Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated Whole-Lung CT Morphometric Phenotyping with Novel Shape Indices: A Pilot Study in Lung Squamous Cell Carcinoma

Emmanuel Joy

dremmjoy@outlook.com

Sri Krishna College of Engineering and Technology

Boney George Thomas

Karunya University

Suganthi Evangeline C

Sri Eshwar College of Engineering

Jenice Aroma R

Thiagarajar College of Engineering

Reshma V.K

Sri Krishna College of Engineering and Technology

Research Article

Keywords: Lung cancer, computer tomography, radiomics, lung segmentation, shape-based features, machine learning, biomarkers, TCGA-LUSC

Posted Date: August 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10585409/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Computed tomography (CT) lung cancer classification plays a crucial role in earlier diagnosis and thereby to offer better treatment to thep. Although radiomics makes it possible to analyse medical images quantitatively, most research only examines the tumour region of interest rather than the overall lung architecture. The factors like Atelectasis, comorbid airway disease, and tumor mass effect can even change the shape of the entire lung and in order to characterize global lung morphology beyond radiomic shape parameters, this pilot study is focused on two composite shape descriptors, the Shape Complexity Index (SCI) and the Axis Disparity Index (ADI), to quantify global lung morphology beyond radiomic shape features.

Results

A fully automated pipeline, integrating DICOM-to-NIfTI conversion, lobe-based lung segmentation using TotalSegmentator, and radiomic feature extraction with PyRadiomics, was implemented on CT series from eight patients in the TCGA-LUSC collection who satisfied predefined image quality criteria. In total, 107 standard radiomic features plus SCI and ADI were computed for each patient. Spearman correlation analysis showed that SCI and ADI captured morphological information that was largely non-redundant with conventional shape descriptors. Patients grouped by median SCI differed substantially in axis length and elongation, with a large effect size for SCI itself (Cohen's d = 1.18). Unsupervised K-means clustering on the shape-feature set, including SCI and ADI, identified two morphologically distinct phenotypes that separated clearly in principal component space. In leave-one-out cross-validated (LOOC) logistic regression using provisional, non-clinical outcome labels, a model combining conventional features with SCI and ADI achieved the same accuracy (0.625) and area under the curve (0.812) as a conventional-features-only baseline. This indicates that the additional biomarkers did not reduce classification performance in this small sample.

Conclusions

This feasibility study shows that an automated whole-lung morphometric pipeline is technically achievable and that SCI and ADI provide structural information that is largely distinct from standard radiomic shape features, without compromising classifier performance. Given the very small, single-cohort sample and provisional outcome labels, these findings are preliminary. Validation in larger, clinically annotated, multi-institutional cohorts is required before any diagnostic or prognostic claims can be made.

Introduction

Lung cancer is one of the leading causes of the top causes of cancer-related deaths worldwide, causing a major global health concern [1]. Computed tomography (CT) is the standard imaging modality for diagnosing and treating lung cancer that could offer 3D visualization of thoracic structures. In traditional practices of investigation, these image interpretations are highly determined by the radiologists through qualitative assessment and increases the risk of manual errors during such diagnosis and reporting. The vision-based assessment approaches generally fail to utilize the quantitative information available in CT images. Hence the need for intelligent models for medical image interpretation is highly evident and Radiomics addresses these limitations by transforming CT images into high-dimensional, mineable data. The use of quantitative image descriptors such as intensity, texture, and shape can reveal patterns that are not discernible to the human eye, enabling the development of predictive models. In general, the Radiomic signatures are associated with clinical outcomes, tumor genomics, and treatment responsiveness across various cancers [2]. Most radiomic studies of lung cancer define the disease using regions of interest limited to the total net tumor volume and its immediate surroundings [3]. Although this tumor-centric approach is informative, it offers only a partial perspective on the condition. Tumors can interact with and reshape the surrounding pulmonary environment through mechanisms such as airway obstruction, atelectasis, or mass effect. Additionally, comorbid conditions such as chronic obstructive pulmonary disease (COPD) or pulmonary fibrosis could independently alter the overall lung shape and structure. From recent investigations, it is found that features derived from the entire lung, rather than solely the tumor, may possess prognostic value independent of tumor-level descriptors [4]. Consequently, morphological differences across the whole lung, reflecting both tumor effects and underlying pulmonary status, represent a potentially underutilized source of diagnostic information. The significance of Quantitative assessment of whole-lung morphology remains with the improved comprehensive characterization of disease status.

Related work

Radiomics, in lung cancer studies, which is widely accepted in quantitative medical image analyses has been used to predict PD-L1 expression status as demonstrated by Lim et al. [5] and even used to identify histopathological subtypes, as reported by Choi et al. [6]. Upon studying the prior works on CT-based lung cancer radiomics, automated lung segmentation, and geometric shape features in quantitative imaging, it is clearly found that the reproducibility of CT-based radiomic features is the common limitation, which can even be sensitive to scanner vendor, slice thickness, reconstruction kernel, and segmentation methods [7, 8]. Initiatives like the Image Biomarker Standardization Initiative (IBSI) aim to handle this by promoting greater standardization methods. Belfiore et al. [9] mentioned the importance of standardization and robustness in radiomic analysis. The proposed work aims at resolving this limitation through using a fully automated, standardized processing pipeline to reduce interpreter person variability. Manual segmentation process takes much time and is generally biased based on the expert knowledge. Most of the early methods relied on image processing techniques such as thresholding, region growing and morphological operations. These methods performed well for normal lung conditions but were not robust in the presence of larger tumors, pleural effusion etc. Deep Learning

segmentation methods especially CNNs built on U-Net architectures improved the robustness significantly. Amini and Klein [10] evaluated TotalSegmentator, a model trained on a large and diverse dataset to segment 104 anatomical structures including lung lobes. More related works have expanded such models to resolve inter lobar fissures for more detailed analysis, as confirmed by Fufin et al. [11].

Shape-based radiomic features describe the geometry of a region of interest independently of internal intensity patterns. Commonly reported 3D shape descriptors include volume, surface area, sphericity, and compactness, together with elongation and flatness, which describe how a shape is distributed along its principal axes and are derived from the eigenvalues $(\lambda_1 \geq \lambda_2 \geq \lambda_3$ of the covariance matrix of voxel coordinates:

$$\text{Elongation} = \sqrt{\lambda_2/\lambda_1}$$

1

$$\text{Flatness} = \sqrt{\lambda_3/\lambda_1}$$

2

By convention, both measures range from 0 to 1, and approach 1 as a shape reaches a spherical symmetry. Though useful, individual descriptors such as these can be limited to characterize complex shapes separately. Nissar and Mir [12] demonstrated the prognostic value of shape-based radiomic features in CT lung cancer analysis, and Raptis et al. [13] showed that shape features contribute much to automated lung cancer detection with the additional radiomic information. Vandaele et al. [14] demonstrated that topological data analysis to capture complementary structural information and improve tumor detection. Rather than proposing replacement descriptors, the present study combines multiple descriptors into a single higher-level entity of descriptors.

Machine learning is the last stage of a typical radiomics workflow, where the extracted feature set is used to build predictive models. Genç and Yıldırım [15] employed ML techniques on CT-derived radiomic features to distinguish round pneumonia from primary lung cancer, and Corso et al. [16] evaluated classification methods commonly used in radiomic analyses. For classification or regression tasks, supervised techniques including logistic regression, support vector machines, random forests, gradient boosting, and deep neural networks are used. Unsupervised methods such as K-means clustering and principal component analysis (PCA) are used for patient grouping without pre-defined labels. Validation techniques, such as the leave-one-out cross-validation (LOOCV) employed in this study, are crucial to reduce overfitting because radiomic feature dimensionality is usually high in relation to available sample sizes. Leng et al. [17] proposed a hybrid framework showing the potential of multi-modal fusion of features for lung cancer classification.

Research gap and contributions

Our literature search indicates a significant research gap in attention to biomarkers that reflect the global morphology of the lungs. The current predominant paradigm has focused on tumors and traditional single-metric shape measurements, which offer considerable areas for new concepts. This has inspired us to develop a framework to test the hypothesis whether the overall shape of the lung provides diagnostic and prognostic value beyond what is currently available.

Given the very small, single-institution sample available (n = 8), this work is presented as a feasibility (pilot) study intended to establish proof of concept than to support any diagnostic or prognostic claim. The contributions of this study are threefold:

1. An automated framework that processes raw DICOM data from the TCGA-LUSC collection through deep-learning-based lung segmentation and radiomic feature extraction without manual intervention.
2. Proposes two novel structural biomarkers, the Shape Complexity Index (SCI) and the Axis Disparity Index (ADI), to summarize higher-order geometric properties of the lung parenchyma.
3. Evaluates these biomarkers through statistical correlation analysis, unsupervised phenotype discovery, and supervised classification with LOOCV.

The methodology used in this pilot study framework is explained in the next section.

Methods

Data Source and Cohort selection

CT imaging data were obtained from the publicly available Cancer Genome Atlas Lung Squamous Cell Carcinoma (TCGA-LUSC) collection, hosted by The Cancer Imaging Archive (TCIA) [18]. This dataset was selected as it represents a clinically and biologically distinct subset of non-small cell lung cancer.

An automated Python script was used to query the entire image directory and identify CT series suitable for three-dimensional analysis. Inclusion criteria required each series to meet the following conditions: (i) confirmation of imaging modality, (ii) adequate axial (Z-axis) coverage, and (iii) presence of valid geometric information. Applying these criteria resulted in a final analysis cohort of eight patients. The complete workflow, from CT image acquisition to whole-lung morphological characterization, is presented in Fig. 1.

Preprocessing

For each patient series, DICOM (.dcm) files were processed using pydicom to construct a 3D NumPy array by stacking all 2D pixel arrays in ascending order by their Z coordinate, as indicated by ImagePositionPatient. Raw pixel values were then converted to Hounsfield Units (HU) using the linear transformation specified in the DICOM header.

$$\mathbf{HU} = \text{PixelValue} \times \text{RescaleSlope} + \text{RescaleIntercept} \quad (3)$$

The Hounsfield Unit (HU) scale provides a standardized measurement of radiodensity, facilitating the comparison of voxel intensities across different scans. On this scale, − 1000 HU represents air, 0 HU represents water, and higher values indicate increasing tissue density. The resulting HU volume was converted to NIfTI format (. nii.gz) using SimpleITK. Anisotropic voxel spacing, including slice thickness and in-plane pixel spacing, was retained from the DICOM header to ensure that all shape and volume calculations were conducted in millimeters.

Lung segmentation and mask post-processing

Lung parenchyma was segmented using TotalSegmentator [10], a deep-learning model built on a U-Net-style architecture. Rather than segmenting the lung as a single structure, a lobe-based strategy was used to capture the anatomy of each of the five lung lobes (left upper, left lower, right upper, right middle, and right lower). TotalSegmentator was run in its faster inference mode to reduce computation time. This mode trades some output resolution for speed and has previously been reported to be less accurate when segmenting parenchyma affected by large masses, which we note as a limitation of the present pipeline. The five lobar masks were combined with a logical OR operation into a single whole-lung binary mask.

Since a segmentation mask generated by a deep-learning model does not automatically inherit the same spatial metadata (origin, spacing, direction) as the source image, the merged lung mask's spatial information was explicitly copied from the source CT NIfTI volume using SimpleITK's CopyInformation method. This step ensures that the mask is spatially aligned with the image and that PyRadiomics computes shape and intensity features in the correct physical coordinate system.

Radiomic feature extraction

Feature extraction was performed using PyRadiomics (v.3.0.1), with the same setup for each configuration. A total of 107 standard features were extracted per patient which are summarized by class in Table 1.

Table 1
Extracted radiomic feature classes and examples

Class	Count	Description and examples
First Order	18	Statistics of voxel intensity distribution (mean, skewness, kurtosis)
Shape	14	3D geometry of the ROI (volume, sphericity, elongation, flatness)
GLCM	24	Grey Level Co-occurrence Matrix features (contrast, correlation, energy)
GLRLM	16	Grey Level Run Length Matrix features (short run emphasis, long run emphasis)
GLSZM	16	Grey Level Size Zone Matrix features (small area emphasis, large area emphasis)
NGTDM	5	Neighbouring Grey Tone Difference Matrix features (coarseness, busyness)
GLDM	14	Grey Level Dependence Matrix features (dependence variance)
Total	**107**	

Proposed shape-based morphological biomarkers

Two composite indices were defined to summarize global lung morphology beyond the individual shape descriptors described in the earlier sections. A conceptual illustration of how axis-length relationships influence SCI and ADI is presented in Fig. 2.

Shape Complexity Index (SCI)

SCI combines elongation, flatness, and the ratio of major-to-minor axis length into a single index intended to summarize the overall geometric irregularity of the lung

$$\varvec{SCI} = Elongation \times Flatness \times \left(\frac{MajorAxisLength}{MinorAxisLength} \right)$$

4

Elongation and flatness decrease while the major-to-minor axis ratio increases as a shape departs from spherical symmetry. SCI, as defined here, therefore combines terms that do not all move in the same direction with increasing anisotropy. We evaluated SCI empirically against conventional features and cluster structure.

Axis Disparity Index (ADI)

ADI quantifies axis-length anisotropy by comparing the longest and shortest lung dimensions, normalized to the longest axis

$$\varvec{A}\varvec{D}\varvec{I} = \frac{MajorAxisLength - LeastAxisLength}{MajorAxisLength}$$

5

An ADI of zero corresponds to a perfectly symmetric organ with equal axis lengths. ADI increases as the disparity between the longest and shortest axes grows, i.e., as the lung becomes more elongated or flattened relative to a sphere.

Experimental Design

Feature preprocessing and normalization

Before constructing the models, the feature matrices were preprocessed. The patient identifiers and non-informative diagnostic metadata generated by PyRadiomics were removed, near-zero-variance features were excluded, as they provided little discriminative value and invalid values were replaced.

Statistical validation

To assess redundancy between the two novel biomarkers and existing radiomic features, Spearman's rank correlation (ρ) was computed between SCI, ADI and each conventional feature and correlations near zero were taken as evidence of low redundancy. Patients were also divided by median SCI into complexity groups, and differences in shape features were compared using the Mann Whitney U test, with effect size quantified by Cohen's d.

Unsupervised phenotype discovery

K-means clustering was applied to the Z-score-normalized shape-feature set, including standard geometric features together with SCI and ADI, to identify natural lung-morphology phenotypes. The optimal number of clusters (k) was selected using the elbow method applied to the within-cluster sum of squares, and the resulting clusters were visualized using principal component analysis (PCA), retaining the first two principal components.

Supervised classification and evaluation

Because no clinically derived binary outcome label was available for this small pilot cohort, synthetic, non-clinical binary labels were assigned to the 8 patients to permit a controlled methodological comparison. The results based on these labels reflect the technical behaviour of the pipeline and should not be interpreted as evidence of clinical predictive validity. Two logistic regression models were compared under leave-one-out cross-validation (LOOCV): a baseline model using only conventional radiomic features, and a combined model adding SCI and ADI to the conventional feature set. Performance was summarized using accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, and specificity, and the performance of the two models was compared using a Wilcoxon signed-rank test.

Results

Cohort characteristics and feature extraction

Automated processing was completed successfully for all 8 patients in the final analysis cohort. For each patient, 107 standard radiomic features as in Table 1 plus the two novel morphological biomarkers (SCI, ADI) were extracted, yielding a final feature matrix of 8 patients × 109 features (including the proposed features) that served as the basis for the statistical and machine-learning analyses reported below.

Statistical validation of novel biomarkers

Spearman correlation analysis was performed to determine whether SCI and ADI represent information distinct from traditional radiomic descriptors. Both biomarkers showed low mean absolute correlations with the conventional radiomic feature set. SCI's correlation with axis-based geometric features was moderate, as expected given that axis length is one of its inputs, but none of its correlations with other radiomic features met conventional thresholds for redundancy. ADI likewise showed low correlations with other radiomic descriptors. These results indicate that SCI and ADI capture morphological information that is largely distinct from conventional radiomic shape descriptors in this cohort.

Patients were divided into low- and high-complexity groups based on the cohort median SCI. The Mann–Whitney U test identified significant between-group differences in Major Axis Length and Elongation. For SCI the effect size between groups was large (Cohen's d = 1.18), indicating a large difference in morphological complexity between the two groups and supporting SCI as a discriminative feature for structural lung phenotyping in this sample.

Unsupervised discovery of morphological phenotypes

K-means clustering on the shape-feature set revealed latent morphological structure within the cohort. The elbow method identified two clusters as optimal thus producing two distinct lung phenotypes, summarized in Table 2.

Table 2
Characteristics of the two unsupervised morphological phenotypes

Mean feature value (Z-scored)	Phenotype 0	Phenotype 1
Shape Complexity Index (SCI)	−4.39	−8.32
Original shape – Least Axis Length	0.52	−1.56
Original shape – Major Axis Length	−0.43	1.29
Original shape – Minor Axis Length	0.44	−1.31
Original shape – Elongation	0.39	−1.18
Original shape – Flatness	0.42	−1.25

In this cohort, a more negative SCI value therefore corresponds to the more anisotropic, elongated phenotype 1. This separation is visually supported by the PCA plot in Fig. 3. The two phenotypes identified by K-means clustering are shown in different colours. Each patient was assigned to a spatially distinct region of this 2D reduced feature space, with clear between-cluster separation, suggesting that global lung morphology carries structural information sufficient to distinguish subgroups of patients in this sample. Phenotype-wise values of Major Axis Length and Shape Complexity Index used for the scatter-plot analysis (in Fig. 2) is given in Table 3.

Table 3
Phenotype-wise values of Major Axis Length and Shape Complexity Index

Phenotype	Major Axis Length (PC-x)	Shape Complexity Index (PC-y)
0	−1.71388	−26.491776
0	−0.35481	−0.007272
0	−0.57141	0.069806
0	−0.20239	−0.000779
0	−0.20402	−0.030378
0	0.460042	0.127865
1	0.596442	−0.411563
1	1.990035	−16.236807

Supervised classification performance

The morphological biomarkers were assessed for their contribution to predictive performance using leave-one-out cross-validation with the synthetic, non-clinical labels described in Section 2.6.4. Two models were compared: a baseline model using only standard radiomic features, and a combined model adding SCI and ADI. Results are shown in Table 4.

Table 4. Performance of the baseline and combined supervised classification models

Performance metric	Baseline model	Combined model
Accuracy	0.625	0.625
AUC	0.812	0.812

Spearman rank correlation analysis was also performed between the proposed biomarkers and the conventional radiomic shape descriptors given in Table 5.

Table 5
Spearman correlation between the proposed novel shape biomarkers and conventional

Radiomic shape descriptor	Spearman correlation (ρ)
Axis Ratio	0.761905
Shape Complexity Index (SCI)	0.761905
Original shape – Voxel Volume	0.452381
Original shape – Minor Axis Length	0.428571
Original shape – Least Axis Length	0.357143
Axis Disparity Index (ADI)	0.071429
Normalized Lung Size	0.047619
Original shape – Major Axis Length	−0.023810
Original shape – Flatness	−0.071429
Original shape – Elongation	−0.142857
Elongation–Flatness Ratio	−0.523810

Overall, the correlation structure suggests that SCI and ADI capture morphological information that is partially distinct from standard radiomic shape features in this small sample. A larger, independent cohort would be needed to confirm this pattern.

Discussions

The empirical behaviour of SCI in this cohort highlights an important conceptual point about how its three multiplicative components relate to shape irregularity. Under the conventions used by PyRadiomics, elongation and flatness range from 0 to 1 and approach 1 as a shape approaches a sphere. Both therefore decrease, as a lung becomes more elongated or flattened.

Holistic shape indices

SCI combines several individual geometric descriptors such as elongation, flatness, axis ratio into a single index subject to the directional caveat above, such holistic indices may help flag geometric deformation of the lung that is not captured by any single conventional shape feature. A longer, more asymmetric phenotype could reflect larger tumors or structural distortion from comorbid pulmonary disease, while a compact phenotype could reflect smaller, peripheral tumors with preserved lung architecture. Testing this interpretation directly against clinical and radiological ground truth is an important next step.

Clinical and research implications

An automated pipeline capable of computing SCI and ADI without manual intervention could, in principle, be used to augment radiology reporting with objective measures of whole-lung structural distortion, and structural complexity metrics of this kind could eventually contribute to models of disease severity or pulmonary function. More broadly, this work illustrates a shift from tumor-centric radiomic analysis toward whole-organ characterization. These are potential directions for future research rather than demonstrated clinical capabilities at this stage.

Study limitations

This feasibility study includes only eight patients, which limits its statistical power and generalizability. LOOCV is an appropriate validation method for this sample size, but all quantitative results should be considered preliminary. The study did not assess real clinical endpoints. Instead, classification relied on synthetic, non-clinical labels to evaluate pipeline performance. The clinical value of SCI and ADI for outcomes such as survival, treatment response, or genomic subtype has not yet been tested All data were drawn from a single cohort, TCGA-LUSC. Validation across different scanners, protocols, and populations in independent, multi-institutional cohorts is still required. Tumor-specific features were excluded from this study. Only whole-lung shape and density were analyzed to focus on organ-level morphology as an initial step. Future research should integrate both tumor-level and whole-lung features in a single model to assess clinical outcomes.

Conclusions

This pilot study presents an automated, end-to-end framework for whole-lung CT morphometric analysis in lung cancer and introduces two novel shape-based biomarkers, the Shape Complexity Index (SCI) and the Axis Disparity Index (ADI), intended to summarize global lung morphology beyond the tumor region of interest. In this small, single-cohort sample (n = 8) drawn from TCGA-LUSC, SCI and ADI captured morphological information that was largely non-redundant with conventional radiomic shape descriptors, contributed to separable unsupervised phenotypes, and did not reduce the performance of a supervised classifier when added to a conventional feature set. The results support the technical feasibility of whole-lung morphometric radiomics and establish a proof-of-concept foundation for the two proposed biomarkers. However, the very small sample size, reliance on provisional and non-clinical outcome labels, and the directional ambiguity observed in the SCI formulation preclude any diagnostic, prognostic,

or predictive claims from this study. Validation in larger, clinically annotated, multi-institutional cohorts, utilizing a refined SCI formulation and actual clinical endpoints, is required before these biomarkers can be considered for further translational development.

Declarations

Acknowledgements

The author acknowledges that the original code and project supporting this study are hosted in the GitHub repository: https://github.com/dremmjoy/lung-ct-radiomics

Data availability

The CT imaging data analysed in this study are publicly available from the TCGA-LUSC collection via The Cancer Imaging Archive (TCIA), https://www.cancerimagingarchive.net/collection/tcga-lusc/ No new datasets were generated during this study.

Consent for publication

Not applicable. This manuscript does not contain any individually identifiable patient data, images, or other personal information.

Competing interests

The authors declare no competing interests.

AI disclosure

Artificial intelligence tools were used for literature search assistance and English language editing during the preparation of this manuscript. All AI-generated content was critically reviewed, verified, and revised by the authors. The authors take full responsibility for the content, scientific validity, and integrity of this work.

References

1. Chen, M., Copley, S. J., Viola, P., Lu, H., & Aboagye, E. O. (2023). Radiomics and artificial intelligence for precision medicine in lung cancer treatment. Seminars in Cancer Biology, 93, 97–113. https://doi.org/10.1016/j.semcancer.2023.05.004
2. Tang, X., Li, Y., Yan, W., Qian, W., Pang, T., Gong, Y., & Yang, Z. (2021). Machine learning-based CT radiomics analysis for prognostic prediction in metastatic non-small cell lung cancer patients with EGFR-T790M mutation receiving third-generation EGFR-TKI osimertinib treatment. Frontiers in Oncology, 11. https://doi.org/10.3389/fonc.2021.719919

3. Wu, L., Lou, X., Kong, N., Xu, M., & Gao, C. (2022). Can quantitative peritumoral CT radiomics features predict the prognosis of patients with non-small cell lung cancer? A systematic review. European Radiology, 33(3), 2105–2117. https://doi.org/10.1007/s00330-022-09174-8
4. Yan, M., Zhang, Z., Tian, J., Yu, J., Dekker, A., Ruysscher, D. de, Wee, L., & Zhao, L. (2025). Whole lung radiomic features are associated with overall survival in patients with locally advanced non-small cell lung cancer treated with definitive radiotherapy. Radiation Oncology, 20(1). https://doi.org/10.1186/s13014-025-02583-1
5. Lim, C. H., Koh, Y. W., Hyun, S. H., & Lee, S. J. (2022). A machine learning approach using PET/CT-based radiomics for prediction of PD-L1 expression in non-small cell lung cancer. Anticancer Research, 42(12), 5875–5884. https://doi.org/10.21873/anticanres.16096
6. Choi, W., Liu, C.-J., Alam, S. R., Oh, J. H., Vaghjiani, R., Humm, J., Weber, W., Adusumilli, P. S., Deasy, J. O., & Lu, W. (2023). Preoperative 18F-FDG PET/CT and CT radiomics for identifying aggressive histopathological subtypes in early-stage lung adenocarcinoma. Computational and Structural Biotechnology Journal, 21, 5601–5608. https://doi.org/10.1016/j.csbj.2023.11.008
7. Hosseini, S. A., Shiri, I., Ghaffarian, P., Hajianfar, G., Avval, A. H., Seyfı, M., Servaes, S., Rosa-Neto, P., Zaidi, H., & Ay, M. R. (2024). The effect of harmonisation on the variability of PET radiomic features extracted using various segmentation methods. Annals of Nuclear Medicine, 38(7), 493–507. https://doi.org/10.1007/s12149-024-01923-7
8. So, A. C. P., Cheng, D., Aslani, S., Azimbagirad, M., Yamada, D., Dunn, R., Josephides, E., McDowall, E., Henry, A.-R., Bille, A., Sivarasan, N., Pennycuick, A., Karapanagiotou, E., & Jacob, J. (2025). A deep learning lung cancer segmentation pipeline to facilitate CT-based radiomics. medRxiv. https://doi.org/10.1101/2025.06.17.25329213
9. Belfiore, M. P., Sansone, M., Monti, R., Marrone, S., Fusco, R., Nardone, V., Grassi, R., & Reginelli, A. (2022). Robustness of radiomics in pre-surgical computed tomography of non-small-cell lung cancer. Journal of Personalised Medicine, 13(1), 83. https://doi.org/10.3390/jpm13010083
10. Amini, E., & Klein, R. (2025). Lung lobe segmentation: performance of open-source MOOSE, TotalSegmentator, and LungMask models compared to a local in-house model. European Radiology Experimental, 9(1). https://doi.org/10.1186/s41747-025-00623-9
11. Fufin, M., Makarov, V., Alfımov, V. I., Ananev, V. V., & Ananeva, A. (2024). Pulmonary fissure segmentation in CT images using image filtering and machine learning. Tomography, 10(10), 1645–1664. https://doi.org/10.3390/tomography10100121
12. Nissar, A., & Mir, A. H. (2024). Proficiency evaluation of shape and WPT radiomics based on machine learning for CT lung cancer prognosis. Egyptian Journal of Radiology and Nuclear Medicine, 55(1). https://doi.org/10.1186/s43055-024-01223-0
13. Raptis, S., Ilioudis, C., & Theodorou, K. (2024). Uncovering the diagnostic power of radiomic feature significance in automated lung cancer detection: an integrative analysis of texture, shape, and intensity contributions. BioMedInformatics, 4(4), 2400–2425. https://doi.org/10.3390/biomedinformatics4040129

14. Vandaele, R., Mukherjee, P., Selby, H. M., Shah, R. P., & Gevaert, O. (2023). Topological data analysis of thoracic radiographic images shows improved radiomics-based lung tumor histology prediction. Patterns, 4(1), 100657. https://doi.org/10.1016/j.patter.2022.100657
15. Genç, H., & Yıldırım, M. (2025). Machine-learning model for differentiating round pneumonia and primary lung cancer using CT-based radiomic analysis. Medicine, 104(37), e44408. https://doi.org/10.1097/md.0000000000044408
16. Corso, F., Tini, G., Lo Presti, G., Garau, N., De Angelis, S. P., Bellerba, F., Rinaldi, L., Botta, F., Rizzo, S., Origgi, D., Paganelli, C., Cremonesi, M., Rampinelli, C., Bellomi, M., Mazzarella, L., Pelicci, P. G., Gandini, S., & Raimondi, S. (2021). The challenge of choosing the best classification method in radiomic analyses: recommendations and applications to lung cancer CT images. Cancers, 13(12), 3088. https://doi.org/10.3390/cancers13123088
17. Leng, Z., Jia, W., Chen, B., Tian, H., & Du, X. (2025). Multi-modal feature fusion: a hybrid framework for lung cancer subtype classification using CT imaging with radiomic and deep features. Journal of Radiation Research and Applied Sciences, 18(3), 101724. https://doi.org/10.1016/j.jrras.2025.101724
18. https://www.cancerimagingarchive.net/collection/tcga-lusc/

Figures

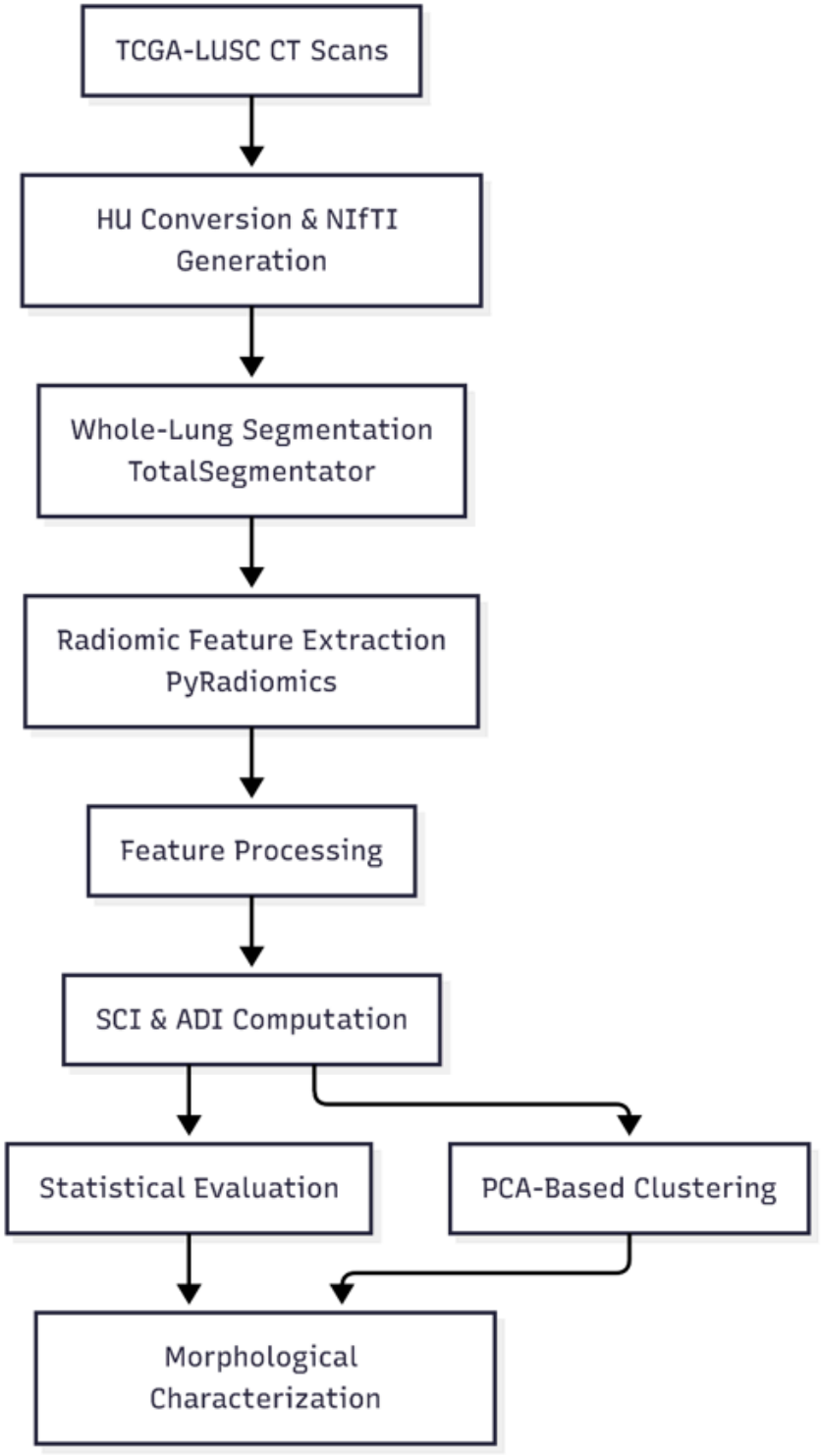


Figure 1

Automated whole-lung CT morphometric radiomics workflow

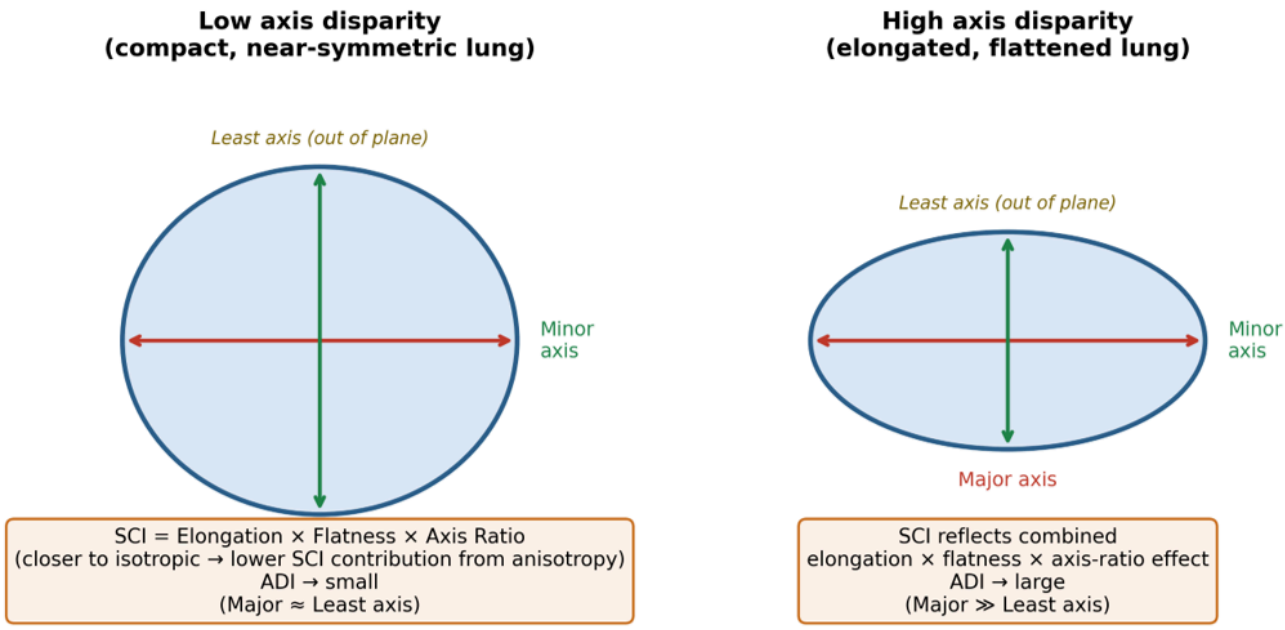


Figure 2

Schematic illustration of the geometric quantities for the Shape Complexity Index (SCI) and Axis Disparity Index (ADI)

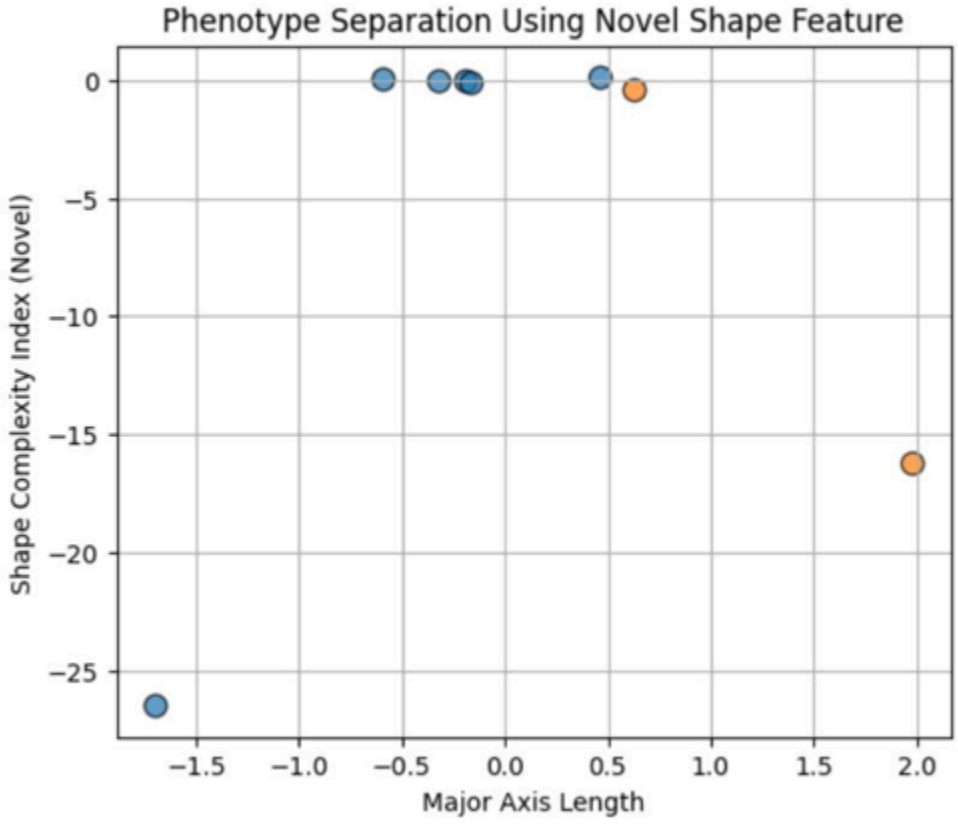


Figure 3

Principal component analysis (PCA) plot of the 8 patients based on shape features.