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

Authors: Hasan Shaikh, Balu Krishna S, et al.

Affiliation: Quantitative Imaging Research and AI Lab, Dept. of Radiation Oncology, CMC Vellore

15th Research Day, CMC Vellore | October 30-31, 2025

Background & Problem Statement

Clinical context and technical challenges in locoregional recurrence prediction

Clinical Problem

LOCOREGIONAL RECURRENCE

50-60%†

Recurrence Rate

High rates of locoregional recurrence persist despite current treatment protocols in head and neck cancer patients

CURRENT LIMITATIONS

⚠

Staging Systems

Traditional TNM staging and clinical parameters demonstrate limited predictive power for individual patient outcomes and cannot effectively guide treatment personalization

Technical Challenge

High-Dimensional Data Problem

Number of Features (p) >> Number of Patients (n)

Feature Space
103 radiomics features extracted per patient

Sample Size
163 patients in final study cohort

Consequences
Feature instability, overfitting risk, and selection bias

† Chang JH, Wu CC, Yuan KS, Wu ATH, Wu SY. Locoregionally recurrent head and neck squamous cell carcinoma: incidence, survival, prognostic factors, and treatment outcomes. *Oncotarget*. 2017;8(33):55600-55612. doi:10.18632/oncotarget.17469

Research Question & Methodology

Systematic approach to sparse and stable signature identification

PRIMARY RESEARCH QUESTION

Can we identify an interpretable, sparse radiomic signature for locoregional recurrence prediction that demonstrates stability and generalizability?

Analytical Workflow

Systematic Methodology

Feature Selection Strategies	Classification Algorithms	Evaluation Framework

• LASSO (L1 regularization)
• SelectKBest (univariate)
• Metaheuristic optimization:
  ◦ Grey Wolf (GWO)
  ◦ Particle Swarm (PSO)
  ◦ Whale (WOA)
  ◦ Genetic Algorithm (GA)
  ◦ Simulated Annealing (SA)
	• Logistic Regression
	• Support Vector Machine
	• Random Forest
	• Decision Tree
	• Naïve Bayes
		• Feature stability analysis
		• Cross-validation performance
		• Held-out test set evaluation
		• Model interpretability
		• Clinical utility assessment

Study Design & Imaging Protocol

Prospective data collection with standardized acquisition parameters

PROSPECTIVE STUDY DESIGN: This study was designed prospectively with predefined imaging protocols established at the initiation of data collection (2020-2024). All imaging acquisitions followed standardized protocols to ensure data quality and reproducibility.

TOTAL PROSPECTIVE COHORT

N = 1,466

Head and Neck Cancer patients treated at CMC Vellore (2020-2024)

CT Imaging Acquisition Parameters

CT Scanners: SIEMENS Biograph 6, SIEMENS SOMATOM Definition AS, GE Healthcare Discovery CT750 HD

Parameter	Specification
Energy Range	100.0 - 130.0 kVp
Exposure Range	5.0 - 350.0 mAs
Slice Thickness	2.5 - 5.0 mm
In-Plane Resolution	0.78125 × 0.78125 mm² to 1.367188 × 1.367188 mm²
Contrast Protocol	Contrast-enhanced and non-contrast imaging included

PRIMARY RADIATION TREATMENT COHORT

N = 367

Patients who received primary radiation ± chemotherapy (n = 1,099 excluded: surgery, palliative care, other treatments)

Patient Selection & Dataset Splitting

Application of inclusion/exclusion criteria to N = 367 primary radiation patients

Inclusion Criteria

1. Diagnosed head and neck cancer
2. Contrast-enhanced CT of head and neck available
3. No treatment before CT scan (treatment-naïve imaging)
4. Treatment with radiation and/or chemoradiation only

Exclusion Criteria (n = 204 excluded)

1. Treatment non-completion (n = 2)
2. Tumor volume unavailable - only primary disease visible (n = 13)
3. Presence of significant image artifacts (n = 3)
4. Did not receive radiation treatment as planned (n = 11)
5. Follow-up of one year not available (n = 175)

↓

FINAL STUDY COHORT

N = 163

Locoregional Recurrence (LRR)

55

33.7%

No Locoregional Recurrence

108

66.3%

↓

Training Cohort

n = 130

(80% of dataset)

Used for feature selection, hyperparameter tuning, and cross-validation

Test Cohort

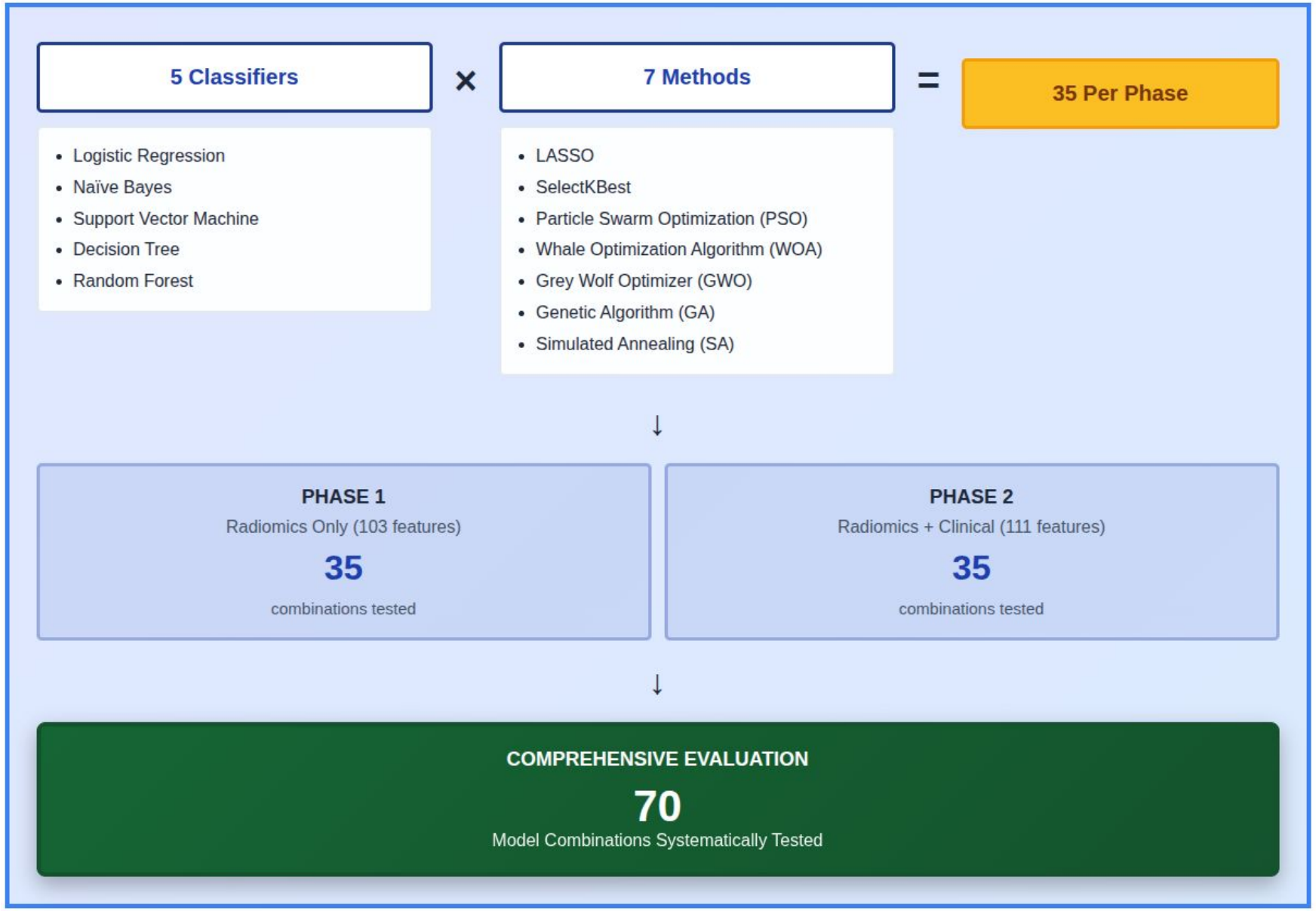
n = 33

(20% of dataset)

Held-out for final model evaluation

Comprehensive Model Evaluation

Systematic testing of 70 model combinations across two feature sets



Optimal Model Selection

Best performing combination from 70 systematic evaluations

WINNING MODEL

Logistic Regression + Grey Wolf Optimizer

Using Radiomics + Clinical Features (111 features → 10 selected)

TEST AUC	TRAIN AUC	SELECTED FEATURES
0.81	**0.79**	**10**
[0.62 - 0.95]	[0.71 - 0.86]	Sparse signature

The Sparse 10-Feature Signature

Limitations & Future Work

Current constraints and research priorities

Limitations

- Single-center (CMC Vellore)
- Indian cohort only
- Modest sample size (n=163)
- Centralized data storage
- Protocol variations across scanners
- Requires external validation

Future Directions

- Multi-center validation
- Federated learning implementation
- Prospective clinical trial
- Protocol harmonization
- Clinical workflow integration
- Longitudinal outcome tracking

Conclusion

Summary of key findings and clinical implications

Key Findings

Developed a **sparse 10-feature signature** (4 clinical + 6 radiomic) for locoregional recurrence prediction in head and neck cancer.

Achieved **AUC 0.81 [0.62-0.95]** on test set with minimal overfitting (train AUC 0.79).

Clinical features + radiomics significantly outperformed radiomics alone (0.81 vs 0.73).

Clinical Impact

This interpretable model enables **risk-stratified treatment planning** and has potential for clinical deployment following multi-center validation and federated learning-based collaborative refinement.

Thank you

Acknowledgement:

IndiaAlliance
DBT wellcome

- Dr. Hannah Mary Thomas T
- Prof. Balu Krishna S