From Pixels to Prognosis: Multimodal Learning for Head and Neck Cancer in the HECKTOR 2025 Challenge

Baixiang Zhao[1,2][0000−0002−3855−8718] and Surajit Ray[1][0000−0003−3965−8136]

[1] University of Glasgow, Glasgow G12 8QQ, UK
[2] China Mobile System Integration Co.,Ltd, Beijing Fengtai District, China
baixiang.zhao@glasgow.ac.uk
surajit.ray@glasgow.ac.uk

Abstract. We describe our methods for the HECKTOR 2025 challenge, which involved three tasks using PET/CT imaging and clinical data: segmentation of primary tumors and lymph nodes, recurrence-free survival prediction, and HPV-status classification. For tumor segmentation, we used a U-Net style SegResNet that achieved Dice scores of 0.52 for primary tumors and 0.38 for lymph nodes on the validation set, ranking in the Top 10. For survival prediction, we developed a multimodal model combining imaging features with clinical data, obtaining a C-index of 0.6482, also placing in the Top 10. The same framework applied to HPV-status classification yielded a balanced accuracy of 0.4655, earning 2nd place. Our results indicate that integrating features across data modalities improves performance, though class imbalance remains challenging. Future work could benefit from incorporating radiotherapy planning data and tumor volume measurements. Code is available at: https://github.com/BaixiangZ/hecktor2025

Keywords: Head and neck cancer · Multimodal learning · Deep learning · Medical image analysis · Survival prediction

1 Introduction

Head and neck cancer presents complex challenges in diagnosis and treatment planning. The HECKTOR 2025 challenge [4] focuses on three key tasks that address different aspects of clinical management.

Task 1: Tumor Segmentation requires detecting and segmenting primary tumors (GTVp) and lymph nodes (GTVn) in PET/CT images. This is difficult because lesions can be small with unclear boundaries, tumors and lymph nodes may appear similar, and PET/CT images have different resolutions and intensity characteristics.

Task 2: Survival Prediction involves predicting recurrence-free survival from FDG-PET/CT images, clinical variables, and radiotherapy dose maps. Challenges include integrating multimodal data and handling censored observations.

Task 3: HPV Status Classification aims to diagnose HPV status from imaging and clinical data, with the additional difficulty of class imbalance.

Our work presents methods for all three tasks, focusing on effective integration of PET/CT imaging with clinical data.

2 Related Work

Previous work in medical image analysis has shown the value of combining multiple data types. Myronenko et al. [3] developed SegResNet for brain tumor segmentation, while Zhao et al. [7] explored spatial guidance methods. For survival analysis, recent approaches have incorporated ranking losses and contrastive learning. Kim et al. [1] addressed class imbalance in medical classification.

The HECKTOR challenge series has advanced head and neck cancer analysis through successive iterations with new data and evaluation protocols [5].

3 Methods

3.1 Data Preprocessing

Imaging Data Preprocessing For all tasks, CT and PET volumes were processed through standardized pipelines. For Task 1, we began with resampling to isotropic 1×1×1 mm voxel spacing using B-spline interpolation for images and nearest-neighbor for labels. We cropped to a fixed neck region (200×200×310) based on PET signal distribution and applied intensity normalization: CT to [250, 250] HU then [0,1], PET to zero mean and unit variance.

For Tasks 2 and 3, CT and PET volumes were loaded with ITK-based readers, intensity-scaled, and resized to 96×96×96 voxels. All images were reoriented to RAS coordinate system and converted to tensor format.

Clinical Data Preprocessing Clinical variables (Age, Gender, Tobacco Consumption, Alcohol Consumption, Performance Status, M-stage, Treatment) were processed identically for training and inference. Continuous variables were imputed with training median and standardized; categorical variables underwent one-hot encoding with handling of missing values. Preprocessing parameters were fitted once on training data to prevent data leakage.

3.2 Model Architectures

Task 1: Segmentation Network We employed SegResNet [3] from MONAI, a U-Net-like architecture with residual connections (Figure 1). The network accepts two-channel input (CT and PET) and produces three-channel output (background, primary tumor, lymph nodes). Both encoder and decoder use residual blocks with group normalization, ReLU activation, and 3×3×3 convolutions.

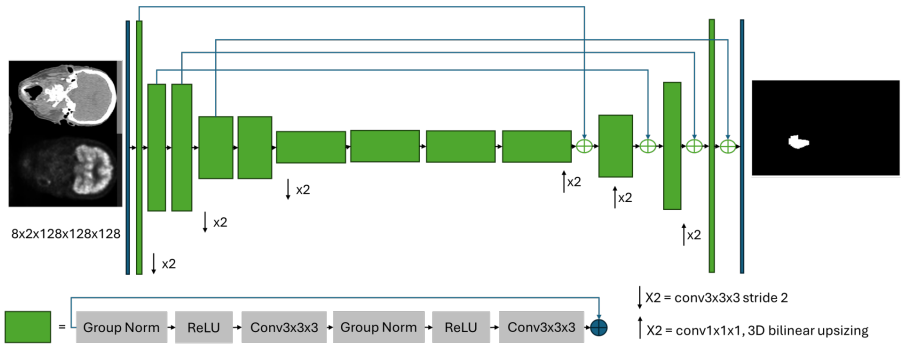

Fig. 1. SegResNet architecture for tumor and lymph node segmentation (Task 1).

Tasks 2 and 3: Multimodal Prediction Networks We developed a unified multimodal framework (Figure 2) with:

- **Imaging branch**: 3D ResNet-18 processing two-channel PET/CT → 512-D features
- **Clinical branch**: Two-layer MLP processing clinical variables → 32-D features
- **Fusion**: Concatenation + MLP → 128-D joint representation
- **Task-specific heads**: Survival risk prediction (Task 2) or HPV classification (Task 3)

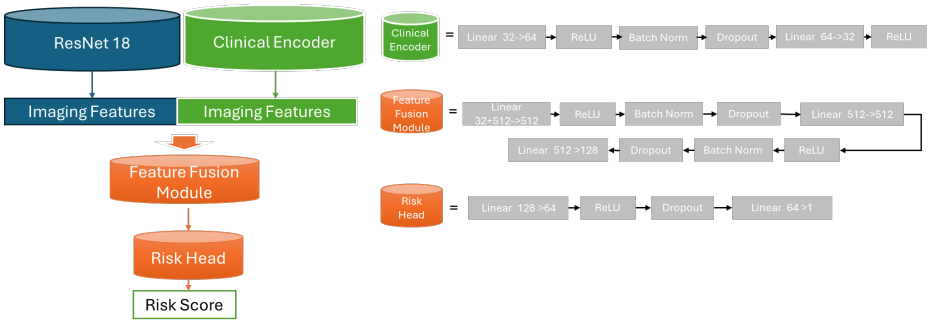

Fig. 2. Multimodal architecture for survival prediction and HPV classification (Tasks 2 and 3).

3.3 Loss Functions

Task 1: Segmentation Loss We used a composite Dice cross-entropy objective over three classes:

$$\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{CE}} \tag{1}$$

Task 2: Survival Prediction Loss Our training objective combined three components:

Cox partial likelihood:

$$\mathcal{L}_{\text{cox}} = -\frac{1}{\sum_i e_i + \varepsilon} \sum_{i:e_i=1} \left(r_i - \log \sum_{j \in R_i} e^{r_j} \right) \tag{2}$$

Pairwise ranking loss:

$$\mathcal{L}_{\text{rank}} = \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \exp\big(\alpha(r_j - r_i)\big) \tag{3}$$

Contrastive regularization:

$$\mathcal{L}_{\text{ctr}} = \frac{1}{|\mathcal{S}|} \sum_{(i,j)\in\mathcal{S}} d_{ij} + \frac{1}{|\mathcal{D}|} \sum_{(i,j)\in\mathcal{D}} \max(0, m - d_{ij}) \tag{4}$$

Total objective:

$$\mathcal{L}_{\text{surv}} = \mathcal{L}_{\text{cox}} + \lambda_{\text{rank}}\mathcal{L}_{\text{rank}} + \lambda_{\text{ctr}}\mathcal{L}_{\text{ctr}} \tag{5}$$

with $\lambda_{\text{rank}} = 0.3$, $\lambda_{\text{ctr}} = 0.1$, $\alpha = 1.0$, $m = 1.0$.

Task 3: Classification Loss Standard cross-entropy loss for binary classification:

$$\mathcal{L}_{\text{class}} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{6}$$

3.4 Training Strategy

All models used stratified 5-fold patient-level cross-validation. Training employed AdamW optimizer with learning rates of 1×10^{-4} (Task 1) and 1×10^{-3} (Tasks 2-3), early stopping with patience of 10 epochs, and appropriate batch sizes (8 for segmentation, 4 for classification). Data augmentation included random cropping, rotation, and intensity variations.

For Task 2, we used iterative joint optimization (20 iterations $\times$ 3 epochs) with gradient clipping and learning rate scheduling. For inference, we employed MONAI's sliding window inference for segmentation and bagged iCARE ensembles for survival prediction.

3.5 Evaluation Metrics

- **Task 1**: Dice similarity coefficient for GTVp and GTVn
- **Task 2**: Concordance index (C-index) measuring survival prediction accuracy
- **Task 3**: Balanced accuracy addressing class imbalance in HPV classification

The C-index is defined as:

$$\text{C-index} = \frac{\sum_{i,j} 1[r_i > r_j] \cdot 1[t_i < t_j] \cdot e_i}{\sum_{i,j} 1[t_i < t_j] \cdot e_i} \tag{7}$$

4 Results

4.1 Task 1: Tumor Segmentation

Our segmentation approach achieved a Dice score of 0.5196 for primary tumors (GTVp) and an aggregated Dice score of 0.3790 for lymph nodes (GTVn) on the validation set. In the final challenge evaluation, this performance secured a Top 10 ranking among all participating teams.

4.2 Task 2: Survival Prediction

The multimodal survival prediction model achieved a C-index of 0.6482 on the validation cohort, demonstrating reasonable concordance between predicted risks and observed survival outcomes. This performance also placed us in the Top 10 teams for this task.

4.3 Task 3: HPV Status Classification

For HPV status classification, our model achieved a balanced accuracy of 0.4655 on the validation set. Despite the challenging nature of this task with significant class imbalance, our approach earned 2nd place in this sub-challenge.

Table 1. Performance summary across all tasks

Task	Metric	Validation Score	Challenge Ranking
Task 1: Segmentation	Dice (GTVp)	0.5196	Top 10
	Dice (GTVn)	0.3790	
Task 2: Survival	C-index	0.6482	Top 10
Task 3: HPV Classification	Balanced Accuracy	0.4655	2nd Place

5 Discussion

5.1 Performance Analysis

Our results demonstrate consistent benefits of multimodal learning across all three tasks. The integration of PET and CT imaging with clinical data provided complementary information that enhanced model performance beyond single-modality approaches.

For Task 1, the performance gap between primary tumors (Dice 0.52) and lymph nodes (Dice 0.38) reflects the greater difficulty in segmenting smaller, more dispersed lymph node structures. Several factors limited our segmentation performance: absence of ensemble strategies, limited architecture exploration, and suboptimal training convergence (training Dice 0.65).

In Task 2, the ensemble of intermediate features provided stability and performance gains. The composite loss function effectively balanced ranking accuracy with feature consistency, though additional tuning of loss weights might yield further improvements.

For Task 3, class imbalance significantly impacted performance, highlighting the need for specialized handling of unbalanced datasets in medical classification tasks.

5.2 Limitations and Future Work

Several limitations present opportunities for future improvement:

Architectural Limitations: We explored limited network variants and did not incorporate ensemble strategies. Future work should evaluate broader architecture families and implement model ensembles.

Data Utilization: We did not incorporate radiotherapy planning data [5] or tumor volume metrics [2], known prognostic factors. Radiotherapy information could provide important complementary signals for survival prediction.

Class Imbalance: For HPV classification, we did not explicitly address class imbalance. Techniques such as those proposed by [1] could improve performance.

Advanced Techniques: Incorporating spatial guidance mechanisms [7] and probability contour refinement [6] could enhance segmentation boundaries. More sophisticated data augmentation and test-time augmentation could improve robustness.

Computational Constraints: Limited hyperparameter optimization and architecture search due to computational resources restricted potential performance. More comprehensive optimization could yield additional gains.

6 Conclusion

We presented comprehensive multimodal frameworks addressing all three tasks of the HECKTOR 2025 challenge. Our approaches demonstrated the value of integrating PET/CT imaging with clinical data for head and neck cancer analysis, achieving competitive results with Top 10 rankings in segmentation and survival prediction, and 2nd place in HPV classification.

The consistent performance improvements from multimodal learning across all tasks underscore the importance of comprehensive data integration in medical image analysis and computational oncology. Future work will focus on incorporating additional data modalities, implementing sophisticated handling of class imbalance, exploring ensemble strategies, and developing specialized architectures for each task.

As the field advances, such integrated approaches will play an increasingly crucial role in personalized cancer care, ultimately contributing to improved diagnosis, prognosis, and treatment planning for head and neck cancer patients.

Acknowledgments. This study was funded by Innovation Cluster Grant and EPSRC Impact Acceleration Account (Grant No EP/X525716/1) at the University of Glasgow. The code is available at: https://github.com/BaixiangZ/hecktor2025

Disclosure of Interests. The authors declare no competing interests.

References

1. Kim, J., Kim, T., Choo, J.: Epic: Effective prompting for imbalanced-class data synthesis in tabular data classification via large language models. Advances in Neural Information Processing Systems **37**, 31504–31542 (2024)
2. Kostakoglu, L., Mattiello, F., Martelli, M., Sehn, L.H., Belada, D., Ghiggi, C., Chua, N., González-Barca, E., Hong, X., Pinto, A., et al.: Total metabolic tumor volume as a survival predictor for patients with diffuse large b-cell lymphoma in the goya study. Haematologica **107**(7), 1633 (2021)
3. Myronenko, A.: 3d mri brain tumor segmentation using autoencoder regularization. In: International MICCAI brainlesion workshop. pp. 311–320. Springer (2018)
4. Saeed, N., Hassan, S., Hardan, S., Aly, A., Taratynova, D., Nawaz, U., Khan, U., Ridzuan, M., Andrearczyk, V., Depeursinge, A., Xie, Y., Eugene, T., Metz, R., Dore, M., Delpon, G., Papineni, V.R.K., Wahid, K., Dede, C., Ali, A.M.S., Sjogreen, C., Naser, M., Fuller, C.D., Oreiller, V., Jreige, M., Prior, J.O., Rest, C.C.L., Tankyevych, O., Decazes, P., Ruan, S., Tanadini-Lang, S., Vallières, M., Elhalawani, H., Abgral, R., Floch, R., Kerleguer, K., Schick, U., Mauguen, M., Bourhis, D., Leclere, J.C., Sambourg, A., Rahmim, A., Hatt, M., Yaqub, M.: A multimodal and multi-centric head and neck cancer dataset for segmentation, diagnosis, and outcome prediction (2025), https://arxiv.org/abs/2509.00367
5. Saeed, N., Hassan, S., Hardan, S., Aly, A., Taratynova, D., Nawaz, U., Khan, U., Ridzuan, M., Eugene, T., Metz, R., et al.: A multimodal head and neck cancer dataset for ai-driven precision oncology. arXiv preprint arXiv:2509.00367 (2025)
6. Zhang, W., Ray, S.: Deep probability contour framework for tumour segmentation and dose painting in pet images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 534–543. Springer (2023)
7. Zhao, B., Soraghan, J., Grose, D., Doshi, T., Di-Caterina, G.: Automatic 3d detection and segmentation of head and neck cancer from mri data. In: 2018 7th European Workshop on Visual Information Processing (EUVIP). pp. 1–6. IEEE (2018)