Could Simplified Deep Learning be More Generalizable? Predicting Recurrence-Free Survival in Head and Neck Cancer With and Without Tumor Segmentation

The Tensor Lab for Computational Medicine

Johns Hopkins Whiting School of Engineering

Koushik Rameshbabu, Laura Chen, Aaron Ge, Dr. Grace Snow

UCSF

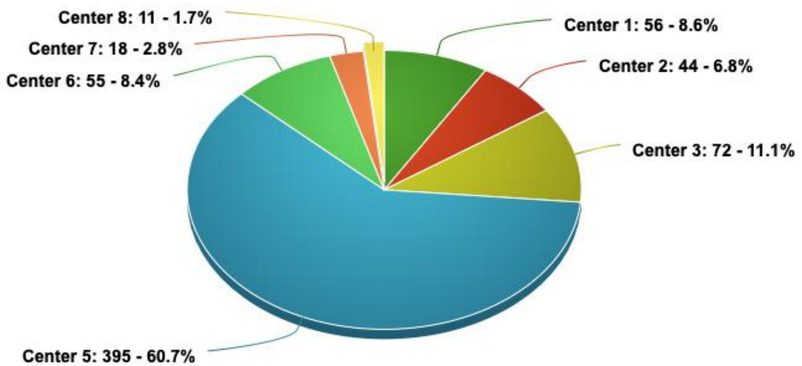
University of Maryland School of Medicine

Introduction

- Head and Neck Cancer (HNC) accounts for over 900,000 new cases and 400,000 deaths annually worldwide, making it one of the most prevalent malignancies.
- Despite advances in radiotherapy and chemotherapy, recurrence rates remain high (30–50%), emphasizing the need for accurate recurrence-free survival (RFS) prediction.
- RFS prediction can help tailor care where low-risk patients may benefit from reduced chemotherapy or surveillance while high-risk patients can receive closer monitoring and timely intervention.
- Many deep learning models rely on segmentation-derived tumor features which require expert annotation. This study investigates whether a simplified multimodal model using only PET/CT and clinical data can achieve comparable predictive performance while improving generalizability.

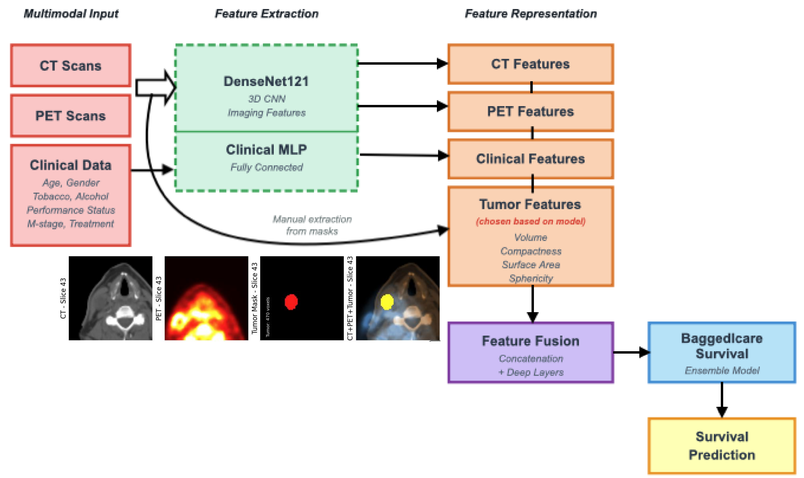
Dataset Description

- The dataset was obtained from the HECKTOR 2025 Challenge, including 651 FDG-PET/CT scans and clinical data from 7 international centers.
- PET and CT images were co-registered, resampled to 1 mm, cropped around tumors, and z-score normalized.
- Clinical variables included age, gender, tobacco, alcohol, performance status, M-stage, HPV status, treatment, and all were standardized.
- Performance status was grouped as Functional (ECOG 0–1 / Karnofsky 80–100) or Restricted (ECOG 2–4).
- Tumor features such as volume, surface area, compactness, and sphericity were extracted from segmentation masks and min–max normalized.
- The cohort had a 20.3% event rate, with 132 relapses and 519 censored cases, providing a balanced 4:1 ratio suitable for survival analysis.

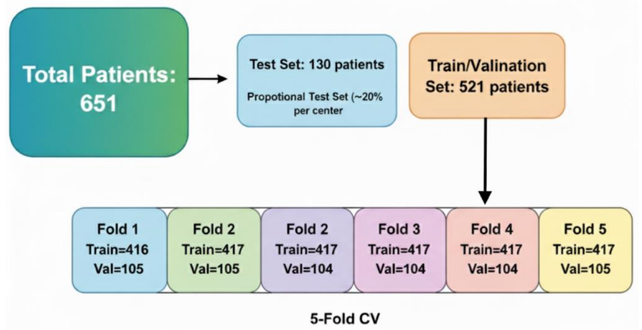


Methodology

We developed a four-stage multimodal deep learning pipeline integrating CT, PET, and clinical data for recurrence-free survival prediction, combining feature extraction, fusion, and ensemble survival modeling to capture both anatomical and clinical prognostic information.



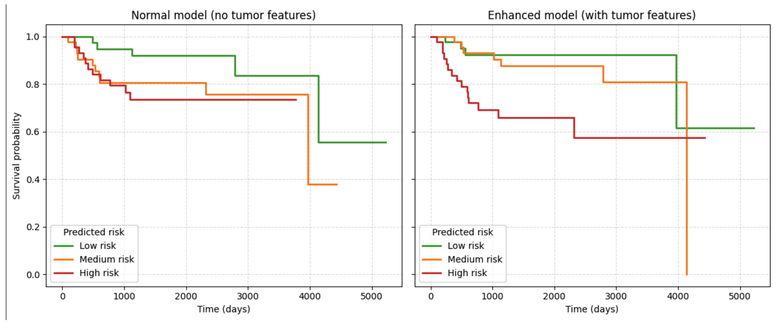
- **Stage 1** – *Multimodal Input*: Integrated co-registered FDG-PET/CT scans with key clinical variables (age, gender, tobacco, alcohol, performance status, M-stage, treatment, HPV status).
- **Stage 2** – *Feature Extraction*: Imaging features were derived using DenseNet121 (3D CNN) and clinical features via a fully connected MLP, with tumor features manually extracted from segmentation masks.
- **Stage 3** – *Feature Fusion & Representation*: Extracted CT, PET, clinical, and tumor features were concatenated to form a unified deep representation.
- **Stage 4** – *Survival Prediction*: The fused features were used in a Bagged ICARE survival ensemble for patient-level recurrence-free survival estimation.
- A stratified center-wise split ensured ~20% of patients per center in the test set (130 total) and 5-fold cross-validation on 521 remaining patients for robust evaluation.



Results

- Model performance was evaluated using the Concordance Index (C-index) and Kaplan–Meier (KM) survival curves.
- The C-index measures how well the model ranks patients by recurrence risk where higher values indicate better agreement between predicted and actual survival orderings.
- The Kaplan–Meier curves visualize patient survival over time, grouped into low, medium, and high predicted-risk tertiles. Greater vertical separation between curves reflects stronger discrimination.

Model type	Mean Validation C-index	Mean Test C-index	Best Validation C-index	Best Test C-index
With Tumor Features	0.6644	0.6603	0.7248	0.6791
Without Tumor Features	0.6371	0.5533	0.6968	0.6367



- ✓ The enhanced model (with tumor features) achieved higher mean C-index values and clearer high-risk separation on KM plots, showing improved early recurrence prediction.
- ✓ The simpler model (without tumor features) remained competitive and more generalizable across centers despite slightly lower C-index performance.

Conclusion

Tumor features enhanced recurrence-free survival prediction and improved the C-index, confirming their prognostic value. However, the simpler PET/CT + clinical model achieved comparable performance and remains more generalizable for real-world clinical deployment where segmentation data may be unavailable.