTVn masks (predicted in S1) provide lesion localization for mask-guided radiomics extraction reused by S2 and S3, and the HPV probability (predicted in S2) is incorporated as a supplementary indicator for S3. Clinical indicators are incorporated in the stage-specific statistical models, being concatenated with image-derived features for HPV status classification and used as covariates in the survival model. This progressive, clinically aligned learning paradigm enables knowledge to flow from tumor/nodal segmentation to diagnosis and then to prognosis, leveraging cross-task synergies while preserving interpretability.

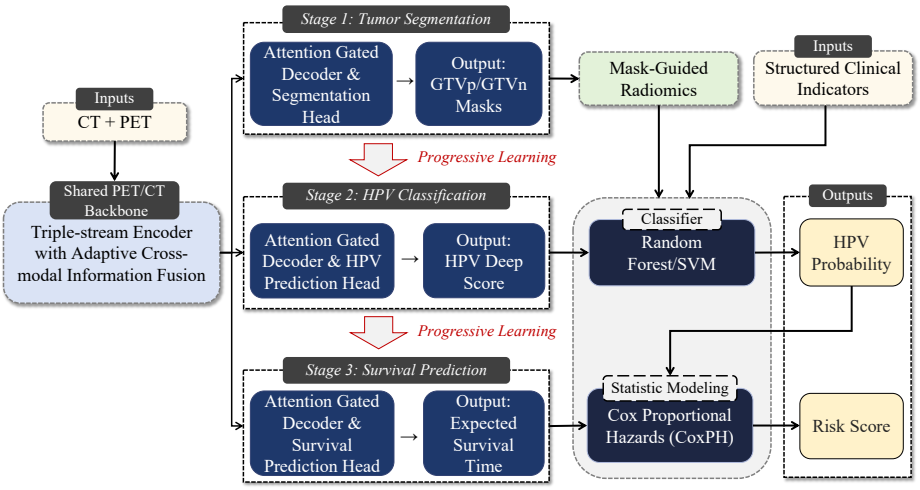

Fig. 1. Overview of our MMPL framework.

2.2 Data preprocessing

Image preprocessing. All PET/CT images undergo a consistent preprocessing procedure for both training and inference to prevent distribution shift. First, the co-registered PET and CT scans are resampled to an isotropic 1.0 mm spacing using B-spline interpolation over a common grid defined by their physical intersection. Subsequently, a Region of Interest (ROI) is automatically localized via a PET-guided strategy. This involves selecting the cranial-most 25% of the image volume along the z-axis, within which PET intensities are z-score normalized. A binary mask is then created by thresholding voxels with z-score greater than 1.0. The centroid of the largest 3D connected component in this mask is designated as the ROI center, where a $200 \times 200 \times 310$ voxel patch is extracted from both PET and CT. The CT intensities are clipped to the $[-1024, 1024]$ HU range and then scaled to $[-1, 1]$, while PET intensities are standardized using a z-score normalization across all non-zero voxels. For model training, random 128^3 sub-patches are cropped from the ROI. The corresponding segmentation labels are resampled using nearest-neighbor interpolation and processed into two separate binary masks for GTVp and GTVn.

Clinical data preprocessing. The provided structured clinical indicators are processed with the following procedure: categorical variables (e.g., sex, smoking status, alcohol status, performance status) are one-hot encoded; continuous variables are standardized via z-score normalization, using statistics fitted on the training split only to prevent information leakage. Missing values are imputed by the median of the training data for continuous variables. All preprocessing settings fitted on the training data are frozen and reused for inference.

2.3 Stage 1: Tumor segmentation

The segmentation head is a 3D convolutional layer attached to the shared backbone. As described in [20], the attention-gated decoder performs four successive upsampling stages from the coarsest features, and the resulting features from the last decoder stage are projected by the segmentation head into two channels to produce voxel-wise probability maps for GTVp and GTVn.

Overlap–Detection–Aware (ODA) loss. To reflect clinical priorities of precise GTVp boundaries and high sensitivity to small GTVn lesions, we use a composite objective coupling Dice with a focal Tversky term [24,25]. For $c \in \{T, N\}$ with binary labels y_c and predictions $\hat{y}_c$,

$$\mathcal{L}_{\text{seg}} = \sum_{c\in\{T,N\}} \Big[\mathcal{L}_{\text{Dice}}(y_c, \hat{y}_c) + \big(1 - \text{TI}(y_c, \hat{y}_c; \alpha_c, \beta_c)\big)^{\gamma_c}\Big] \; + \; \mathcal{L}_{\text{stab}}, \tag{1}$$

where $\mathcal{L}_{\text{Dice}}(y, \hat{y}) = 1 - \dfrac{2\langle y, \hat{y}\rangle + \varepsilon}{\|y\|_1 + \|\hat{y}\|_1 + \varepsilon}$ and

$$\text{TI}(y, \hat{y}; \alpha, \beta) = \frac{\text{TP} + \varepsilon}{\text{TP} + \alpha\,\text{FP} + \beta\,\text{FN} + \varepsilon}, \tag{2}$$

with $\text{TP} = \langle y, \hat{y}\rangle$, $\text{FP} = \langle 1 - y, \hat{y}\rangle$, and $\text{FN} = \langle y, 1 - \hat{y}\rangle$. We use near-symmetric parameters for GTVp $(\alpha_T, \beta_T, \gamma_T) = (0.5, 0.5, 1.33)$ and recall-biased parameters for GTVn $(\alpha_N, \beta_N, \gamma_N) = (0.3, 0.7, 1.50)$. A stabilization term improves optimization on empty or low-signal crops:

$$\mathcal{L}_{\text{stab}} = \sum_{c\in\{T,N\}} \Big(\mathcal{L}_{\text{BCE}}(y_c, \hat{y}_c) + \mathcal{L}_{\text{Dice}}(1 - y_c,\, 1 - \hat{y}_c)\Big) \tag{3}$$

Inference procedure. We compute the PET-guided 1.0 mm ROI (as described in Section 2.2) and run overlapping sliding-window inference; window-wise probabilities are averaged across overlaps. Class-specific thresholds are applied, labels are composed with tumor-first priority, and the result is mapped back to the native CT grid. Post-processing consists of 3D hole filling and removal of connected components smaller than 50 voxels per class.

Mask-guided radiomics. Using the predicted masks of GTVp and GTVn, we extract radiomics features from both PET and CT to provide structured, interpretable descriptors to later stages. In particular, features cover morphology/shape, first-order statistics, and texture families (e.g., GLCM, GLRLM, GLSZM) computed on the original images as well as on standard filtered images (e.g., LoG, wavelet sub-bands), with fixed-bin discretization and filter settings as specified in [17]. Subsequently, features are standardized by z-score normalization using statistics fitted on the training data to prevent information leakage.

2.4 Stage 2: HPV status classification

The HPV prediction head is a shallow MLP head attached to the attention-gated decoder at four scales (1/8, 1/4, 1/2, full resolution). At each scale, the output of the last convolutional layer is fed into a global average pooling layer to form a scale-wise vector. The four vectors are concatenated and fed into the HPV prediction head, yielding a scalar logit whose sigmoid is supervised with a class-balanced binary cross-entropy. The logit is used as the deep score for downstream integration with clinical indicators and radiomics features.

Supervised classifier. In line with clinical diagnosis practice, we aggregate three evidence sources: (i) the deep score from the HPV prediction head; (ii) the pre-extracted, *mask-guided* PET/CT radiomics from Stage 1, and (iii) structured clinical indicators as encoded and standardized as in Section 2.2. The concatenated vector is fed into a supervised ensemble classifier (random forest [26], scikit-learn) to output the HPV probability p_{HPV}. To enhance robustness, class imbalance is handled with SMOTE within each training fold [27]. Hyperparameters (e.g., number of trees, depth, minimum samples per leaf) are selected on validation folds to maximize balanced accuracy.

2.5 Stage 3: Survival prediction

The survival prediction head is also a shallow MLP head that uses the same multi-scale features as Stage 2. Unlike Stage 2, the prediction head maps the four vectors to a K-dimensional vector of conditional survival probabilities over consecutive time intervals and is trained with a censoring-aware discrete-time negative log-likelihood [28]. Given a fixed setting of time intervals, the expected survival time is $\widehat{T} = \sum_{i=1}^{K} S_i \, \Delta t_i$, where S_i denotes the cumulative survival probability up to interval i, and Δt_i is the duration of interval i.

Cox proportional hazards (CoxPH) integration. For the final risk score prediction, we integrate (i) the predicted survival time, (ii) the pre-extracted, *mask-guided* PET/CT radiomics from Stage 1, (iii) the final HPV probability from Stage 2, and (iv) structured clinical indicators (as encoded and standardized as in Section 2.2) via a CoxPH model [29]. Feature selection is performed within each training fold to avoid information leakage: Univariate Cox analysis retains clinical indicators with $p<0.05$, while Least Absolute Shrinkage and Selection Operator (LASSO) regression is applied to select radiomics features.

3 Experimental Setup

Dataset overview. HECKTOR 2025 uses a multi-center, multimodal head-and-neck cancer dataset [23]. The corpus contains 1,123 pretreatment FDG-PET/CT studies from ten institutions. Major contributors are MD Anderson (444 cases; 39.6%), CHUB (216 cases; 19.2%), and University Hospital Zurich

(101 cases; 9.0%). Harmonized clinical data include RFS time and censoring indicator, HPV status, demographics, and staging. Across the full cohort, RFS information is available for 1,052 patients: 843 censored and 209 events; HPV status is available for 873 patients: 587 positive and 286 negative [23]. Our method was developed with the organizer-released training subset, including 726 patients from seven centers for all three tasks, while the remaining data was retained by the organizers for online validation and testing.

Training subset. All internal validation, model selection, and ablations were conducted on the organizer-released training subset available to us. The effective label availability in the training data is as follows:

- Tumor Segmentation (Task 1): 680 cases with valid GTVp/GTVn masks.
- Survival Prediction (Task 2): 678 cases with RFS labels (542 censored and 136 non-censored).
- HPV status classification (Task 3): 588 cases with HPV labels (58 HPV-negative and 530 HPV-positive).

Data splits. We used two complementary validation schemes on the organizer-released training subset. (i) *Five-fold patient-level cross-validation:* patients were randomly partitioned into five disjoint folds at the *patient* level. Task-specific folds were induced by intersecting the global split with each task's label-available subset (segmentation, HPV, RFS). (ii) *Five-fold center-out validation:* the seven centers were grouped into five folds by holding out one large center or a pair of smaller centers per fold to balance validation size; all patients from held-out centers form the validation set for that fold, ensuring that the validation distribution is center-disjoint from training.

Challenge protocol. We followed the official HECKTOR 2025 rules: Our method was developed merely on the released training subset, and the performance on the held-out data was assessed by the organizers. The testing labels are not public, and no external data was used.

4 Results

All quantitative results reported below are the mean across five-fold patient-level cross-validation on the released training data, which guided our model selection during development. For the final challenge submission, we additionally trained five models using five-fold center-out validation. The final results ensemble the output of ten models. The results on the held-out testing data were computed by the organizers and have not been released yet.

4.1 Performance of Tumor Segmentation (Task 1)

The evaluation metrics follow the official HECKTOR protocol: GTVp is tracked by mean Dice, whereas GTVn emphasizes both volumetric overlap and lesion

Table 1. Five-fold cross-validation results for Task 1.

Method	Dice(GTVp)	Dice(GTVn)	F1(GTVn)
Dice-only	0.7643	0.6616	0.7164
ODA (Dice+Focal Tversky+stabilizers)	0.8161	0.8206	0.8391

Table 2. Five-fold cross-validation results for Task 3. AUC is used for model selection within each fold. Sensitivity and specificity are computed at the validation threshold that maximizes balanced accuracy.

Method	AUC	Sensitivity	Specificity
Deep-only	0.9535	0.9863	0.5338
Deep+Rad+Clin	0.9601	0.9795	0.7667

detectability. Accordingly, we report Dice results for GTVp and GTVn, whereas for GTVn, the aggregated F1 score is reported to evaluate detection performance.

Table 1 presents the five-fold cross-validation results on the training data, which demonstrates that compared to using Dice loss alone (Dice-only), ODA improves GTVn detectability and overlap (higher lesion F1 and GTVn Dice) and also raises GTVp Dice.

4.2 Performance of HPV status classification (Task 3)

During development, model selection within each fold was based on AUC (threshold-independent). After selecting the model, we chose a decision threshold on the corresponding validation split to maximize balanced accuracy, and then reported sensitivity and specificity at that threshold.

Table 2 presents the five-fold cross-validation results for the deep learning model (Deep-only) and the additional integration with radiomics features and clinical indicators (Deep+Rad+Clin). Compared to the deep-only baseline, adding radiomics features and clinical indicators yields a higher AUC and substantially improves specificity with only a minor degrade in sensitivity, thereby increasing balanced accuracy at the selected thresholds.

4.3 Performance of RFS prediction (Task2)

Following the official HECKTOR protocol, the survival prediction performance is evaluated with C-index [30]. **Table 3** presents the five-fold cross-validation results for the deep learning model (Deep-only) and the CoxPH integration model, which demonstrates that the integration of radiomics features, clinical indicators, and HPV probability yields a higher C-index.

Table 3. Five-fold cross-validation results for Task 2.

Method	C-index
Deep-only	0.6974
CoxPH Integration	0.7107

5 Conclusion and Limitations

In the study, we have outlined MMPL, a Multi-stage Multimodal Progressive Learning framework that mirrors the clinical workflow of *segmentation* → *diagnosis* → *prognosis*. Built on a unified PET/CT backbone with task-specific heads, MMPL links the stages through a *prior-guided progressive learning strategy* at both the parameter and the output levels. This clinically aligned learning paradigm enables knowledge to flow from segmentation to diagnosis and then to prognosis, promotes cross-task synergy. As demonstrated in the five-fold cross-validation, MMPL improved lymph node detectability and overall segmentation quality with the proposed ODA objective, increased HPV specificity at comparable sensitivity when combining deep score with mask-guided radiomics and clinical indicators, and raised the survival prediction via CoxPH integration. Our final submission ensembles ten models trained with both patient-level and center-out cross-validation and achieved top-tier validation performance across all three tasks (testing results remain pending).

Although the challenge provides radiotherapy dose maps for outcome modeling, our current implementation does not incorporate dose information because only a very small number of cases in the available training subset included usable dose maps. Future work will integrate dose maps and examine their interaction with PET/CT images and clinical indicators.

References

1. Sung, H., Ferlay, J., Siegel, R.L., et al.: Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **71**(3), 209–249 (2021).
2. Colevas, A.D., Cmelak, A.J., Pfister, D.G., et al.: NCCN Guidelines® Insights: Head and Neck Cancers, Version 2.2025. *Journal of the National Comprehensive Cancer Network* **23**(2) (2025).
3. Ang, K.K., Harris, J., Wheeler, R., et al.: Human papillomavirus and survival of patients with oropharyngeal cancer. *New England Journal of Medicine* **363**(1), 24–35 (2010).
4. Amin, M.B., Edge, S., Greene, F., et al. (eds.): *AJCC Cancer Staging Manual*, 8th ed. Springer, New York (2017).
5. Meng, M., Bi, L., Fulham, M., Feng, D. and Kim, J.: Merging-diverging hybrid transformer networks for survival prediction in head and neck cancer. In: MICCAI 2023, LNCS **14225**, 400-410 (2023)

6. Murugesan, G.K., et al.: Head and Neck Primary Tumor Segmentation Using Deep Learning (nnU-Net framework). In: *HECKTOR 2021, LNCS* **13209**, 224-235 (2022).
7. Oreiller, V., Andrearczyk, V., Jreige, M., Castelli, J., Prior, J.O., Vallières, M., Visvikis, D., Hatt, M., Depeursinge, A.: Head and neck tumor segmentation in PET/CT: the HECKTOR challenge. *Medical Image Analysis* **77**, 102336 (2022).
8. Andrearczyk, V., Oreiller, V., Boughdad, S., et al.: Automatic head and neck tumor segmentation and outcome prediction relying on FDG-PET/CT images: Findings from the second edition of the HECKTOR challenge. *Medical Image Analysis* **90**, 102972 (2023).
9. Li, G.Y., Chen, J., Jang, S.-I., Gong, K., Li, Q.: SwinCross: Cross-modal Swin Transformer for Head-and-Neck Tumor Segmentation in PET/CT Images. arXiv:2302.03861 (2023).
10. Jo, K.H., et al.: 18F-FDG PET/CT parameters enhance MRI radiomics for predicting HPV status in OPSCC. *Yonsei Medical Journal* **64**(12), 992–1002 (2023).
11. Woo, C., et al.: Development and testing of a machine-learning model for HPV status using ^{18}F-FDG PET/CT-derived parameters in OPSCC. *Korean Journal of Radiology* **24**(1), 1–12 (2023).
12. Fanizzi, A., et al.: Explainable CNN-based prediction of HPV status in OPSCC. *Scientific Reports* **14**, 16134 (2024).
13. Vallières, M., Kay-Rivest, E., Perrin, L.J., et al.: Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Scientific Reports* **7**, 10117 (2017).
14. Huynh, B.N., Groendahl, A.R., Tomic, O., et al.: Head and neck cancer treatment outcome prediction: conventional radiomics vs. deep-learning radiomics on pretreatment PET/CT. *Frontiers in Medicine* **10**, 1217037 (2023).
15. Gu, B., et al.: Prediction of 5-year progression-free survival in advanced nasopharyngeal carcinoma with pretreatment PET/CT using multi-modality deep learning-based radiomics. *Frontiers in oncology*, **12**, 899351 (2022).
16. Meng, M., Peng, Y., Bi, L., Kim, J.: Multi-task Deep Learning for Joint Tumor Segmentation and Outcome Prediction in Head and Neck Cancer. In: *HECKTOR 2021, LNCS* **13209**, 160-167 (2022).
17. Meng, M., Bi, L., Feng, D.D., Kim, J.: Radiomics-Enhanced Deep Multi-task Learning for Outcome Prediction in Head and Neck Cancer. In: *HECKTOR 2022, LNCS* **13626**, 135-143 (2023).
18. Meng, M., Gu, B., Bi, L., Song, S., Feng, D.D., Kim, J.: DeepMTS: Deep Multi-Task Learning for Survival Prediction in Patients With Advanced Nasopharyngeal Carcinoma Using Pretreatment PET/CT. *IEEE Journal of Biomedical and Health Informatics* **26**(9), 4497–4507 (2022).
19. Gu, B., et al.: Multi-task deep learning-based radiomic nomogram for prognostic prediction in locoregionally advanced nasopharyngeal carcinoma. *European journal of nuclear medicine and molecular imaging*, **50**(13), 3996-4009 (2023).
20. Meng, M., Gu, B., Kim, J.: Adaptive segmentation-to-survival learning for survival prediction from multi-modality medical images. *npj Precision Oncology* **8**, 63 (2024).
21. Andrearczyk, V., Oreiller, V., Boughdad, S., et al.: Overview of the HECKTOR Challenge at MICCAI 2020: Automatic Head and Neck Tumor Segmentation in PET/CT. In: *Head and Neck Tumor Segmentation*. LNCS, Springer (2021).
22. Andrearczyk, V., Oreiller, V., Abobakr, M., et al.: Overview of the HECKTOR Challenge at MICCAI 2022: Automatic Head and Neck Tumor Segmentation and

Outcome Prediction in PET/CT. In: *Head and Neck Tumor Segmentation and Outcome Prediction (HECKTOR 2022)*. LNCS, Springer (2023).
23. Saeed, N., Hassan, S., Hardan, S., et al.: A Multimodal and Multi-centric Head and Neck Cancer Dataset for Tumor Segmentation and Outcome Prediction. arXiv:2509.00367 (2025).
24. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation. In: MLMI, MICCAI Workshops, pp. 379–387 (2017).
25. Abraham, N., Khan, N.: A novel focal Tversky loss function with improved Attention U-Net for lesion segmentation. In: ISBI Workshops, pp. 1–4 (2019).
26. Breiman, L.: Random forests. *Machine Learning* **45**, 5–32 (2001).
27. Chawla, N.V., Bowyer, K.W., Hall, L.O., et al.: SMOTE: Synthetic Minority Oversampling Technique. *Journal of Artificial Intelligence Research* **16**, 321–357 (2002).
28. Kvamme, H., Borgan, Ø.: Continuous and discrete-time survival prediction with neural networks. *Lifetime Data Analysis* **27**, 710–736 (2021).
29. Cox, D.R.: Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34* (2), 187-202 (1972).
30. Harrell, F.E. Jr., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A.: Evaluating the yield of medical tests. *JAMA* **247**(18), 2543–2546 (1982).