A Hybrid Vision Transformer-Based Framework for Multi-Endpoint Survival Prediction in Oropharyngeal Cancer

MU Shemanto[1], AL Imran[2], HA Azhari[3], Manju Sharma[4], X. Sharon Qi[5]

[1] Evercare Hospital Chattogram, Chattogram, Bangladesh,[2] Department of Computer Science & Engineering, United International University, Dhaka, Bangladesh

[3] Institute of Natural Science (INS), United International University, Dhaka, Bangladesh,[4] Department of Radiation Oncology, University of California, San Francisco, CA, USA

[5] Department of Radiation Oncology, University of California, Los Angeles, Los Angeles, CA, USA, Corresponding Author: X. Sharon Qi (xqi@mednet.ucla.edu)

2026 JOINT AAPM | COMP MEETING
JULY 19–22 | VANCOUVER, BC

Abstract

Purpose: To develop a hybrid **Vision Transformer (ViT)-radiomics** framework for multi-endpoint survival prediction in oropharyngeal cancer (OPC).

Methods: 512 OPC patients from the TCIA database were analyzed using pretreatment CT images, GTV segmentations, handcrafted radiomic features, ViT-derived deep features, and clinical variables. These complementary features were integrated into a Multi-Task Logistic Regression (MTLR) model to simultaneously predict overall survival (OS), local failure-free survival (LFFS), regional failure-free survival (RFFS), and distant failure-free survival (DFFS).

Results: The integrated Clinical + Radiomics + ViT model achieved the highest predictive performance (OS AUC = 0.844) and significantly improved risk stratification compared with conventional models (log-rank p < 0.001).

Conclusion: Integrating handcrafted radiomics with ViT-derived features improves survival prediction and supports personalized radiotherapy for patients with OPC.

Introduction

- Oropharyngeal cancer (OPC) requires accurate survival prediction to support personalized treatment and follow-up.
- Handcrafted radiomics captures tumour heterogeneity but has limited ability to model global imaging context.
- Vision Transformers (ViTs) learn complementary deep imaging representations from CT scans, enhancing feature extraction.
- Integrating clinical, radiomic, and ViT-derived features may improve prediction of multiple survival outcomes.
- Objective: Develop and evaluate a hybrid Clinical + Radiomics + ViT framework for simultaneous prediction of OS, LFFS, RFFS, and DFFS.

Methods and Materials

- **Dataset:** OPC-Radiomics (TCIA), 512 patients after quality control.
- **Inputs:** Pretreatment contrast-enhanced CT, GTV segmentations, and 16 clinical variables.
- **Image Preprocessing:** CT resampled to 2 × 2 × 2 mm³, tumour-centered cropping, and intensity normalization.
- **Feature Extraction:** 851 IBSI-compliant radiomic features extracted using 3D Slicer.
- **Deep Features:** Global imaging representations learned using a 3D Vision Transformer (ViT).
- **Multimodal Fusion:** Clinical, radiomic, and ViT features integrated into a unified feature set.
- **Prediction Model:** Multi-Task Logistic Regression (MTLR) for simultaneous prediction of OS, LFFS, RFFS, and DFFS.
- **Evaluation:** 5-fold stratified cross-validation using C-index, time-dependent AUC, Kaplan–Meier analysis, and 1000-bootstrap confidence intervals.

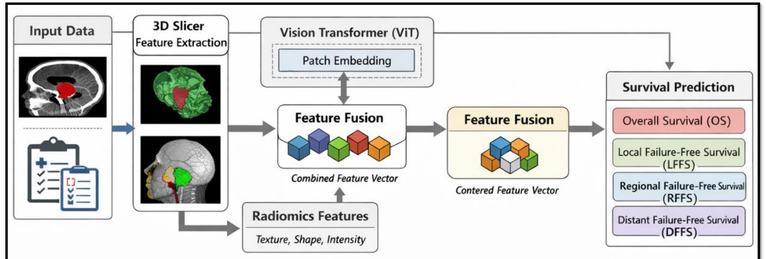

Fig. 1: Overview of the integrated radiomics and Vision Transformer framework for multi-task survival prediction.

Table 1: Area under the receiver operating characteristic curve (AUC) and 95% confidence intervals for predicting patient death (PD), local failure (LF), regional failure (RF), and distant failure (DF) within 1, 3, and 5 years after diagnosis using different feature combinations. Bolded values indicate the best performance for the corresponding outcome.

Model / Feature Combination	Outcome	1-year AUC (95% CI)	2-year AUC (95% CI)	3-year AUC (95% CI)
Clinical Only	PD	0.764 (0.702–0.826)	0.802 (0.761–0.842)	0.796 (0.755–0.836)
	LF	0.750 (0.672–0.829)	0.759 (0.690–0.827)	0.773 (0.705–0.840)
	RF	0.766 (0.671–0.861)	0.771 (0.685–0.858)	0.780 (0.696–0.864)
	DF	0.802 (0.730–0.856)	0.799 (0.743–0.856)	0.791 (0.693–0.847)
Clinical + Radiomics	PD	0.821 (0.765–0.877)	0.818 (0.776–0.877)	0.810 (0.771–0.849)
	LF	0.796 (0.717–0.876)	0.828 (0.768–0.889)	0.839 (0.783–0.896)
	RF	0.823 (0.723–0.923)	0.840 (0.763–0.917)	0.841 (0.766–0.916)
	DF	0.816 (0.730–0.902)	0.782 (0.717–0.847)	0.777 (0.711–0.842)
Clinical + ViT	PD	0.806 (0.749–0.862)	0.834 (0.797–0.871)	0.816 (0.777–0.854)
	LF	0.799 (0.730–0.867)	0.817 (0.760–0.874)	0.833 (0.780–0.887)
	RF	0.839 (0.776–0.903)	0.844 (0.783–0.905)	0.843 (0.781–0.906)
	DF	0.839 (0.766–0.912)	0.829 (0.775–0.882)	0.816 (0.761–0.872)
Clinical + Radiomics + ViT	**PD**	**0.799 (0.749–0.862)**	**0.834 (0.797–0.871)**	**0.821 (0.777–0.854)**
	LF	**0.812 (0.730–0.867)**	**0.817 (0.760–0.874)**	**0.833 (0.780–0.887)**
	RF	**0.821 (0.776–0.903)**	**0.844 (0.783–0.905)**	**0.838 (0.781–0.906)**
	DF	**0.828 (0.766–0.912)**	**0.829 (0.775–0.882)**	**0.844 (0.761–0.872)**

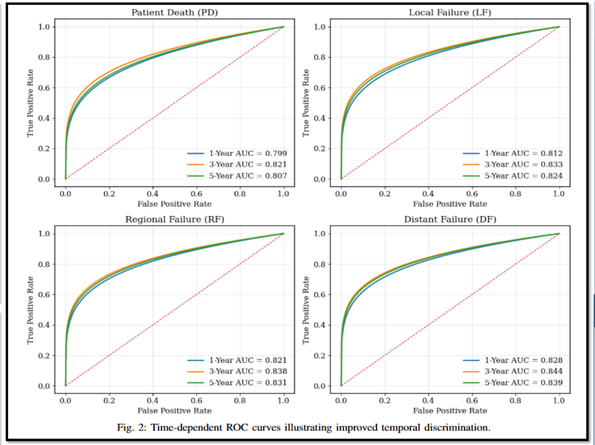

Fig. 2: Time-dependent ROC curves illustrating improved temporal discrimination.

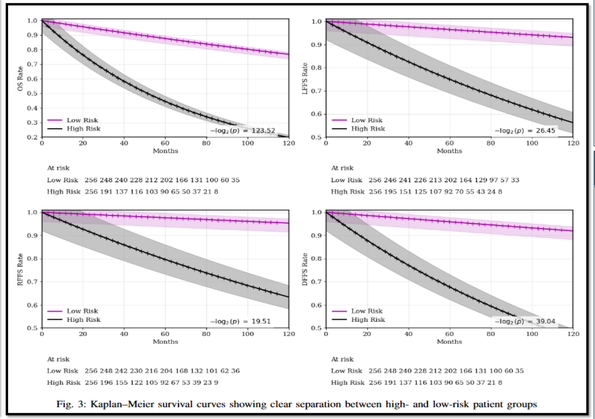

Fig. 3: Kaplan–Meier survival curves showing clear separation between high- and low-risk patient groups

Results

The proposed Clinical + Radiomics + Vision Transformer (ViT) framework (Fig. 1) achieved the best performance across all survival endpoints, with C-index values of 0.789 (OS), 0.781 (LFFS), 0.792 (RFFS), and 0.786 (DFFS). Compared with the Clinical + ViT model, incorporating handcrafted radiomic features significantly improved prognostic accuracy (ΔC-index = 0.016, p = 0.003). As summarized in Table 1, the multimodal model consistently achieved the highest time-dependent AUCs (0.816–0.854). The ROC curves (Fig. 2) demonstrate strong temporal discrimination, while the Kaplan–Meier curves (Fig. 3) show clear separation between high- and low-risk groups (log-rank p < 0.001). Overall, these results demonstrate that multimodal integration significantly enhances survival prediction and patient risk stratification in OPC.

Innovation, Impact, and Discussion

- Novel hybrid framework integrating clinical data, handcrafted radiomics, and Vision Transformer (ViT) features.
- Multi-Task Logistic Regression (MTLR) simultaneously predicts OS, LFFS, RFFS, and DFFS within a single model.
- Multimodal feature fusion significantly improved prognostic accuracy and patient risk stratification compared with single-modality approaches.
- Demonstrated particularly strong performance in HPV-positive patients and those receiving chemoradiotherapy.
- Supports personalized prognosis, risk-adaptive radiotherapy, and optimized post-treatment surveillance.
- Provides an interpretable and clinically actionable AI framework for precision oncology in oropharyngeal cancer.

Conclusions

- Hybrid ViT-radiomics framework integrating clinical, handcrafted radiomic, and Vision Transformer (ViT) features for multi-endpoint survival prediction.
- Achieved superior prognostic accuracy and patient risk stratification compared with conventional and single-modality approaches.
- Supports personalized radiotherapy planning, risk-adaptive treatment, and clinical decision-making in oropharyngeal cancer.
- Demonstrates the value of combining handcrafted and deep imaging features for precision oncology.
- Future work: Independent external cohort validation, incorporation of advanced deep learning and multimodal foundation models, and to further improve clinical robustness and generalizability.

Contact

Prof. X. Sharon Qi, Ph.D., DABR
Department of Radiation Oncology
University of California, Los Angeles (UCLA)
Email: xqi@mednet.ucla.edu
Website: https://www.uclahealth.org/node/110181

References

1. [1] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.
2. [2] M. Chen, K. Wang, and J. Wang, "Vision transformer-based multilabel survival prediction for oropharynx cancer after radiation therapy," In ternational Journal of Radiation Oncology* Biology* Physics, vol. 118, no. 4, pp. 1123–1134, 2024.
3. [3]J. Y. Y. Kwan, J. Su, S. H. Huang, L. S. Ghoraie, W. Xu, B. Chan, K. W. Yip, M. Giuliani, A. Bayley, J. Kim, et al., "Radiomic biomarkers to refine risk models for distant metastasis in hpv-related oropharyngeal carcinoma," International Journal of Radiation Oncology* Biology* Physics, vol. 102, no. 4, pp. 1107–1116, 2018.