Before We Treat, Can We Tell? A Locoregional Recurrence Signature in Head & Neck Cancer

Hasan Shaikh*†[1], Amal Joseph Varghese*[1], Balu Krishna S[1], Julia Priyadarshini Rao[1], Ezhil Sindhanai[1], Rajendra Benny Kuchipudi[1], Manu Mathew[1], Jino Wilson Victor[1], Rajesh I[1], Simon Pavamani[1], and Hanny Mary Thomas T[1]

[1]Quantitative Imaging Research and Artificial Intelligence Lab, Department of Radiation Oncology, Christian Medical College, Vellore, Tamil Nadu 632004, India

Background:

Locoregional recurrence remains a critical challenge in head and neck cancer management, affecting 50-60% of patients despite current treatment protocols. Existing clinical staging systems provide limited predictive accuracy for individual patients and limit personalized treatment decisions. This prediction challenge is compounded by the high-dimensional nature of radiomics data, where hundreds of imaging biomarkers must be analyzed in relatively small patient cohorts: a classic $p \gg n$ problem that suffers from feature instability and selection bias. In this setting, the main failure is feature instability: small resamples or fold changes yield different subsets, followed by selection leakage that can overstate accuracy. The key technical question is therefore not "which classifier," but "which sparse, stable subset of radiomics+clinical features carries signal that holds up on new patients?" We compare three concrete selection mechanisms LASSO (embedded shrinkage), SelectKBest (univariate filter), and metaheuristic subset search (Particle Swarm Optimization, Whale Optimization Algorithm, Grey Wolf Optimizer, Genetic Algorithm, Simulated Annealing), used with common classifiers, and ask whether a compact consensus signature emerges that generalizes under strict train/test separation. Our hypothesis is that, in small-n HNC radiomics, a low-dimensional, linearly exploitable signature that combines shape/size and texture heterogeneity with a few clinical covariates will outperform higher-capacity models that tend to overfit and chase noise.

Aim and Objective:

This study aims to (i) systematically compare LASSO, SelectKBest, and metaheuristics subset search across multiple classifier families across both radiomics and radiomics + clinical inputs for LRR prediction; (ii) quantify generalization using a fixed Train/Validation/Test protocol and report ROC-AUC with 95% CIs and (iii) identify the single best-performing model and report its final selected features for interpretability and replication.

Materials and Methods:

Baseline CTs were contoured by the treating radiation oncologist; radiomics were extracted from the GTV using PyRadiomics, yielding 103 features spanning first-order intensity, 3D shapes, and texture families (GLCM, GLRLM, GLDM, NGTDM). We also considered 8 clinical variables (Age, Weight, primary site/Location, T/N/M stage, AJCC 7th stage, HPV/p16). We evaluated four configurations: (i) radiomics only, (ii) radiomics + clinical, (iii) hybrid feature selection on radiomics (Bootstrap-LASSO followed by metaheuristic search), and (iv) the same hybrid pipeline on radiomics + clinical. Classifiers included Logistic Regression, Naive Bayes, SVM, Decision Tree, and Random Forest. We performed a fixed 80/20 patient-level split. All feature selection and hyperparameter tuning were performed exclusively within the training set using stratified 5-fold cross-validation. Final models were refit on the entire training set with the consensus features and evaluated once on the held-out test set. Performance was summarized as ROC-AUC with 95% bootstrap CIs.

*These authors contributed equally to this work.

†Corresponding author: hasan.shaikh.inst@cmcvellore.ac.in

Results:

We analyzed 163 prospectively recruited HNC patients (55 LRR; 108 disease-free) treated with chemoradiation from 2020 to 2024. Adding clinical variables to radiomics improved generalization across models. The best held-out performance was Logistic Regression trained on a GWO-selected sparse subset in the radiomics + clinical setting: Test AUC = 0.81 [0.62, 0.95]; Train AUC = 0.79 [0.71, 0.86]. The best-performing model selected 10 features: 4 clinical (Age, AJCC_Stage, T_Stage, Location) and 6 radiomics reflecting tumor shape (Maximum2DDiameterSlice, MinorAxisLength) and texture heterogeneity (LargeDependenceEmphasis, RunLengthNonUniformityNormalized, Idm, Imc1). Random Forest was competitive but showed a gap between high Train AUC and lower Test AUC ($\approx$ 0.64-0.74 across settings). SVM improved markedly when clinical variables were added (from $\approx$ 0.32-0.36 with radiomics only to 0.78 with radiomics + clinical).

Conclusion:

Key finding: A simple logistic regression using a 10-feature GWO-derived radiomics + clinical signature achieved the strongest held-out discrimination, supporting small, stability-audited feature sets in this $p \gg n$ context. Next step: We will validate this model on data from other hospitals and check calibration and clinical benefit (decision-curve analysis).