Journal of Intelligent Decision Making and Information Science

www.jidmis.org
eISSN: 3079-0875

Comparative Analysis of CT Imaging Features in Lung Cancer and Pulmonary Tuberculosis

Maya Genisa*[1], Johari Yap Abdullah[2], Diniwati Mukhtar[3], Chandra Prasetyo Utomo[4], Aryenti[5]

[1,3] Biomedical Sciences Programme, School of Pascasarjana, YARSI University, Jakarta 10510, Indonesia
[2] Craniofacial Imaging Laboratory, School of Dental Sciences, Universiti Sains Malaysia, Kota Bharu 16150, Malaysia
[4] IT Department, YARSI University, Jakarta 10510, Indonesia
[5] Faculty of Medicine, YARSI University, Jakarta 10510, Indonesia
Email: maya.genisa@yarsi.ac.id*, johariyap@usm.my

ARTICLE INFO

Article history:
Received 03 July 2026
Received in revised form 06 Aug 2026
Accepted 16 Aug 2026

Keywords: lung cancer; pulmonary tuberculosis; computed tomography; radiomics; local phase; image derivatives; quantitative imaging.

ABSTRACT

Background: Lung cancer (CA) and pulmonary tuberculosis (TB) can exhibit overlapping appearances on computed tomography (CT), complicating differential diagnosis, particularly in TB-endemic settings such as Indonesia. This study evaluated whether nine quantitative CT attributes could distinguish between the two conditions.

Methods: A retrospective, cross-sectional analysis was conducted using 71 regions of interest (ROIs) from 18 patients, comprising 40 ROIs from 10 patients with lung cancer and 31 ROIs from 8 patients with pulmonary tuberculosis. Nine pixel-wise attributes were evaluated: attenuation (HU), contrast, sharpness, roughness, edge, phase, first derivative, second derivative, and smoothness. Group differences were assessed using the Mann–Whitney U test with Holm correction. Discriminatory performance was evaluated using the area under the receiver operating characteristic curve (AUC) with 95% bootstrap confidence intervals. Patient-level analyses and leave-one-patient-out cross-validation were also performed.

Results: Three attributes differed significantly between the groups after Holm correction. Phase was more negative in TB than in CA (−0.225 vs. −0.138; adjusted p = 0.006; AUC = 0.74, 95% CI 0.62–0.85). The second derivative was lower in TB (0.041 vs. 0.061; adjusted p = 0.015; AUC = 0.72, 95% CI 0.59–0.82). Attenuation was also lower in TB (−687.1 vs. −619.7 HU; adjusted p = 0.024; AUC = 0.70, 95% CI 0.58–0.83). At the patient level, phase (p = 0.021) and second derivative (p = 0.012) remained significant, whereas attenuation did not (p = 0.068). A cross-validated model combining phase, second derivative, and attenuation achieved an AUC of 0.66.

Conclusion: Phase, second derivative, and attenuation demonstrated discriminatory potential, but substantial overlap between the groups and limited model performance preclude their use as independent diagnostic markers. Larger cohorts, standardized image processing, and independent external validation are required to establish their reproducibility and potential clinical utility.

1. Introduction

Lung cancer and pulmonary tuberculosis are two of the leading causes of respiratory death, and on CT they can look alike. Lung cancer was the most frequently diagnosed cancer in 2022, with almost 2.5 million new cases worldwide [1]. TB affected an estimated 10.8 million people in 2023 and caused 1.25 million deaths; Indonesia carried 10% of the global burden, second only to India [2].The two diseases overlap in their CT appearance. Tuberculomas and post-primary TB can present as solid nodules or masses with irregular margins, mimicking malignancy. Conversely, cancer can coexist with

or arise in TB scars. Typical TB signs, such as the tree-in-bud pattern of small centrilobular nodules on branching airways, are suggestive but not specific [3]. Misclassification delays cancer treatment or exposes TB patients to unnecessary invasive procedures. Quantitative image analysis (radiomics) converts images into measurable features that can support diagnosis [4]. Radiomics models have separated lung adenocarcinoma from tuberculoma with high accuracy in solitary solid nodules [5], and feature definitions have been standardised by the Image Biomarker Standardization Initiative [6]. Beyond conventional texture statistics, local signal descriptors such as phase from the monogenic signal [7] and first- and second-order intensity derivatives describe the structure of a lesion independently of its brightness. This study compared nine pixel-wise CT attributes — attenuation, contrast, sharpness, roughness, edge, phase, first derivative, second derivative and smoothness — between CA and TB lesions. The aims were (1) to test which attributes differ between the two diseases at ROI and patient level, (2) to estimate how well each attribute, alone and combined, separates them, and (3) to describe how their spatial patterns differ in representative cases.

2. Methodology

2.1 Study design and subjects

This was a retrospective, cross-sectional comparison of chest CT regions of interest (ROIs) from patients with lung cancer (CA, n = 40) and pulmonary tuberculosis (TB, n = 31). The ROIs came from 10 CA patients (CA-01 to CA-10) and 8 TB patients (TB-01 to TB-08); each patient contributed four ROIs, except TB-08 with three (Figure 1). [Study protocol code

USM/JEPeM/22080524.]

2.2 CT acquisition

The CT images were collected using a Siemens Somatom Definition AS+ (Siemens, Erlangen, Germany) with a 1-mm slice thickness and a matrix of 512 x 512 pixels. Data was collected at the Radiology Department, Hospital USM. The scans were acquired using 120 kVp, with a field of view (FOV) of 256 x 256 mm^2. The attenuation values were kept in Hounsfield units (HU) without rescaling.

2.3 Region of interest

For each patient, four rectangular ROIs enclosing the lesion and its immediate surroundings were placed: two on axial slices and two on coronal slices (sagittal for CA-07). ROI size followed lesion size and ranged from 29 × 32 to 112 × 101 pixels. Analysis was performed with the Lung CT Screening – TB/CA/Normal software, version 1.0 (Universitas YARSI; research use only), which reports the mean and SD of each attribute map within the RO*I*.

2.4 Image attributes

Nine pixel-wise attribute maps were computed for every ROI. The mean and standard deviation (SD) of each map within the ROI were recorded. The value of each attribute was calculated by normalising the sum of the attribute values inside the ROI with the ROI area (pixels).

Attribute	What it measures	Unit / range
HU (original)	X-ray attenuation, i.e. tissue density	HU
Contrast	Local intensity difference between a pixel neighbourhood and its surroundings	0–1
Sharpness	How abrupt local intensity transitions are	0–1
Roughness	Local irregularity of intensity (texture heterogeneity)	0–1
Edge	Strength of boundaries between structures	dimensionless
Phase	Local phase of the image signal, separating line-like from edge-like structure independently of brightness [7]	radians (−π to π)
First derivative	Magnitude of the intensity gradient	HU per pixel
Second derivative	Local curvature of intensity (blob- and ridge-like structure)	0–1
Smoothness	Local homogeneity of intensity; high values mean uniform regions	0–1

2.5 Analysis

Data verification. Every tabulated ROI mean was checked against the value printed on the software's output panel. Eight transcription errors were found and corrected before analysis (Supplementary Table S1). Three of them changed group results materially: a roughness value entered as 0.501 instead of 0.0501, a phase value entered as −2.199 instead of −0.219, and a contrast value entered as 0.0498 instead of 0.498.

Statistics. Attributes were summarised as mean ± SD and median (interquartile range). Groups were compared at ROI level with the two-sided Mann–Whitney U test; Welch's t-test was run as a check. P-values were adjusted for nine comparisons with the Holm method. The difference between groups is also given as a relative difference:

$\Delta rel = (\bar{x}TB - \bar{x}CA) / |\bar{x}CA| \times 100\%$ Discrimination was measured by the AUC with a 95% bootstrap confidence interval (2,000 resamples), oriented so that values above 0.5 indicate separation in either direction. Because ROIs from one patient are not independent, each attribute was also averaged per patient and the 10 CA and 8 TB patient means compared with the Mann–Whitney U test. Finally, an L2-penalised logistic regression combining the significant attributes, and one combining all nine, were evaluated with leave-one-patient-out cross-validation. Significance was set at $\alpha = 0.05$. Analyses used Python 3.13 with SciPy 1.18.

3. Results

3.1 Study cohort

The 18 patients showed the expected spread of appearances (Figure 1). Most CA cases presented as a solitary, round or lobulated solid mass (for example CA-01, CA-05, CA-07 and CA-10); CA-09 showed

diffuse bilateral opacities. TB cases showed cavities (TB-02, TB-08), scattered or diffuse small nodules (TB-03, TB-04) and, in TB-08, a destroyed left lung with pleural thickening.

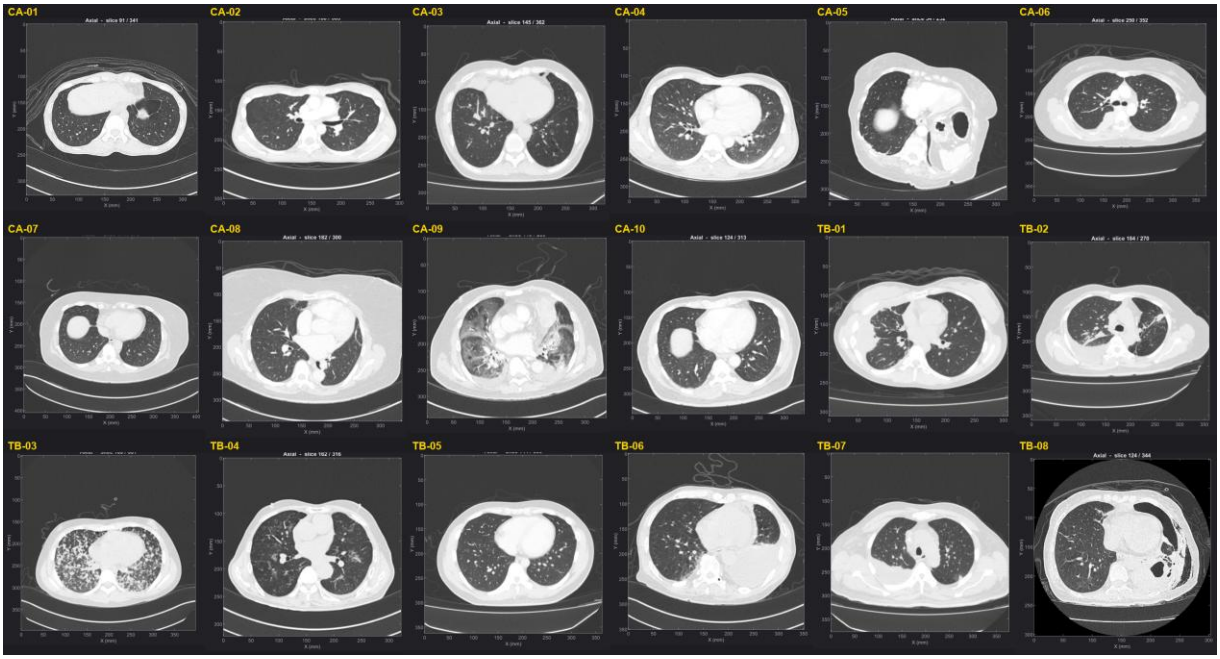


Figure 1. Axial CT overview of every patient in the cohort.

3.2 Group comparison

Phase, second derivative and HU differed significantly between CA and TB after Holm correction; the other six attributes did not (Table 1, Figure 2).

Table 1. ROI-level attribute values (mean ± SD), group comparison and discrimination. CA: 40 ROIs, 10 patients; TB: 31 ROIs, 8 patients.

Attribute	CA	TB	Δrel (%)	p	p (Holm)	AUC (95% CI)	p (patient level)
Phase	−0.138 ± 0.082	−0.225 ± 0.101	−63.5	< 0.001	0.006	0.74 (0.62–0.85)	0.021
Second derivative	0.061 ± 0.028	0.041 ± 0.018	−33.0	0.002	0.015	0.72 (0.59–0.82)	0.012
HU	−619.7 ± 126.3	−687.1 ± 96.8	−10.9	0.003	0.024	0.70 (0.58–0.83)	0.068
Sharpness	0.159 ± 0.068	0.170 ± 0.058	+7.4	0.085	0.511	0.62 (0.49–0.74)	0.083

Roughness	0.044 ± 0.015	0.048 ± 0.014	+7.9	0.610	1.000	0.54 (0.39–0.67)	0.697
First derivative	92.6 ± 28.0	88.2 ± 22.0	−4.7	0.647	1.000	0.53 (0.40–0.67)	0.697
Contrast	0.401 ± 0.088	0.401 ± 0.088	+0.1	0.711	1.000	0.53 (0.39–0.68)	0.573
Smoothness	0.774 ± 0.080	0.765 ± 0.078	−1.2	0.857	1.000	0.51 (0.38–0.64)	0.965
Edge	0.403 ± 0.104	0.419 ± 0.126	+4.1	0.972	1.000	0.50 (0.36–0.65)	0.573

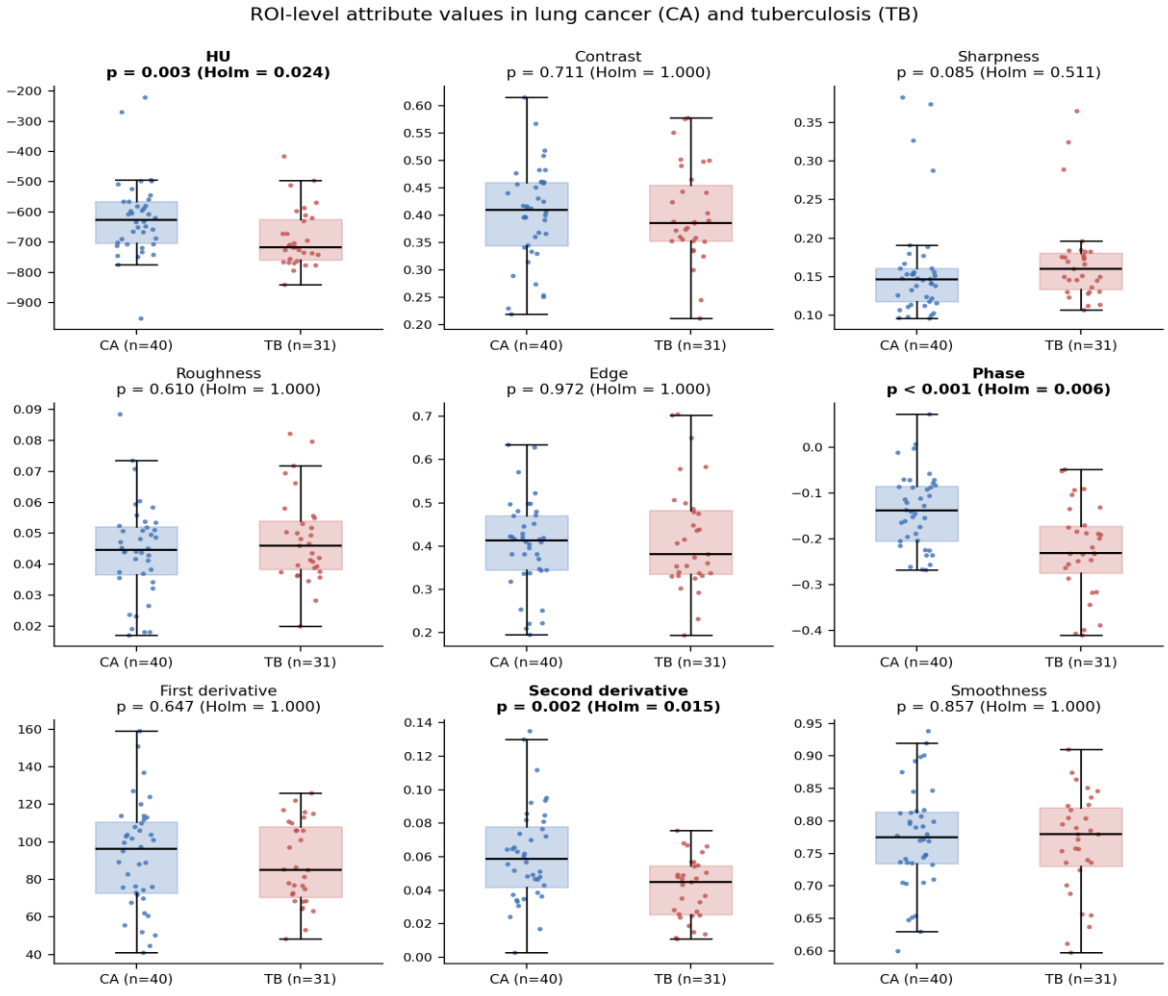


Figure 2. ROI-level distribution of each attribute in CA and TB.

Phase was more negative in TB: median −0.231 (IQR −0.275 to −0.173) vs −0.139 (−0.204 to −0.085) in CA; Cohen's d = −0.97.

Second derivative was lower in TB: median 0.045 (0.026–0.055) vs 0.059 (0.042–0.078); d = −0.82.

HU was lower in TB: median −717 (−759 to −625) vs −626 (−703 to −566) HU; d = −0.59.

Roughness, contrast, edge, first derivative and smoothness were nearly identical in the two groups (relative differences 0.1–7.9%). Welch's t-test gave the same three significant attributes.

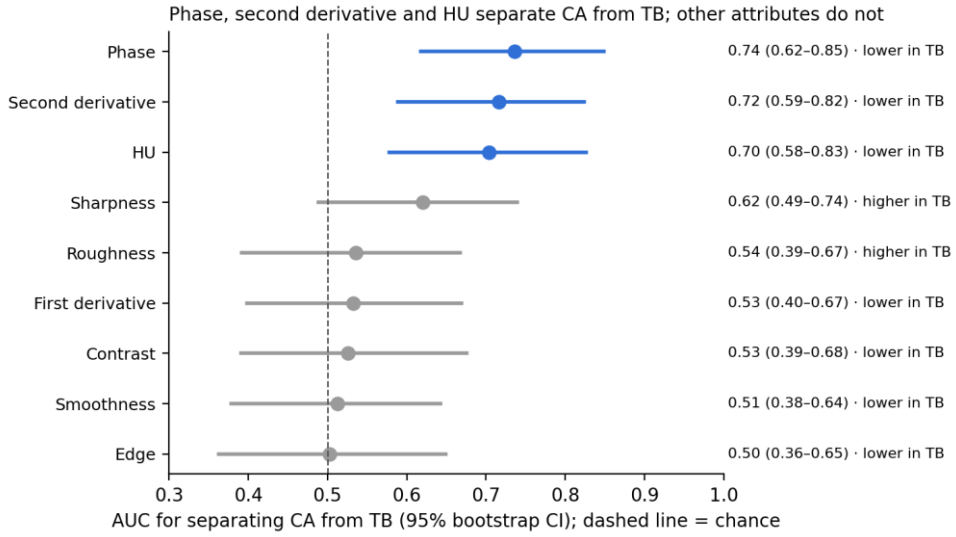


Figure 3. AUC with 95% bootstrap confidence interval for each attribute.

3.3 Patient-level analysis and combined model

To examine whether the ROI-level findings were maintained after accounting for the clustering of multiple ROIs within individual patients, the mean value of each attribute was calculated for each patient. At the patient level, phase ($p = 0.021$) and second derivative ($p = 0.012$) remained significantly different between the CA and TB groups, whereas attenuation (HU) did not reach statistical significance ($p = 0.068$).

The differences between the ROI-level and patient-level findings indicate that the apparent discriminatory performance of individual attributes depends partly on the unit of analysis. Although attenuation differed significantly between the groups at the ROI level after Holm correction, this difference was not statistically significant when the analysis was based on patient-level means.

The leave-one-patient-out cross-validated logistic regression model combining phase, second derivative, and attenuation achieved an AUC of 0.66. In comparison, the model incorporating all nine attributes achieved an AUC of 0.56. These results suggest that adding more attributes did not improve discrimination in this small dataset.

The observed AUC values indicate limited to moderate separation between the two diagnostic groups. Substantial overlap remained between individual ROI measurements, and the combined model did not demonstrate sufficient discriminatory performance to support clinical diagnostic use.

Furthermore, the absence of external validation and calibration analysis limits conclusions regarding the model's performance in other patient populations.

3.4 Representative cases

One case per group (CA-06, PT78F288; TB-02, PT603670) was examined map by map. Their ROI means were close, but the SDs and spatial patterns differed clearly (Table 2, Table 3, Figures 4 and 5).

Table 2. ROI mean (SD) of each attribute in the representative cases.

Attribute	CA case PT78F288	TB case PT603670
HU	−666 (325)	−597 (274)
Contrast	0.461 (0.231)	0.466 (0.196)
Sharpness	0.189 (0.125)	0.146 (0.0757)
Roughness	0.0604 (0.0439)	0.0580 (0.0345)
Edge	0.498 (0.596)	0.500 (0.498)
Phase	−0.140 (0.832)	−0.184 (0.847)
First derivative	104 (113)	97.2 (89.6)
Second derivative	0.0658 (0.150)	0.0683 (0.133)
Smoothness	0.710 (0.284)	0.736 (0.233)

Most case means differed by less than 5%. HU, sharpness and second derivative ran opposite to the group trend, showing that a single ROI mean is unreliable on its own. The CA case had the higher SD in 8 of 9 attributes; phase SD was similar (0.832 vs 0.847).

Table 3. Spatial pattern of each map in the representative cases.

Map	CA (PT78F288, coronal)	TB (PT603670, axial)
Original	One compact, wedge-shaped solid mass with linear strands radiating outward	Several irregular, branching, tubular opacities, small satellite nodules and an air channel within the lesion
Contrast	One large continuous high-contrast region with a uniform interior	Patchy, fragmented high-contrast regions
Sharpness	Hot spots clustered around the mass	Many small scattered hot spots matching satellite micronodules
Edge and first derivative	One strong outline of the mass plus straight strands	Many tortuous, worm-like edges and small ring-shaped edges

Second derivative	One focal round bright peak (solid core)	An elongated, beaded chain of moderate peaks
Roughness	Concentrated at the margin of the mass	A central hot spot with a surrounding network
Phase	One large coherent region over the mass	Many small islands of alternating phase
Smoothness	A dark band tracing a single outline	Many winding dark lines

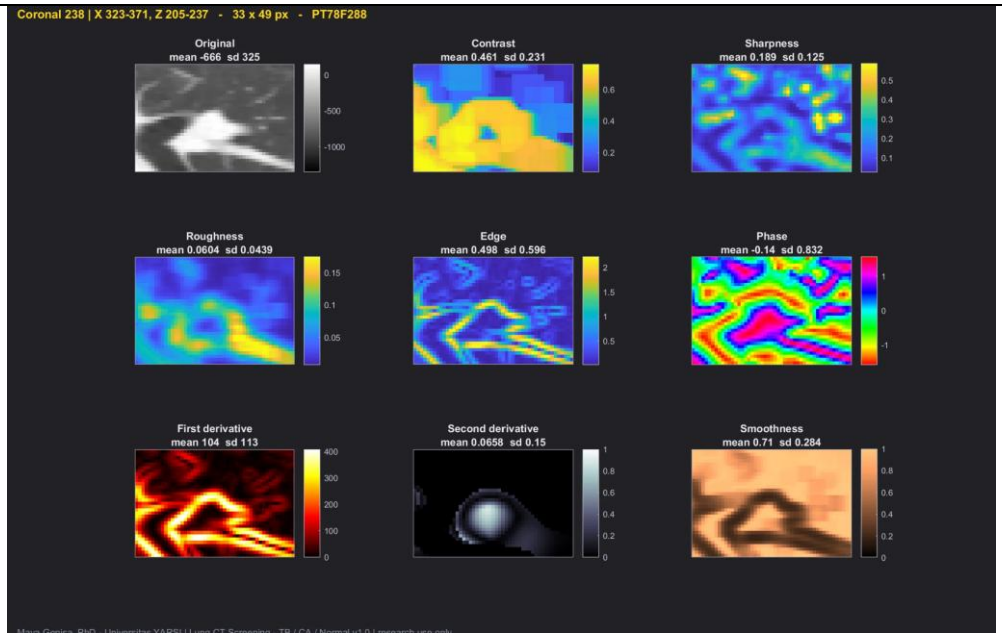


Figure 4. Representative CA case (CA-06, PT78F288).

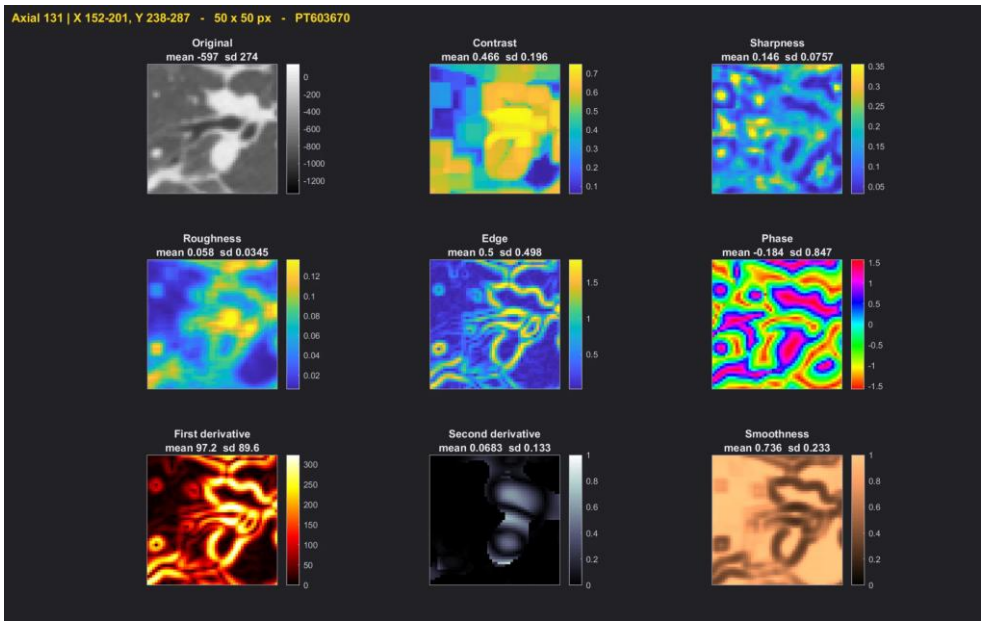


Figure 5. Representative TB case (TB-02, PT603670).

4. Discussion

Phase, second derivative and attenuation separated CA from TB, each with an AUC of about 0.70–0.74; the six texture and edge attributes did not. Phase and second derivative held at patient level, which makes them the most robust candidates in this cohort.

4.1 Interpretation of the discriminating attributes

Phase. Local phase describes the type of structure — bright line, dark line or edge — independently of brightness [7]. A more negative mean phase in TB points to more dark, line-like structure, consistent with the air-containing bronchi, cavities and air channels seen in TB lesions (Figure 1, TB-02 and TB-08). The representative TB case showed a fragmented phase map; the CA case showed one coherent phase region over a solid mass.

Second derivative. Second-order intensity derivatives respond to blob- and ridge-like structure. The higher values in CA fit a compact, solid core, seen as one focal peak in the representative CA case and as round masses in most CA patients. In TB the response was weaker and, in the representative case, formed a beaded chain along tubular airway-centred lesions, as in the tree-in-bud pattern [3].

Attenuation. CA ROIs were denser (−620 vs −687 HU). Both means are low because each ROI includes aerated lung, so HU largely reflects how much solid tissue fills the box. This dependence on ROI placement probably explains why HU lost significance at patient level.

4.2 Attributes that did not discriminate

Roughness, contrast, sharpness, edge, first derivative and smoothness overlapped almost completely. In the uncorrected data, roughness appeared 40.6% higher in TB; this came from a single value entered ten times too large. Texture heterogeneity, often assumed to separate TB from cancer, therefore did not do so here once the data were verified. This shows how much a single transcription error can change a small radiomics study.

4.3 Comparison with previous work

Radiomics studies have separated lung adenocarcinoma from tuberculoma with high accuracy. Zhuo et al. combined 15 radiomics features with clinical and morphological signs (lobulation, vacuole, nodule diameter) and reported an AUC of 0.99 in validation [5]. Our single-attribute AUCs of 0.70–0.74 are lower, as expected from ROI means of nine maps in 18 patients. Both studies point to morphology — compact versus branching, solid versus air-containing — as central, and our phase and derivative measures describe exactly those properties.

4.4 Clinical relevance

The clinical interpretation of these findings requires caution. Although phase, second derivative, and attenuation showed discriminatory potential at the ROI level, the observed AUC values and substantial overlap between the groups indicate that these attributes cannot independently establish a diagnosis of lung cancer or pulmonary tuberculosis. In particular, attenuation did not remain statistically significant in the patient-level analysis, highlighting the importance of accounting for within-patient clustering when interpreting imaging biomarkers.

The findings are therefore best regarded as exploratory evidence that selected quantitative CT attributes may complement conventional radiological assessment. Their potential contribution to clinical decision-making must be evaluated in larger, independently collected cohorts and in conjunction with established diagnostic procedures. No conclusions regarding improved patient outcomes or reduced diagnostic errors can be drawn from the present study alone.

4.5 Limitations

Several limitations should be considered when interpreting the findings of this study.

First, the cohort was small and comprised only 18 patients, including 10 patients with lung cancer and 8 with pulmonary tuberculosis. This limited sample size reduces statistical power, increases uncertainty in the estimated effect sizes and AUC values, and restricts the generalizability of the findings. Differences among lung cancer subtypes and pulmonary tuberculosis presentations were not evaluated separately.

Second, the dataset contained 71 ROIs obtained from the 18 patients. Because multiple ROIs from the same patient may share anatomical and pathological characteristics, they cannot be considered fully independent observations. Although patient-level aggregation was performed to examine the robustness of the findings, the analysis included only 18 independent patient-level units. Consequently, the patient-level comparisons remain subject to substantial uncertainty.

Third, the ROIs were two-dimensional rectangular regions that included the lesions and surrounding lung tissue. Their measurements may therefore depend on ROI dimensions, imaging plane, lesion size, and the proportion of the ROI occupied by abnormal tissue. The use of different imaging planes and ROI sizes may have introduced additional variability. Future studies should consider three-dimensional lesion segmentation and a standardized ROI-selection protocol.

Fourth, the combined logistic regression models were evaluated using leave-one-patient-out cross-validation, but no external validation was performed. The three-feature model was constructed using attributes identified as significant in the same dataset, which introduces a risk of optimistic performance estimates if feature selection was not repeated independently within each training fold. The reported AUC of 0.66 should therefore be interpreted as exploratory rather than as evidence of established diagnostic performance. The calibration of the predicted probabilities was also not assessed, and the study does not establish whether the model produces reliable absolute risk estimates.

Fifth, eight transcription errors were identified and corrected by checking the tabulated values against the software output. Although this verification improved the consistency of the analyzed data, the need for corrections highlights the importance of implementing systematic data-quality checks in quantitative imaging studies. One additional TB-08 ROI available in the software output was not included in the analysis, and the effect of this omission was not evaluated.

Finally, the quantitative attributes were generated using a research-use software application, and their computational definitions may depend on the specific implementation. Reproducibility across software versions, scanners, acquisition protocols, and reconstruction settings was not assessed.

These limitations restrict direct comparisons with other radiomics studies and prevent conclusions regarding generalizability across clinical settings.

Taken together, the findings should be interpreted as preliminary evidence of differences in selected CT attributes between lung cancer and pulmonary tuberculosis. Larger multicentre studies, standardized image processing, independent validation, and more comprehensive model evaluation are required before clinical application can be considered.

4.6 Future work

4.6 Future Work

Future research should evaluate the observed CT attributes in larger, multicentre cohorts that include more diverse patient populations, lung cancer subtypes, and pulmonary tuberculosis presentations. An independent external validation cohort is needed to assess whether the discriminatory performance can be reproduced beyond the original study population.

Image-processing procedures should also be standardized, including ROI selection, lesion segmentation, image reconstruction settings, and the computational definitions of the quantitative attributes. Three-dimensional lesion segmentation may reduce the influence of surrounding lung tissue on ROI-level measurements.

In addition to ROI means, future models should incorporate spatial-pattern descriptors that capture lesion organization and structural heterogeneity. Candidate features include the number of separate second-derivative peaks, the number of ring-shaped edges, phase entropy, and the fractal dimension of the edge map. Their incremental value should be tested against the existing attributes using patient-level data partitioning and independent validation.

Future model evaluation should extend beyond AUC to include confidence intervals, sensitivity, specificity, predictive values, and calibration measures, where supported by an adequately sized dataset. These assessments would help determine whether quantitative CT attributes provide reproducible and clinically meaningful information when used alongside established radiological and microbiological investigations.

CONCLUSION

This study identified phase, second derivative, and attenuation as quantitative CT attributes that differed between lung cancer and pulmonary tuberculosis at the ROI level. Their individual AUC values ranged from 0.70 to 0.74, indicating limited to moderate discriminatory performance. At the patient level, phase and second derivative remained statistically significant, whereas attenuation did not.

The findings suggest that differences in local image structure may provide useful information for characterizing pulmonary lesions. However, substantial overlap between the two groups and the limited performance of the combined model indicate that these attributes are insufficient for independent diagnosis. The findings should therefore be interpreted as exploratory and should not be considered evidence of clinical diagnostic validity.

Further research involving larger multicentre cohorts, standardized image-processing procedures, spatial-pattern descriptors, and independent external validation is necessary to establish the reproducibility, generalizability, and potential clinical utility of quantitative CT attributes for differentiating lung cancer from pulmonary tuberculosis.

Acknowledgment

We thank YARSI University and DIKTI for supporting this research through project contract number: 0652/LL3/AL.04/2026, and all partners/collaborators, including USM, for supporting this work.

Figure legends

Figure 1. Axial CT overview slices of all 18 patients (CA-01 to CA-10, TB-01 to TB-08), shown on the slice used for an axial ROI. Labels are study codes; no identifying information is shown.

Figure 2. Box plots of the nine attributes for the 40 CA and 31 TB ROIs. Boxes show the median and interquartile range, whiskers extend to 1.5 × IQR, and dots are individual ROIs. Titles give the Mann–Whitney p-value and the Holm-adjusted p-value; bold titles are significant after adjustment.

Figure 3. Area under the ROC curve for separating CA from TB with each attribute alone, with 95% bootstrap confidence intervals (2,000 resamples). AUC is oriented so that 0.5 is chance; the text gives the direction of the difference. Blue: significant after Holm correction.

Figure 4. Representative lung cancer case (CA-06, PT78F288; coronal slice 238, ROI X 323–371, Z 205–237; 33 × 49 pixels). Panels show the original CT (HU) and the eight derived maps, each with its ROI mean and SD. Note the single compact mass, one continuous outline in the edge and first-derivative maps, and one focal second-derivative peak.

Figure 5. Representative pulmonary tuberculosis case (TB-02, PT603670; axial slice 131, ROI X 152–201, Y 238–287; 50 × 50 pixels). Panels as in Figure 4. Note the branching, multifocal opacities, ring-shaped edges, fragmented phase and beaded second-derivative pattern.

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. doi:10.3322/caac.21834.

[2] World Health Organization. Global Tuberculosis Report 2024. Geneva: World Health Organization; 2024.

[3] Rossi SE, Franquet T, Volpacchio M, Giménez A, Aguilar G. Tree-in-bud pattern at thin-section CT of the lungs: radiologic-pathologic overview. Radiographics. 2005;25(3):789-801. doi:10.1148/rg.253045115.

[4] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563-577. doi:10.1148/radiol.2015151169.

[5] Zhuo Y, Zhan Y, Zhang Z, et al. Clinical and CT radiomics nomogram for preoperative differentiation of pulmonary adenocarcinoma from tuberculoma in solitary solid nodule. Front Oncol. 2021;11:701598. doi:10.3389/fonc.2021.701598.

[6] Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295(2):328-338. doi:10.1148/radiol.2020191145.

[7] Felsberg M, Sommer G. The monogenic signal. IEEE Trans Signal Process. 2001;49(12):3136-3144. doi:10.1109/78.969520.