Original Article
da Silva ACC, et al. Radiol Bras. 2026;59:e20250005
doi: 10.1590/0100-3984.2025.0005

Evaluation of a radiomics workflow in chest CT: a pilot study of lung lesion segmentation, feature extraction, and clustering

Avaliação de um fluxo de radiômica em TC de tórax: estudo piloto de segmentação de lesões pulmonares, extração de características e clusterização

Ana Carolina Costa da Silva[1,a], Paulo Henrique Ruis Garcia[2,3,b], Karina Yukimi Pexito Sakurai[4,c], Douglas Carli Silva[3,d], Maria Vitória David Ludwig[5e], Fabiana Reis Decicino Campos[5f], Márcio Valente Yamada Sawamura[4,5g], Hye Ju Lee[5hf]

Abstract

Objective: To evaluate a radiomics workflow applied to chest computed tomography (CT), examining its feasibility and performance across the steps of lung lesion segmentation, feature extraction, and clustering, as well as exploring its potential to distinguish between benign and malignant lung lesions.

Materials and Methods: This was a retrospective study including CT images of 32 histopathology-confirmed lesions (11 benign; 21 malignant). Lesions were segmented by using syngo.via Frontier Radiomics software, after which > 900 radiomic features were extracted with the same software. Data processing—including normalization, dimensionality reduction (principal component analysis), and unsupervised (k-means) clustering—was conducted in Python.

Results: The best clustering solution was obtained with a k = 2 (silhouette score = 0.4668), effectively separating benign lesions, which exhibited low heterogeneity, from malignant lesions, which exhibited complex patterns of texture and shape.

Conclusion: The application of a radiomics workflow proved feasible and potentially useful for differentiating between benign and malignant lung lesions on CT, underscoring its value as a complementary tool in thoracic oncology. However, future studies using multicenter datasets, blinded evaluations, and more structured integration environments are warranted in order to support clinical validation and prototyping.

Keywords: Radiomics; Lung neoplasms; Tomography, X-ray computed; Radiographic image interpretation, computer-assisted; Cluster analysis; Biomarkers; Machine learning.

Resumo

Objetivo: Avaliar um fluxo de trabalho em radiômica aplicado à tomografia computadorizada de tórax, examinando a viabilidade e desempenho nas etapas de segmentação de lesões pulmonares, extração de características e agrupamento, além de explorar o potencial para distinguir lesões pulmonares benignas de malignas.

Materiais e Métodos: Estudo retrospectivo incluindo 32 lesões pulmonares com diagnóstico histopatológico (11 benignas, 21 malignas). As segmentações das lesões foram realizadas com o syngo.via MM Radiomics Frontier, seguido da extração de > 900 características radiômicas pelo mesmo software. Os dados foram processados em Python com normalização, redução de dimensionalidade (PCA) e agrupamento não supervisionado (k-means).

Resultados: O modelo k-means com k = 2 apresentou melhor desempenho (coeficiente de silhueta = 0,4668), distinguindo adequadamente lesões benignas, com baixa heterogeneidade, das malignas, caracterizadas por maior complexidade textural e morfológica.

Conclusão: A aplicação da radiômica mostrou-se factível e potencialmente útil na diferenciação entre lesões pulmonares benignas e malignas em exames de TC, reforçando seu valor como ferramenta complementar em oncologia torácica. No entanto, recomenda-se a realização de estudos futuros com conjuntos de dados multicêntricos, avaliações cegas e ambientes de integração mais estruturados para validação e prototipagem clínica.

Unitermos: Radiômica; Neoplasias pulmonares; Tomografia computadorizada por raios X; Interpretação de imagem radiográfica assistida por computador; Análise por conglomerados; Biomarcadores; Aprendizado de máquina.

Running title: **Radiomics in chest CT: a pilot study**

1. Siemens Healthineers, Computed Tomography, São Paulo, SP, Brazil. 2. Institute of Energy and Nuclear Research, University of São Paulo, São Paulo, SP, Brazil. 3. Siemens Healthineers, Education Services, São Paulo, SP, Brazil. 4. Institute of Radiology, Hospital das Clínicas, University of São Paulo School of Medicine, São Paulo, SP, Brazil. 5. Hospital Sírio-Libanês, São Paulo, SP, Brazil.

Correspondence: Ana Carolina Costa da Silva. Siemens Healthineers, Av. Mutinga, 3800, São Paulo, SP, Brazil, 05110-000. Email: ana.carolina.tnlr@gmail.com

a. https://orcid.org/0000-0003-1994-9008; b. https://orcid.org/0009-0004-5154-6035; c. https://orcid.org/0009-0000-1731-4135; d. https://orcid.org/0000-0002-8550-6312; e. https://orcid.org/0000-0001-7020-4306; f. https://orcid.org/0009-0003-4414-5113; g. https://orcid.org/0000-0002-9424-9776; h. https://orcid.org/0000-0001-8263-9059.

Submitted: 21 Jan 2026. Revised: 23 Mar 2026. Accepted: 23 Mar 2026.

Editor in charge: Dr. Valdair Francisco Muglia.

How to cite this article:

da Silva ACC, Garcia PHR, Sakurai KYP, Silva DC, Ludwig MVD, Campos FRD, Sawamura MVY, Lee HJ. Evaluation of a radiomics workflow in chest CT: a pilot study of lung lesion segmentation, feature extraction, and clustering. Radiol Bras. 2026,59:e20260005

INTRODUCTION

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, accounting for approximately 1.8 million deaths annually[1]. Early detection is critical to improving survival rates, and chest computed tomography (CT) plays a pivotal role in lung cancer screening and diagnosis, particularly among high-risk populations[2].

Radiomics has emerged as a powerful approach that extracts quantitative features from standard medical images. These features, derived from pixel intensity distributions and spatial relationships, can capture subtle textural, morphological, and structural properties of tissue, which may correlate with histological and genetic tumor characteristics[3,4]. Radiomic features are typically categorized as first-order statistics (e.g., histogram-based features), second-order texture metrics (e.g., gray level co-occurrence matrix [GLCM] and gray level run-length matrix [GLRLM]), and higher-order features, which incorporate transformations such as wavelets or filters to capture multiscale patterns.

The central premise of radiomics is that medical images contain hidden biological information not always apparent in visual interpretation. By combining radiomic data with clinical or molecular information, it is possible to generate predictive models that may support clinical decision-making, as well as facilitating diagnosis, prognostication, and treatment planning[5,6].

In the context of lung cancer, chest CT is particularly well-suited for radiomic analysis, as it allows detailed visualization of pulmonary nodules and masses. However, the success of radiomics depends heavily on a standardized, reproducible workflow, including consistent acquisition parameters, robust segmentation, and careful feature selection. Variability in any of these steps can compromise the stability and generalizability of radiomic models[7,8].

The aim of this study was to present a pilot project that explores the best practices in radiomics applied to chest CT for the detection and classification of lung lesions, including benign, primary malignant, and secondary malignant (metastatic) lesions. This work highlights the importance of a structured workflow, interdisciplinary collaboration, and explainability in exploratory radiomics analyses.

MATERIALS AND METHODS

This retrospective study was approved by the local institutional review board (Reference no. 78384324.5.0000.5461). Because the data were collected retrospectively, the requirement for informed consent was waived. Chest CT images were obtained from oncologic patients under investigation for a new finding of a pulmonary nodule during follow-up CT. All lesions were submitted to CT-guided biopsy. A total of 32 pulmonary lesions were included in this pilot study, comprising 11 benign lesions and 21 malignant lesions. Malignant lesions were further classified as either primary lung tumors or metastatic lesions. No treatment or clinical outcome was considered for classification. The dataset, including diagnoses and subclassifications, is presented in Table 1.

Chest CT examinations were performed using multimodal scanners equipped with proprietary iterative reconstruction algorithms: Biograph 128 (Siemens Healthineers, Erlangen, Germany); Somatom Definition AS+ (Siemens Healthineers); Somaton Definition AS (Siemens Healthineers); Somatom Definition Edge (Siemens Healthineers); and Discovery MI (GE Healthcare, Milwaukee, WI, USA). Tube voltage and current modulation were adjusted according to vendor specific technology; in the sample as a whole, the mean value was 120 kVp (range, 90–140 kVp). All images were acquired in helical mode during a full-inspiration breath-hold, with the following parameters: slice thickness, 1.0–2.0 mm with matching reconstruction intervals; 512×512 matrix; and a field of view adjusted to the thoracic cavity. To minimize variability during lesion segmentation

Table 1—Dataset composition: histopathological diagnosis and subclassification of lung lesions.

Parameter	(N = 32)
Sex, n	
Male	12
Female	20
Age (years), mean (range)	65.7 (41–91)
Benign lesions,* n	11
Primary malignant lesions,† n	14
Secondary malignant lesions,‡ n	7

*Organizing necrotic tissue with chronic granulomatous inflammation, aspergillus, necrotic inflammatory nodule, or a chronic nodular inflammatory process with fibrosis and lymphoplasmacytic infiltrate.
†Lung adenocarcinoma or poorly differentiated non-small cell lung cancer.
‡Metastasis from Hodgkin lymphoma, renal cancer, leiomyosarcoma, breast cancer, or melanoma.

and radiomic feature extraction, all scans were acquired through the use of unenhanced chest CT protocols and the images were reconstructed using standard soft-tissue and lung kernels , thus allowing adequate evaluation of parenchymal texture and lesion morphology.

We segmented the lesions by using the software syngo.via MM Radiomics Frontier (Siemens Healthineers), employing a semi-automated approach followed by manual refinement. All segmentations were performed by an experienced thoracic radiologist, to ensure anatomical accuracy and consistency. Only unenhanced axial image reconstructions were used for segmentation and radiomics analysis.

Following segmentation, a comprehensive set of > 900 radiomic features was extracted for each lesion using the syngo.via MM Radiomics Frontier software, which makes use of the PyRadiomics package and offers the option of feature computation/extraction, as well as allowing the results to be exported in the comma-separated values format [9]. The extracted features encompassed first-order statistics, shape descriptors, and high-order texture metrics derived from matrices such as GLRLM, GLCM, gray level size-zone matrix, and wavelet transformations. Feature selection was also used to reduce not-a-number values in order to explore redundancy and correlation among radiomic features, we conducted a pairwise Pearson correlation analysis, visualized on a clustered heatmap. After feature selection, 26 variables remained and were used for the final analysis.

The analysis was performed using custom Python scripts implemented through the feature analysis module, with statistical packages for analytics. The analytical pipeline included the following steps:

Feature normalization was conducted with StandardScaler to standardize all features to have zero mean and unit variance, thus ensuring equal contribution during clustering.

Dimensionality reduction was performed via principal component analysis (PCA) to reduce feature space complexity and allow two-dimensional visualization of cluster separation.

Unsupervised clustering was carried out by using the k-means algorithm, with the number of clusters (k) varying from 2 to 5 to explore different grouping structures.

Cluster validity was assessed by calculating the silhouette score, a metric that quantifies how well each sample fits within its assigned cluster compared with other clusters. Higher scores (closer to 1) indicate clustering structures that are more cohesive and well-defined.

Supervised models with clustered data were tested to evaluate the feasibility of results and the potential of the study to validate the workflow using radiomics analysis.

The Radiomics workflow is illustrated and summarized in the Figures 1 and 2.

To evaluate the classification performance of the models, four supervised machine learning algorithms were trained and tested: Random Forest, Logistic Regression, Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost). The receiver operating characteristic (ROC) curves were plotted with the predicted probabilities for the positive class, and the areas under the curve (AUCs) were used as the primary metric for model comparison. All models were evaluated on the same test dataset using the same feature set. This approach allows for a consistent assessment of the ability of each model to distinguish between classes, regardless of the decision thresholds.

RESULTS

Radiomics extraction provides a large number of data, and feature selection is a crucial step in the radiomics workflow. The goal is to reduce irrelevant or redundant features and retain those that are relevant and useful, reducing dimensionality so that analysis process can be performed more efficiently, decreasing the resource demand, work time, and risk of overfitting[10]. To achieve that reduction, we conducted a pairwise Pearson correlation analysis to explore redundancy and correlation among radiomic features, visualized through a clustered heatmap (Figure 3). The matrix displays the correlation coefficients between features. Hierarchical clustering was applied to groups of highly correlated features, helping to identify patterns and potential collinearities in the dataset. This step is crucial for feature selection, because highly correlated variables may contribute re-

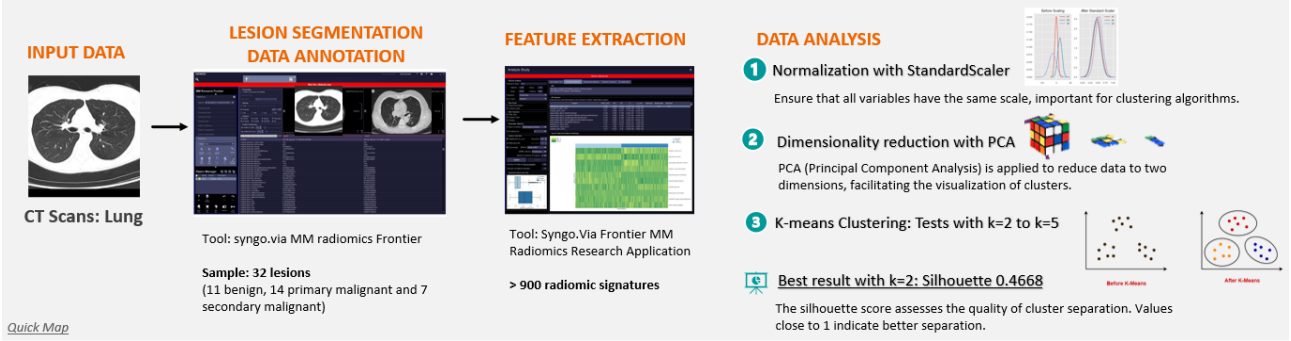

Figure 1. Quick map project, created by the authors.

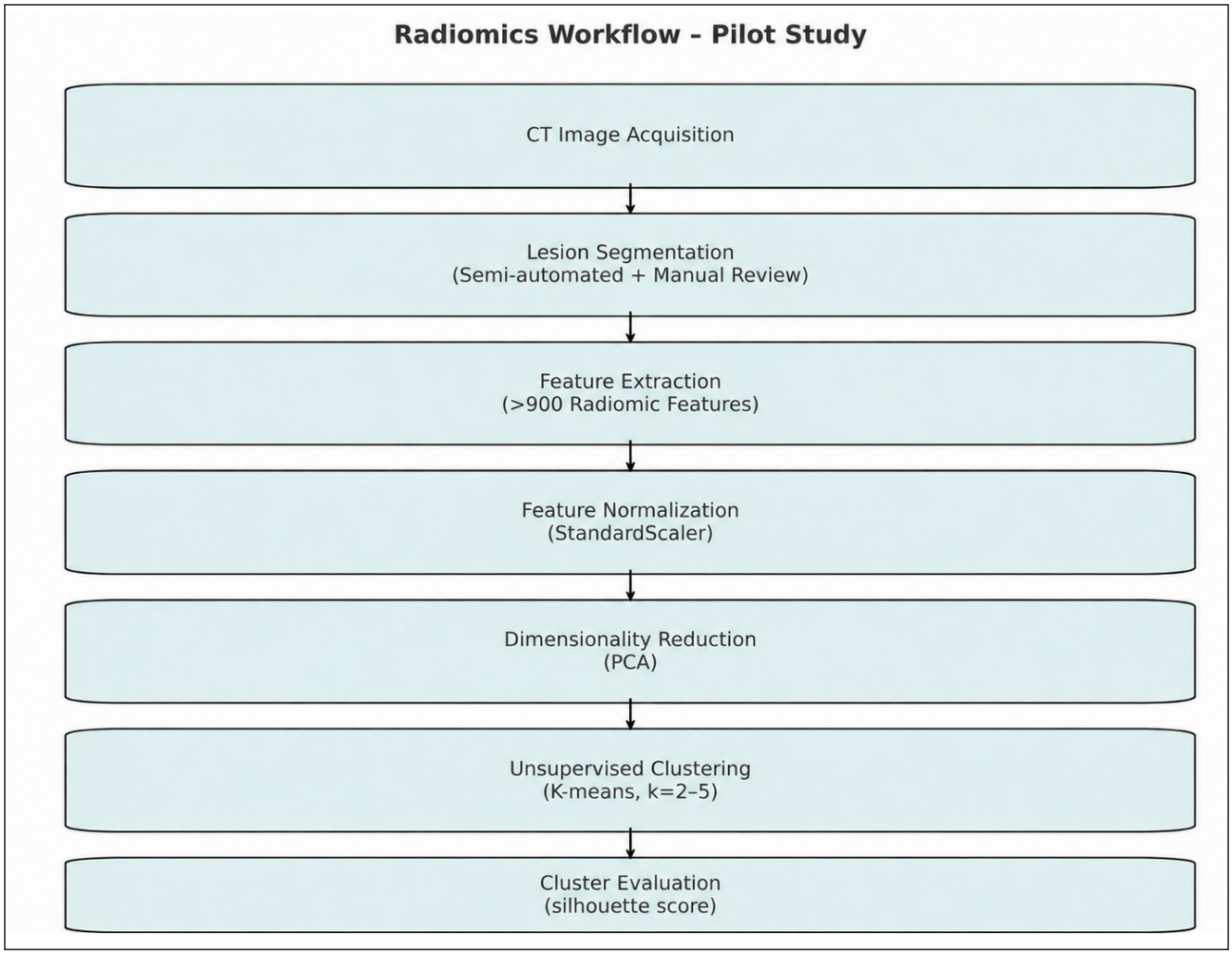

Figure 2. Workflow pilot study, created by the authors.

dundant information to downstream models and affect their generalizability.

In the differentiation between benign and malignant lesions, the analysis highlighted several feature clusters with strong internal correlations, suggesting opportunities for dimensionality reduction or selection of representative features within each cluster.

After this first feature selection, k-means analysis was applied to data to classify into groups. Among the tested configurations, the clustering solution with k = 2 produced the highest silhouette score (0.4668), indicating the most distinct and internally coherent separation of data points. This suggests that the dataset naturally tends to form two well-defined clusters. In contrast,

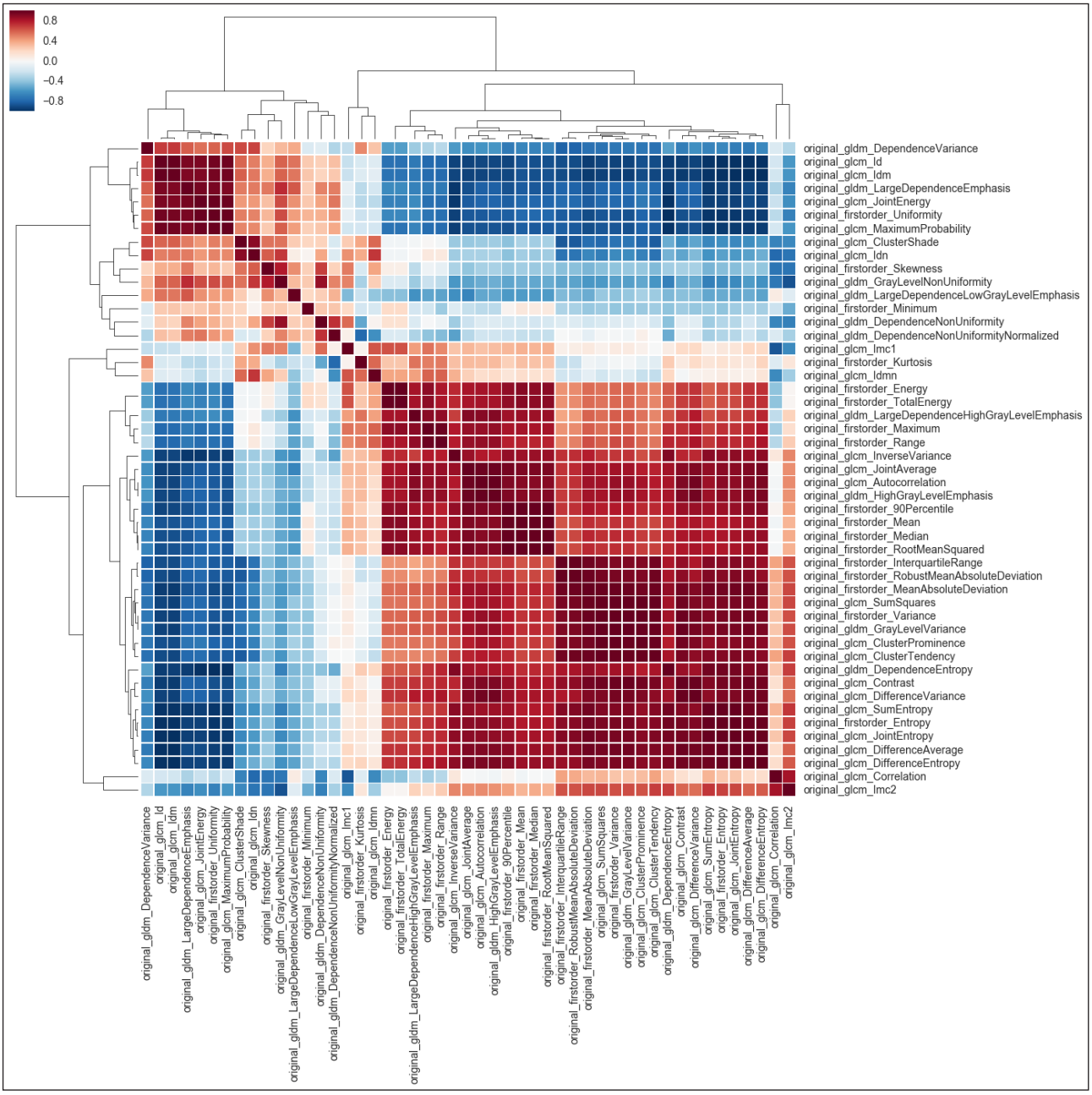

Figure 3. Clustered heatmap of pairwise Pearson correlation coefficients among radiomic features. Red indicates a strong positive correlation, and blue indicates a strong negative correlation. Hierarchical clustering was applied to group similar features, enabling the identification of redundant or collinear variables for potential feature selection.

lower silhouette scores were observed for other cluster configurations: k = 3 (0.2763), k = 4 (0.2889), and k = 5 (0.3860), reflecting less cohesive groupings and overlapping between clusters. The PCA projections supported those findings, showing visually interpretable cluster separation when k = 2, which appeared to align with the binary classification of lesions as benign or malignant.

In the subgroup analysis of malignant lesions, no statistically significant differences were observed between primary and metastatic malignant tumors. The unsupervised k-means clustering with k = 2, which yielded the highest silhouette score (0.4668), grouped primary and secondary malignant lesions predominantly within the same cluster, indicating a high degree of similarity between these subtypes. Consistently, the PCA projections demonstrated substantial overlap between primary and metastatic malignant lesions, with no clear separation into distinct clusters. These findings are further sup-

ported by the moderate performance of the supervised classification models, with the best model (XGBoost) achieving an AUC of 0.72, suggesting limited discriminatory ability between malignant subtypes within the constraints of the current feature set and the small sample size (n = 32).

The ROC curve analysis (Figure 4) compares the performance of four classification models: Random Forest, Logistic Regression, SVM, and XGBoost. Among them, the XGBoost model achieved the best performance, with an AUC of 0.72, followed by the SVM model, with an AUC of 0.69. In contrast, the Random Forest and Logistic Regression models demonstrated poor discriminatory power, with AUCs of 0.49 and 0.32, respectively, both below the level of chance (AUC = 0.5).

Preliminary clinical interpretation of clusters

The two clusters identified through unsupervised k-means clustering (k = 2) demonstrated a clinically meaningful pattern when compared to the biopsy-confirmed classification of lesions.

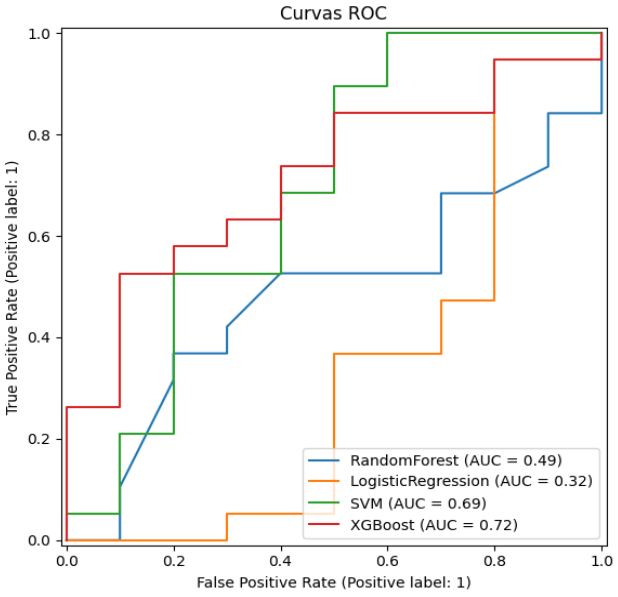

Figure 4. ROC curves for the four classification models tested. The AUC is shown in parentheses for each model. The XGBoost model outperformed the others, achieving an AUC of 0.72, followed by the SVM model (AUC = 0.69), whereas the Random Forest and Logistic Regression models performed near or below random classification levels.

Cluster 1 was predominantly composed of benign lesions, typically presenting more homogeneous texture, smoother borders, and lower intensity variation, as captured by lower first-order entropy and less complex texture metrics (e.g., GLCM contrast and GLRLM run entropy).

Cluster 2 was composed mainly of malignant lesions, including primary and secondary tumors. These lesions exhibited higher heterogeneity, irregular shapes, and greater variation in pixel intensity and texture, consistent with biological aggressiveness and structural disorganization, characteristics well reflected in higher-order radiomic features.

DISCUSSION

Our results suggest that radiomics features alone may be sufficient to distinguish between benign and malignant pulmonary lesions, even in the absence of clinical or molecular data. However, the cluster containing malignant lesions included primary and secondary (metastatic) tumors, which were not consistently separated by the current feature set or clustering method, suggesting that additional stratification strategies may be necessary to resolve finer subclass differences in future analyses.

From a clinical perspective, our findings indicate that radiomics could play a role as a decision-support tool for early triage of lung lesions detected on CT scans, helping to flag cases that require further diagnostic workup. Nevertheless, the inability to distinguish between different malignant subtypes at this stage underscores the need for the following[9]:

Integration of clinical metadata, such as patient medical history/smoking status and overall disease severity (e.g., tumor–node–metastasis staging)

Inclusion of molecular markers (tumor histology)

Refinement of feature selection strategies focused on tumor biology

This pilot study demonstrated the practical feasibility of implementing a structured radiomics workflow using chest CT data within a clinical setting. Beyond technical validation, it also underscored critical elements that influence the robustness and reproducibility of radiomics analyses, particularly the need for standardized acquisition parameters and consistent segmentation practices. These findings are aligned with data in the current literature, supporting the idea that even subtle variations in CT protocol (e.g., slice thickness, reconstruction filters) can have a significant impact on feature stability and model performance[11,12].

The clustering results when k = 2 was applied revealed that the extracted features captured relevant differences between benign and malignant lesions. This supports the underlying premise of radiomics: that imaging contains quantifiable patterns reflecting biological behavior, which may not be apparent through visual inspection alone. The concordance between the clustering structure and the ground truth classification suggests that radiomics

descriptors hold significant discriminatory power and may serve as the foundation for future decision-support tools.

Dimensionality reduction methods such as PCA are essential for selecting a specific number of features for analysis, because they help identify redundant or highly correlated variables that contribute similarly to the outcome. These methods also facilitate data processing by reducing the computational resources required. However, a key drawback is the potential loss of model explainability, which may be a significant barrier to clinical adoption.

Several key challenges were identified throughout the project:

Workflow—Although semi-automated segmentation tools were used for this study, a manual review, which is time-intensive and limits scalability, is still required for the ultimate accuracy of segmentation. This highlights the importance of developing or integrating an automated segmentation tool validated for clinical use[13,16].

Protocol harmonization—To ensure reproducibility, adherence to standardized acquisition parameters is essential. Inconsistent voxel sizes, slice thicknesses, or reconstruction kernels introduce variability that can obscure biologically relevant patterns. A resampling tool can be used to ensure isotropic voxels with equal distances between neighboring voxels in all directions. This step is crucial to ensure that the data extracted from the images is equivalent across the different types of studies analyzed[12–14].

Multidisciplinary collaboration—Effective radiomics research demands the integration of expertise from radiology, data science, and imaging physics. This interdisciplinary effort was crucial for designing the analytical pipeline, selecting appropriate features, and interpreting the results in a clinically meaningful context[15].

The dataset used in this pilot study comprised only 32 cases, which precluded a statistically reliable division into independent training and testing subsets. Considering the limited sample size and the exploratory aim of the study, we employed an unsupervised clustering approach. This method does not require predefined training or validation partitions, given that the objective was not predictive modeling but the identification of intrinsic data patterns. Despite these constraints, the findings demonstrate the potential of radiomics features to discriminate between benign and malignant lung lesions. In addition, this pilot study contributed to ongoing discussions on best practices for CT-based radiomics, particularly in the field of thoracic oncology[16,17].

Key methodological pillars include the adoption of standardized image acquisition and reconstruction protocols, ensuring image harmonization across different vendors. In addition, the careful selection of appropriate techniques for feature selection and dimensionality reduction is critical to maintain a balance between efficient data processing and model interpretability[16]. Overall, these insights serve as a foundation for expanding this work into larger, multicentric studies. They also emphasize the value of radiomics not only as a research tool but as a potential enabler of precision imaging and data-driven decision-making in routine clinical care.

Finally, enriching the analysis with clinical variables like age, sex, and medical history may provide a more comprehensive model that better reflects real-world diagnostic decision-making.

CONCLUSION

This pilot study validated the feasibility of deploying a radiomics pipeline in the clinical CT environment, specifically for the analysis of lung lesions. Through semi-automated segmentation feature extraction and data analysis, we demonstrated that radiomics descriptors can effectively differentiate benign from malignant lesions using unsupervised clustering techniques. The results underscore the potential of radiomics as a noninvasive, quantitative tool to support clinical decision-making, particularly in the early detection and characterization of lung malignancies. However, the study also highlighted key operational challenges, including the need for workflow automation, strict protocol standardization, and integration of multidisciplinary expertise, that must be addressed to ensure scalability and reproducibility.

Future research should prioritize multicenter studies to expand the dataset and ensure broader population diversity, thereby enhancing the generalizability of the findings. Integrating clinical and molecular data, as well as refining feature selection strategies, will further improve the discriminatory performance of radiomics models, particularly among malignant subtypes. Collectively, this work contributes to the establishment of a structured framework for CT-based radiomics in thoracic oncology and provides a foundation for its future translational application in personalized medicine.

REFERENCES

1. International Agency for Research on Cancer. Global Cancer Observatory: Cancer Today [homepage on the Internet]. Lyon: International Agency for Research on Cancer; 2024 [cited 2025 Jun 24]. Available from: https://gco.iarc.who.int/today/
2. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395-409.
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563-77.
4. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14:749-62.

5. Zhang Y, Huang W, Jiao H, et al. PET radiomics in lung cancer: advances and translational challenges. EJNMMI Phys. 2024;11:81.
6. Ibrahim A, Primakov S, Beuque M, et al. Radiomics for precision medicine: current challenges, future prospects, and the proposal of a new framework. Methods. 2021;188:20-9.
7. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295:328-38.
8. Traverso A, Wee L, Dekker A, et al. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys. 2018;102:1143-58.
9. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77:e104-7.
10. Zhang W, Guo Y, Jin Q. Radiomics and its feature selection: a review. Symmetry (Basel). 2023;15:1834.
11. Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. Sci Rep. 2017;7:588.
12. Mali SA, Ibrahim A, Woodruff HC, et al. Making radiomics more reproducible across scanner and imaging protocol variations: a review of harmonization methods. J Pers Med. 2021;11:842.
13. Volpe S, Isaksson LJ, Zaffaroni M, et al. Impact of image filtering and assessment of volume-confounding effects on CT radiomic features and derived survival models in non-small cell lung cancer. Transl Lung Cancer Res. 2022;11:2452-63.
14. Ma Z, Zhang J, Liu X, et al. Comparative analysis of repeatability in CT radiomics and dosiomics features under image perturbation: a study in cervical cancer patients. Cancers (Basel). 2024;16:2872.
15. Shur JD, Doran SJ, Kumar S, et al. Radiomics in oncology: a practical guide. Radiographics. 2021;41:1717-32.
16. Avery E, Sanelli PC, Aboian M, et al. Radiomics: a primer on processing workflow and analysis. Semin Ultrasound CT MR. 2022;43:142-6.
17. Riberdy V, Guida A, Rioux J, et al. Radiomics in preclinical imaging research: methods, challenges and opportunities. NPJ Imaging. 2025;3:45.