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

From Voxels to Mutations: Stability-Aware Multi-Target CT Imaging Prediction of EGFR and KRAS Mutations

Mohammad Salmanpour[1,2,3]; Somayeh S. Mehrnia[4]; Fatemeh Razavi[5]; Helia Abedini[6]; Niloufar Rahimi Moghadam[7]; Shahram Taeb[8]; Arman Rahmim[1,2]

1) Department of Basic and Translational Research, BC Cancer Research Institute; 2) Department of Radiology, University of British Columbia; 3) Technological Virtual Collaboration Company (TECVICO CORP.), Vancouver, BC, Canada; 4) Department of Radiology Technology, Faculty of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences; 5) Department of Biology, SR.C., Islamic Azad University; 6) Department of Computer and Information Technology, Islamic Azad University; 7) Blood Transfusion Research Center, High Institute for Research and Education in Transfusion Medicine, Tehran, Iran; 8) Department of Radiology, School of Paramedical Sciences, Guilan University of Medical Sciences, Rasht, Iran.

ABSTRACT

- This multicenter study developed a CT-based radiogenomic framework to non-invasively predict EGFR and KRAS mutations in **non-small cell lung cancer** (NSCLC). Using data from 1,082 patients across 12 public cohorts, the study compared **handcrafted radiomics features (HRF), deep radiomics features (DRF), and combined features** under supervised (SL) and semi-supervised learning (SSL).
- HRFs combined with SSL showed the best generalizability. The top model achieved strong internal performance, with a cross-validation ROC-AUC of 0.93, and maintained better external performance than DRF models, with an external ROC-AUC of 0.77. SSL significantly improved performance compared with SL.
- Overall, **standardized HRFs with SSL** offer a more robust and generalizable approach for dual EGFR/KRAS mutation prediction across heterogeneous multicenter CT datasets.

CONTACT

Mohammad Salmanpour, Ph.D.
Department of Basic and Translational Research, BC Cancer Research Institute, Vancouver, BC, Canada
Contact: msalman@bccrc.ca

INTRODUCTION

Lung cancer is the leading cause of cancer-related mortality, with NSCLC accounting for approximately 85% of cases [1]. CT-based radiogenomics may non-invasively identify actionable EGFR and KRAS mutations [2] when tissue-based genomic testing is limited [3].

METHODS AND MATERIALS

Data and Imaging:

- Included 1,082 NSCLC patients from 12 CT cohorts, with genomic data for 136 patients and expert-validated tumor segmentations by two physicians (Chart 1).
- Predicted outcomes: **EGFR-mutant, KRAS-mutant, and wild-type**.

Features and Tools

- **496 IBSI-compliant HRFs and 511 DRFs** extracted with the **PySERA** framework.

Supervised Learning (SL)

- Tested 56 dimensionality-reduction methods and 29 classifiers on NSCLC data (n=123) using five-fold cross-validation, with external validation on TCGA-LUAD (n=13)

Semi-Supervised Learning (SSL)

- In each fold, the model was trained on labeled patients, pseudo-labeled 946 unlabeled cases from 10 centers, and combined both datasets to expand training.

Robust Model Selection

Models were ranked using a stability-aware composite score combining normalized mean performance (M_i) and stability (S_i), derived from variability (σ_i) across five parameters, with scores ranging from 0 to 2.

$$M_i^{norm} = \frac{M_i - M_i^{min}}{M_i^{max} - M_i^{min}} \quad (1)$$

$$\sigma_i^{norm} = \frac{\sigma_i - \sigma_i^{min}}{\sigma_i^{max} - \sigma_i^{min}} \quad (2)$$

$$S_i = 1 - \sigma_i^{norm} \quad (3)$$

$$Score = \frac{1}{10}\sum_{i=1}^{5}(M_i^{norm} + S_i) \quad (4)$$

SHapley Additive exPlanations (SHAP) Analysis

- SHAP quantified HRF contributions to predicting wild-type, KRAS, and EGFR mutation classes, improving model interpretability

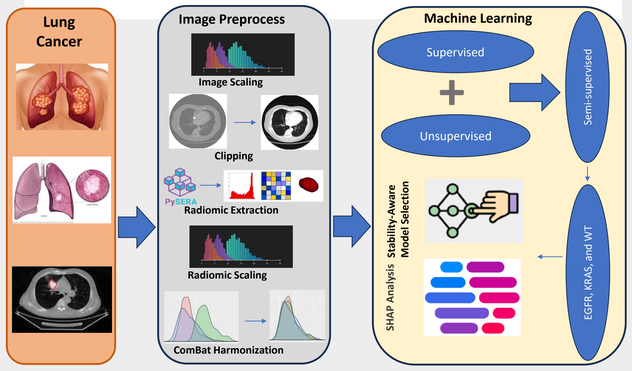

Chart 1. Overall study workflow. RFFI: Random Forest Feature Importance; HGB: HistGradient Boosting; LGBM: LightGBM; PI: Permutation Importance; MI: Mutual Information; RandF: Random Forest; FAHG: Feature Agglomeration for Hierarchical Grouping; SHAP: SHapley Additive exPlanations; EGFR: Epidermal growth factor receptor; KRAS: Kirsten rat sarcoma virus; WT: Wild-Type.

RESULTS

- **HRF-based models with SSL** showed superior robustness and generalizability across data splits.
- **HRF-based feature** importance linked with Histogram Gradient Boosting achieved (Chart 2):
  - Cross-validation ROC-AUC: **0.93 ± 0.02**
  - Cross-validation accuracy: **0.88 ± 0.01**
  - External testing ROC-AUC: **0.77 ± 0.07**
  - External testing accuracy: **0.77 ± 0.00**
- **DRF-based models** or a mixture of **DRF and HRFs** had **high cross-validation performance** (~0.97) but markedly reduced external AUC (~60%).
- Across all algorithms, **SSL** significantly **outperformed SL** ($P < 0.05$).
- SHAP analysis across five model combinations identified **morphology and texture** as key predictors, with **Morph-VolDens-AABB** most influential overall. Distinct multiscale features characterized wild-type, KRAS, and EGFR classes (Table 1).

Chart 2. Mean ± standard deviation performance of the top five semi-supervised models based on (**top**) handcrafted radiomics features, (**middle**) deep radiomics features, and (**bottom**) combined handcrafted-deep radiomics features, evaluated using five-fold cross-validation and external testing.

Table 1. Top 10 SHAP-ranked features for each class, showing class-specific feature importance scores and feature-category abbreviations. Morph-VolDens-AABB, Morphology-Volume Density (Axis-Aligned Bounding Box); Stat-CoV, Statistics-Coefficient of Variation; GLRLM-LRE-2Davg, Gray Level Run Length Matrix-Long Runs Emphasis (2D Averaged); MI-I3-Int, Moment Invariants-Third Principal Invariant (Intensity-weighted); Morph-AreaDens-MVEE, Morphology-Area Density (Minimum Volume Enclosing Ellipsoid); Morph-AreaDens-OMBB, Morphology-Area Density (Oriented Minimum Bounding Box); Morph-Comp1, Morphology-Compactness 1; Diag-MorphVox-InterpROI, Diagnostics-Morphological Mask Voxel Count (Interpolated ROI); GLCM-DiffAvg-2.5Dcomb, Gray Level Co-occurrence Matrix-Difference Average (2.5D Combined); Diag-IntBBdimZ-InterpROI, Diagnostics-Intensity Mask Bounding Box Dimension Z (Interpolated ROI); GLCM-ClustProm-3Dcomb, Gray Level Co-occurrence Matrix-Cluster Prominence (3D Combined); Stat-QCOD, Statistics-Quartile Coefficient of Dispersion; Stat-MinInt, Statistics-Minimum Intensity; MI-I2-Int, Moment Invariants-Second Principal Invariant (Intensity-weighted); GLCM-JointAvg-2Dcomb, Gray Level Co-occurrence Matrix-Joint Average (2D Combined); GLDZM-HGZE-2D, Gray Level Distance Zone Matrix-High Gray Level Zone Emphasis (2D); GLDZM-SDHGE-3D, Gray Level Distance Zone Matrix-Small Distance High Gray Level Emphasis (3D); GLDZM-SDHGE-2.5D, Gray Level Distance Zone Matrix-Small Distance High Gray Level Emphasis (2.5D); GLDZM-SDHGE-2D, Gray Level Distance Zone Matrix-Small Distance High Gray Level Emphasis (2D); Morph-Elong, Morphology-Elongation; Morph-Comp2, Morphology-Compactness 2.

No KRAS or EGFR mutation		KRAS mutation		EGFR mutation		
Feature Name	SHAP Value	Feature Name	SHAP Value	Feature Name	SHAP Value	
Morph-VolDens-AABB	0.301	MI-I3-Int	0.281	GLDZM-SDHGE-3D	0.294	0.33
Stat-CoV	0.207	Stat-CoV	0.167	GLDZM-SDHGE-2.5D	0.25	0.29
GLRLM-LRE-2Davg	0.154	GLCM-ClustProm-3Dcomb	0.122	GLDZM-SDHGE-2D	0.125	0.25
Morph-AreaDens-MVEE	0.101	Stat-QCOD	0.116	Morph-Elong	0.117	0.21
Morph-AreaDens-OMBB	0.079	Stat-MinInt	0.098	Morph-Comp2	0.092	0.17
Morph-Comp1	0.078	Morph-VolDens-AABB	0.095	GLRLM-LRE-2Davg	0.075	0.13
Diag-MorphVox-InterpROI	0.078	MI-I2-Int	0.092	Morph-VolDens-AABB	0.069	0.09
GLCM-DiffAvg-2.5Dcomb	0.075	GLRLM-LRE-2Davg	0.089	Stat-CoV	0.067	0.05
Diag-IntBBdimZ-InterpROI	0.075	GLCM-JointAvg-2Dcomb	0.079	Morph-AreaDens-OMBB	0.063	0.01
MI-I3-Int	0.069	GLDZM-HGZE-2D	0.068	Morph-AreaDens-MVEE	0.046	0

CONCLUSIONS

HRF-based radiogenomics with SSL demonstrated the greatest robustness, interpretability, and external generalizability in multicenter NSCLC mutation prediction. DFR and fused models showed potential but require task-specific optimization and improved integration to perform reliably across institutions.

REFERENCES

1. Rebecca L. Siegel, T.B.K., Angela N. Giaquinto M, Hyuna Sung, Ahmedin Jemal DVM,, Cancer statistics, 2025. CA Cancer J Clin, 2025. 75(1): p. 10-45.
2. Rizzo S, P.F., Buscarino V, De Maria F, Raimondi S, Barberis M, et al, CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. European radiology, 2016. 26(1): p. 32-42.
3. Chaddha U, A.A., Ghori U, Kheir F, Debiane L, McWilliams A, et al. , Safety and Sample Adequacy for Comprehensive Biomarker Testing of Bronchoscopic Biopsies: An American Association of Bronchology and Interventional Pulmonology (AABIP) and International Association for the Study of Lung Cancer (IASLC) Clinical Practice Guideline. Journal of Thoracic Oncology, 2025.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com