Computer Methods and Programs in Biomedicine 257 (2024) 108474

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Lung nodule classification using radiomics model trained on degraded SDCT images

Jiaying Liu [a,*], Anna Corti [a], Valentina D.A. Corino [a,b], Luca Mainardi [a]

[a] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Politecnico di Milano, Via Giuseppe Ponzio, 34, 20133 Milan, Italy
[b] Cardiotech Lab, Centro Cardiologico Monzino IRCCS, Milan, Italy

ARTICLE INFO

Keywords:
Standard-dose CT degradation
Low-dose CT
Lung nodules
Machine learning
Radiomics

ABSTRACT

Background and objective: Low-dose computed tomography (LDCT) screening has shown promise in reducing lung cancer mortality; however, it suffers from high false positive rates and a scarcity of available annotated datasets. To overcome these challenges, we propose a novel approach using synthetic LDCT images generated from standard-dose CT (SDCT) scans from the LIDC-IDRI dataset. Our objective is to develop and validate an interpretable radiomics-based model for distinguishing likely benign from likely malignant pulmonary nodules.
Methods: From a total of 1010 CT images (695 SDCTs and 315 LDCTs), we degraded SDCTs in the sinogram domain and obtained 1950 nodules as the training set. The 675 nodules from the LDCTs were stratified into 50%-50% partitions for validation and testing. Radiomic features were extracted from nodules, and three feature sets were assessed using: a) only shape and size (SS) features, b) all features but SS features, and c) all features. A systematic pipeline was developed to optimize the feature set and evaluate multiple machine learning models. Models were trained using degraded SDCT, validated and tested on the LDCT nodules.
Results: Training a logistic regression model using three SS features yielded the most promising results, achieving on the test set mean balanced accuracy, sensitivity, specificity, and AUC-ROC scores of 0.81, 0.76, 0.85, and 0.87, respectively.
Conclusions: Our study demonstrates the feasibility and effectiveness of using synthetic LDCT images for developing a relatively accurate radiomics-based model in lung nodule classification. This approach addresses challenges associated with LDCT screening, offering potential implications for improving lung cancer detection and reducing false positives.

1. Introduction

Lung cancer is the deadliest malignancy in the world, being responsible for 18% of total cancer deaths in 2020 [1]. Being often diagnosed at advanced stages, it is associated with a low 5-year survival rate (being only 13% in Europe) [2]. Several trials have recently proved the efficacy of low-dose computed tomography (LDCT) screening in reducing lung cancer mortality [3–5]. However, LDCT screening is associated with high false positive rates: up to 96.4% of the positive screening results were false positive [3]. With the implementation of screening programs in the US and ongoing efforts in Europe to assess the benefits and cost savings of lung cancer screening through trials, the workload on radiologists increases significantly. To alleviate the burden on radiologists and overcome the high number of false positives, extensive research has been conducted to offer objective and efficient diagnostic tools [6].

Radiomics analysis is based on the premise that image-captured characteristics of tumors, such as shape and texture, are potential tumor biomarkers that can be used to stratify patients, predict patient clinical outcomes or further characterize tumor phenotypes. In this context, radiomic features can enable the development of computer-aided diagnostic tools supporting clinical decisions [7,8]. Moreover, radiomics has the advantage of being more explainable than deep learning features, which are intrinsically more abstract and difficult to interpret [8,9].

Several studies applied radiomics with machine learning for nodule malignancy prediction in lung cancer screening, as detailed in Table 1. Other studies using deep learning approaches, such as Artificial Neural Networks and Convolutional Neural Networks [10], achieved good performances but they were out the scope of this paper as we intend to

* Corresponding author.
E-mail address: jiaying.liu@polimi.it (J. Liu).

https://doi.org/10.1016/j.cmpb.2024.108474
Received 16 May 2024; Received in revised form 11 October 2024; Accepted 22 October 2024
Available online 23 October 2024

develop a more explainable model. Although the extensive research in the field of lung cancer screening, some studies report low sensitivity rates (e.g., 56.0% in [11], 70.0% in [9], 10.2% in [12]), others considered small datasets (e.g., 72 nodules in [13], 113 nodules in [14]) or deliberately selected balanced datasets to address class imbalance [9, 15]. All studies but one used LDCT images in their research.

Publicly available datasets, as the National Lung Screening Trial (NLST) and the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), have been extensively used for the development of nodule detection and classification but present distinct limitations [3,18]. NLST is a randomized controlled trial designed to determine if lung cancer screening using LDCT or chest radiography reduces mortality. Despite its size, containing around 25,000 participants, it lacks tumor segmentation [3]. Nevertheless, segmentation is a critical step in a radiomic approach, as features are extracted from the segmented volumes. On the other hand, LIDC-IDRI consists of diagnostic and lung cancer screening CT scans with marked-up annotated lesions from experienced radiologists. However, the number of LDCT images is relatively limited (315 out of the 1010 CTs, considering images acquired with tube current times seconds < 80 *mAs*).

In order to extract radiomic features from the nodules and build a radiomic-based classifier, the nodules require to be segmented. Thus, we intend to develop a classifier capable of distinguishing likely benign nodules from likely malignant ones during the screening stage. In this scenario, without the annotations from the screening NLST dataset, we adopted an innovative strategy leveraging the annotations from the non-screening LIDC-IDRI dataset by degrading standard-dose CT (SDCT) scans. The aim of this study is to develop an interpretable radiomics-based model using degraded SDCT images for the classification of likely benign pulmonary nodules from likely malignant ones and validate the classifier by testing on real LDCT images. Being of its first, this approach not only addresses the scarcity of annotated LDCT data but also offers a novel solution to explore the potential inclusion and utilization of SDCT images in the model development for detecting screening pulmonary nodules, especially in situations where LDCT screening is not widely available or incorporated into protocols.

2. Methods

2.1. SDCT degradation and image preprocessing

To develop a LDCT-based lung classifier utilizing both SDCT and LDCT images from the LIDC-IDRI dataset, we degraded the SDCT images. This degradation step was essential to ensure compatibility between the different image types and facilitate the development of a classifier capable of effectively analyzing both SDCT and LDCT modalities.

Additionally, all images (both SDCT and LDCT) were preprocessed to reduce the image-related sources of variability. The preprocessing steps were thresholding, to ensure all pixel values were positive with negative pixel values set to zero, and denoising, using a 3D Gaussian filter with a $3 \times 3 \times 3$ voxel kernel and $\sigma = 0.5$. Following the thresholding step, SDCT images were degraded to obtain simulated LDCT. All steps were performed using MATLAB version 2022b.

SDCT degradation (Fig. 1) was performed using a low-dose simulation technique in the sinogram domain, as proposed by Zeng et al. [19]. Briefly, standard-dose sinogram data were obtained by using Radon transformation (360 projection angles with 1-degree increments); subsequently, simulated Poisson and Gaussian noise was added to mimic the effect of electronic and quantum noise, generating low-dose sinogram data. Eq. (1) describes the calculation for the low-dose sinogram data ($p_{ld,sim}$):

$$p_{ld,sim} = log\frac{I^0_{ld,sim}}{Poisson\left(I^0_{ld,sim}(s) * \exp\left(-p_{sd}\right)\right) + Gaussian\left(m_e, \sigma_e^2\right)} \quad (1)$$

where $I^0_{ld,sim}$ is the simulated low-dose scan incident flux, p_{sd} is the standard-dose sinogram data obtained from the original SDCT images using Radon transformation, and m_e and σ_e^2 are the mean and variance of the electronic noise, respectively. Specifically, $I^0_{ld,sim}$ is computed by multiplying the incident flux with the ratio between the high and simulated low-dose incident flux, k. This ratio is computed based on the proposed equation Eq. (2):

Table 1
Summary of recent studies applying radiomics for nodule malignancy prediction in lung cancer screening.

Study	Dataset (nodule size)	LDCT?	Model description	Sen.	Spc.	Acc.	AUC	Limitations
Choi et al. 2018 [13]	72 patients LIDC-IDRI ($\geq$ 3 mm)	yes	2 radiomic features, LASSO features selection and SVM classification	87.2%	81.2%	84.6%	0.89	small dataset
Peikert et al. 2018 [15]	726 nodules from NLST ($\geq$ 7 mm)	yes	8 radiomic features selected and trained with LASSO model	90.4%	85.5%	-	0.94	selected data to balance the dataset
Alahmari et al. 2018 [11]	467 cases from NLST ($\geq$ 4 mm)	yes	18 Rider features, ReliefF feature selection, Random Forest	49.0%	93.0%	-	0.82	low sensitivity
Causey et al. 2018 [9]	664 nodules from LIDC-IDRI ($\geq$ 3 mm)	no	a logistic regression model trained only on a size metric (square root of largest cross-sectional area)	69.7%	95.5%	82.6%	0.94	low sensitivity
Mao et al. 2019 [16]	294 nodules (6–15 mm)	yes	11 radiomic features, LASSO feature selection, linear regression model	81.0%	92.2%	89.8%	0.90	small sample size
Garau et al. 2020 [14]	113 nodules COSMOS (all sizes)	yes	5 radiomic features, SVM-LASSO test set: 72 nodules from LIDC-IDRI.	90.0%	FPR = 35.5%	79.1%	0.86	small sample size
Liu et al. 2021 [17]	141 nodules (not specified)	yes	3 radiomic features, radiomic monogram	86.4%	85.7%	86.0%	0.98	small sample size, potential selection bias
Rundo et al. 2021 [12]	703 nodules from bioMILD trial (all sizes)	yes	32 radiomic features, Borderline-SMOTE over-sampling method, elastic net regularization for logistic regression, tested with a blind test	67.0%	86.2%	76.6%	0.82	low sensitivity

LIDC-IDRI - the Lung Image Database Consortium and Image Database Resource Initiative; NLST - National Lung Screening Trial; Sen. – sensitivity; Spc. – specificity; Acc. – accuracy; AUC – area under the ROC curve; FPR – false positive rate.

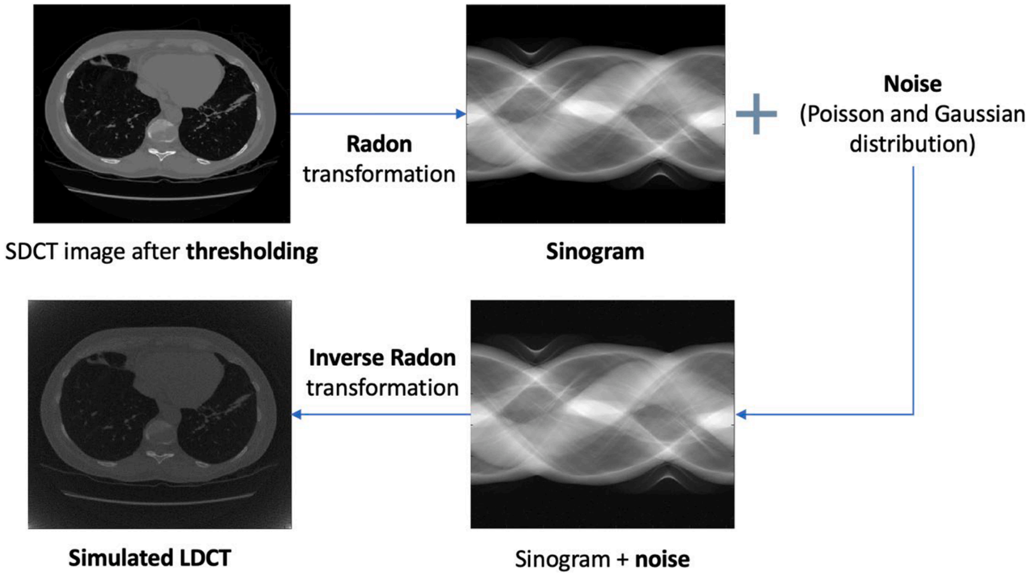

Fig. 1. Pipeline to simulate low-dose computed tomography (LDCT). The Radon transformation is applied to the standard-dose computed tomography (SDCT) image, then electronic and quantum noise effects are included by noise injection. The final step involves inverse Radon transformation to reconstruct the LDCT image.

$$k\ (mAs) = a * (mAs) + b \tag{2}$$

where $a = 0.0095$ and $b = 0.0309$. In the present study, three different low-dose scenarios were simulated using different k-values corresponding to 40, 50, and 75 *mAs* to model variability in dose levels. We assigned $k1(40mAs)$ to half of the data points, $k2(50mAs)$ to a quarter, and $k3\ (75mAs)$ to the remaining quarter. This distribution was based on the proportions found in a subset of LDCT images from the LIDC-IDRI dataset. Electronic noise was modeled using Gaussian noise with a standard deviation $\sigma = \sqrt{11}$, consistent with the scanner noise characteristics outlined in Zeng et al [19]. and Ma et al [20]. Finally, the simulated LDCT images were reconstructed from the low-dose sinogram data using the inverse Radon transform. For parameters settings see more details in [19].

2.2. Dataset description

The dataset used in this study derives from LIDC-IDRI, available in [21]. It contains 1018 thoracic CT scans acquired from 1010 participants, with nodule annotations performed by one to four experienced radiologists. After excluding the duplicated participants, a total of 2625 nodules, each assigned as $\geq$ 3 mm (lesions with greatest in-plane dimension in the range 3–30 mm) marked by at least one radiologist, was included in this study.

The dataset includes some participants with SDCT and others with LDCT images. Tube current ranged from 40 to 650 *mAs*, while tube voltage ranged from 120 to 140 *kV*. Even if there is no common consensus, previous studies considered as LDCT those images acquired with tube current-time product ranging from 20 to 100 *mAs* and voltages spanning from 80 to 140 *kV* [22,23]. For our analysis, LDCT images were defined as those acquired with a tube current $\leq$ 80 *mAs*, similar to the threshold used in the NLST [22].

Each nodule was evaluated by the radiologists for the likelihood of malignancy on a scale of 1 to 5: 1 – highly unlikely, 2 – moderately unlikely, 3 – indeterminate, 4 – moderately suspicious, 5 – highly suspicious.

Given that multiple radiologists might provide scores for a single nodule, a consensus was reached by averaging the scores to derive a single score for each nodule. In our study, nodules with an average score $\geq$ 4 were considered malignant. The summary of nodule characteristics in each subset is presented in Table 2.

Table 2
Summary of the LIDC-IDRI dataset used in this study.

	SDCT	LDCT
Nr. participants	695	315
Nr. nodules	1950	675
Averaged likelihood of malignancy		
[1;2]	280	86
[2;3]	766	222
[3;4]	679	298
[4;5]	225	69
Non-malignant (average likelihood <4)	1725	606
Malignant (average likelihood ≥4)	225	69

SDCT: standard-dose CT; LDCT: low-dose CT.

2.3. Pipeline description

Fig. 2 details the workflow of the study, consisting of radiomic feature extraction, feature selection, model training, validation and tuning, and model evaluation. In this study, nodules in SDCT subset were used as the training set, while nodules in the LDCT subset were stratified into 50–50% partitions, designated as the validation and test set, respectively.

2.3.1. Radiomic feature extraction

Radiomic features were extracted from the region of interest (ROI), the segmented nodules, using Pyradiomics [24]. When the nodule had more than one annotation, the consensus mask was used to extract the radiomic features. A total of 851 features were extracted from each nodule, including 14 shape features, 18 first-order features, 75 texture features (24 gray-level co-occurrence matrix (GLCM) features, 14 gray-level dependence matrix (GLDM) features, 16 gray-level run length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM), and 5 neighboring gray tone difference matrix (NGTDM) features), and 744 wavelet-based features. For more details on radiomic features, see Pyradiomics documentation [24].

2.3.2. Feature selection

Feature selection step was performed on three different feature subsets independently: a) only shape and size (SS) features, b) all features but SS features, and c) all features. These three sets were chosen to assess the relevance of nodule morphological characteristics on the

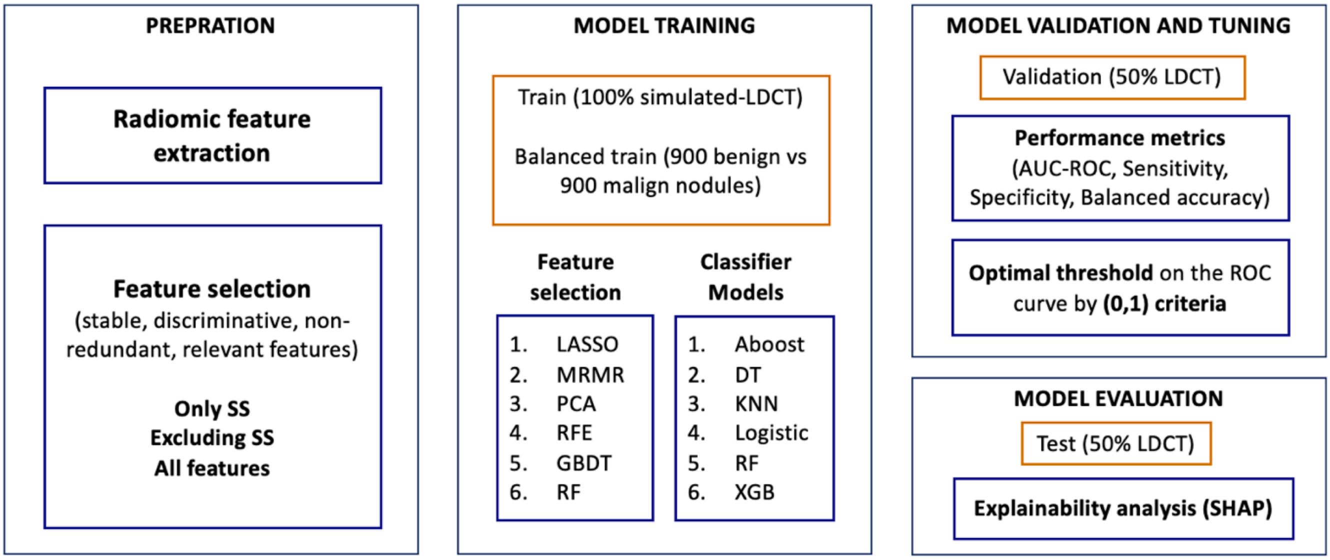

Fig. 2. Pipeline consisting of radiomic feature extraction, feature selection (using a) only shape and size (SS) features, b) all features but SS features, and c) all features), model training and validation, and explainability analysis.

prediction.

All feature selection steps were performed on the training set and then applied to the test set. Additionally, the selection of stable, discriminant and non-redundant features were computed before balancing the training set (see Section 2.3.3), while the selection of more relevant features was performed after obtaining a balanced dataset.

2.3.2.1. Selection of stable and discriminant features. Firstly, a stability analysis was performed to identify the features that exhibit consistent behavior under small transformations. This analysis was conducted by considering nodules that had at least four annotations (705 nodules): features were computed for each ROI and the inter-class correlation coefficient (ICC) was obtained. Features were deemed stable if ICC was > 0.75, as defined by Koo et al. [25]. Second, a discriminative analysis was conducted to discard features that have similar values under large transformations and, therefore, can be considered non-informative. The analysis consisted in translating the original ROIs 150% of the size of the box surrounding the ROI (Fig. 3) and computing the ICC: features were considered discriminative (i.e., informative) when ICC was < 0.5.

2.3.2.2. Selection of non-redundant features. To address redundancy, a pairwise correlation analysis was performed: when a pair of features had an absolute Spearman's correlation coefficient above 0.85, only one of the two was kept. In particular, the one with a lower mean correlation (computed through Spearman's correlation coefficient) with all the other n-2 features was selected. Afterwards, the non-redundant features underwent normalization using z-score standardization.

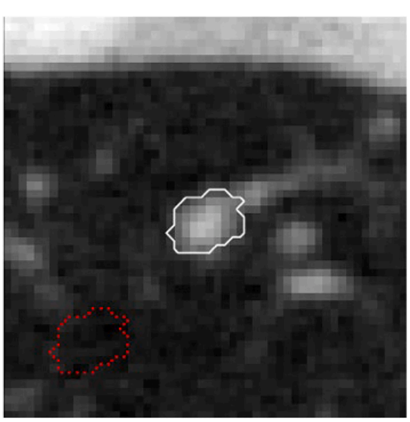

Fig. 3. Example of a large translation for discriminative analysis. The white contour represents the consensus masks, while the red dots indicate the translation masks.

2.3.2.3. Selection of more relevant features. Six different feature selection techniques were assessed after balancing the training set (see Section 2.3.3): principal component analysis (PCA), recursive feature elimination (RFE), feature selection using random forest (RF), least absolute shrinkage and selection operator (LASSO), feature selection using gradient-boosted decision tree (GBDT) and minimum redundancy maximum relevance (MRMR). Specifically, we explored varying numbers of selected features (5, 10, 15, and 20) using RFE, RF, GBDT, and MRMR. For PCA, dimensionality reduction was examined with 5, 10, 15, and 20 components, retaining 56%, 67%, 73%, and 78% of variance, respectively, for the feature subset including all the features and 56%, 68%, 74% and 79% of variance, respectively, for the feature subset without SS features. For LASSO, features were selected based on non-zero coefficients with a penalty value of 0.1.

2.3.3. Balancing the training set

To address the imbalanced training dataset, we first applied a data augmentation technique introduced in [26] consisting in the application of ROI perturbations on the 225 malignant nodules of the training set: 15% dilation and erosion, as well as random contour adjustments. Fig. 4 illustrates the augmentation process, including erosion (a), dilation (b), and contour randomization (c). As a result, the size of the minority class in the training set quadruplicated (from 225 to 900). Afterwards, a random undersampling approach was adopted to reduce the non-malignant nodules (majority class) from 1725 to 900, to match the number of nodules in the minority class post-augmentation. Random undersampling was opted in our methodology since it has shown to perform well in the literature according to [27].

2.3.4. Model development

For each subset of selected features, six machine learning models were trained and validated: random forest (RF), AdaBoost (ABoost), XGBoost (XGB), decision tree (DT), K-nearest neighbors (KNN), and logistic regression (LR). These algorithms were chosen based on the studies mentioned in Table 1 and their extensive utilization in the field of medical image analysis and radiomics in oncology, aligning with

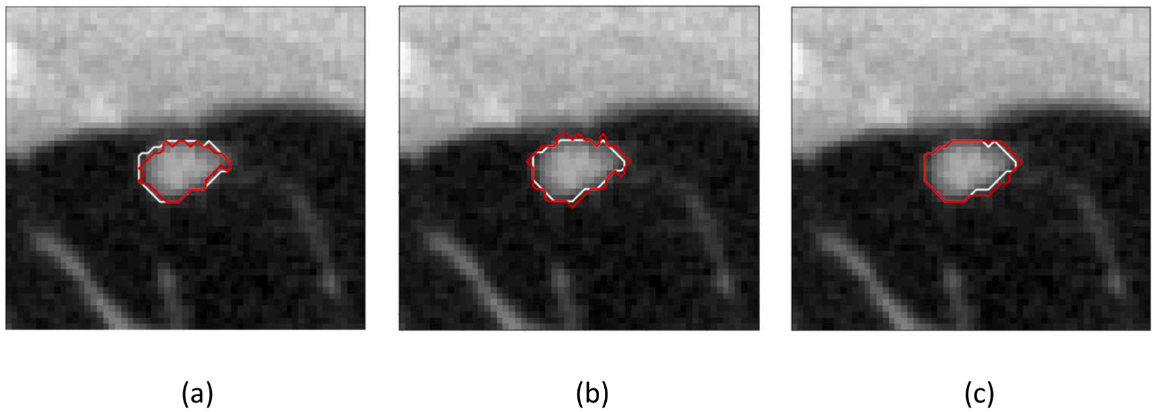

Fig. 4. Perturbation technique applied to the minority class for data augmentation: (a) erosion, (b) dilation, (c) random contour. The white contour represents the original consensus mask, while the red represents the perturbation.

Parmar et al.'s review [28]. For hyperparameter tunning, grid search was used. Details of the parameter grid for each model can be found in the Supplementary Material. During the validation phase, the optimal threshold on the ROC curve was determined by the minimum Euclidean distance to the point (0,1). Bootstrapping was performed 1000 times to obtain the mean value and the 95% confidence interval (CI) for each metric. Model development was implemented using the Scikit-Learn Python library [29].

Finally, the developed model was evaluated on the test set and SHAP analysis (SHapley Additive exPlanations) [30] was used to explain the predictions.

3. Results

The analysis focused on the three different feature sets. To determine the optimal number of features for each feature selection method combined with the different classifiers, we selected the configuration that yielded the highest area under the ROC curve (AUC-ROC) value on the validation set (see Supplementary Material). Additionally, sensitivity, specificity, balanced accuracy and geometric mean were computed to evaluate models' performances on the validation and test set after applying the optimal threshold on the ROC curve determined during the validation phase.

3.1. Model using only SS features

When considering only the SS features, three features were identified as stable, discriminant and non-redundant: *minor_axis_length, sphericity* and *elongation*. Consequently, further feature selection was not required. In this scenario (Fig. 5), RF and LR achieved similar scores on the validation set, with RF obtaining balanced accuracy, sensitivity, specificity and AUC-ROC scores of 0.80, 0.76, 0.83 and 0.88, respectively, and LR yielding 0.80, 0.76, 0.84 and 0.87 for the same metrics, respectively.

Considering the simplicity of LR compared to RF, LR was evaluated on the test set and achieved balanced accuracy, sensitivity, specificity and AUC-ROC scores of 0.82, 0.74, 0.89 and 0.82, respectively. SHAP analysis (Fig. 6) on LR revealed that *minor_axis_length*, corresponding to the second-largest axis length of the ROI-enclosing ellipsoid, was the most important feature, with lower values (indicating smaller nodules) associated to reduced predicted malignancy risk. The second most significant feature was *sphericity* which measures the roundness of the shape of the tumor region relative to a sphere. Higher *sphericity* values (indicating rounder nodules) were associated with reduced predicted malignancy risk. Additionally, *elongation* measures the relationship between the two largest principal components in the ROI shape. Higher *elongation* values suggested a more circle-like nodule shape, correlating with a lower likelihood of malignancy.

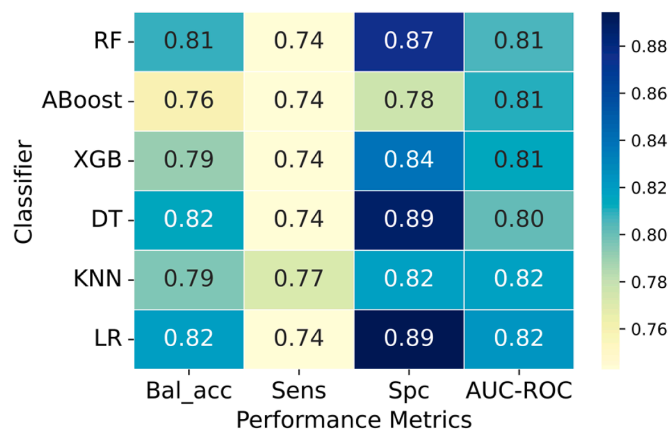

Fig. 5. - Heatmap showing performance metrics (balanced accuracy (bal_acc), sensitivity (sens), specificity (spc), AUC-ROC) on the validation set for different classifiers (random forest (RF), AdaBoost (ABoost), XGBoost (XGB), decision tree (DT), k-nearest neighbors (KNN), logistic regression (LR)) using only SS features.

3.2. Model without SS features

When using all features but SS features, 111 features resulted stable, discriminative, and non-redundant. Subsequently, the selection of more relevant features methods was performed, leading to a varying number of features utilized for model training. Fig. 7 depicts heatmaps of sensitivity and balanced accuracy on the validation set for each feature selection and classifier combination considering the optimal number of selected features. The best performance was obtained by LR trained with the 20 features selected by RF, being associated with a balanced accuracy of 0.79, a sensitivity of 0.82, specificity of 0.76, and AUC-ROC of 0.83. Evaluating the same model on the test set reported balanced accuracy, sensitivity, specificity and AUC-ROC scores of 0.60, 0.43, 0.77 and 0.53, respectively. Given these results on the test set, which were insufficient for meaningful interpretation, SHAP analysis was not performed.

3.3. Model using all features

Considering the set including all the 851 radiomic features, 113 were considered stable, discriminative, and non-redundant, thus undergoing further selection of more relevant features. Fig. 8 depicts heatmaps of sensitivity and balanced accuracy on the validation set. The best performance on the validation set was obtained with the combination of MRMR feature selection (5 features) and ABoost classifier, leading to a

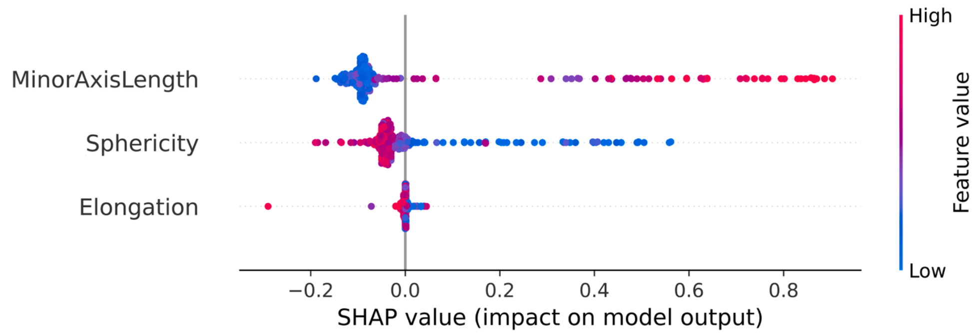

Fig. 6. Beeswarm plot of the test set generated from SHAP analysis of the developed model using logistic regression.

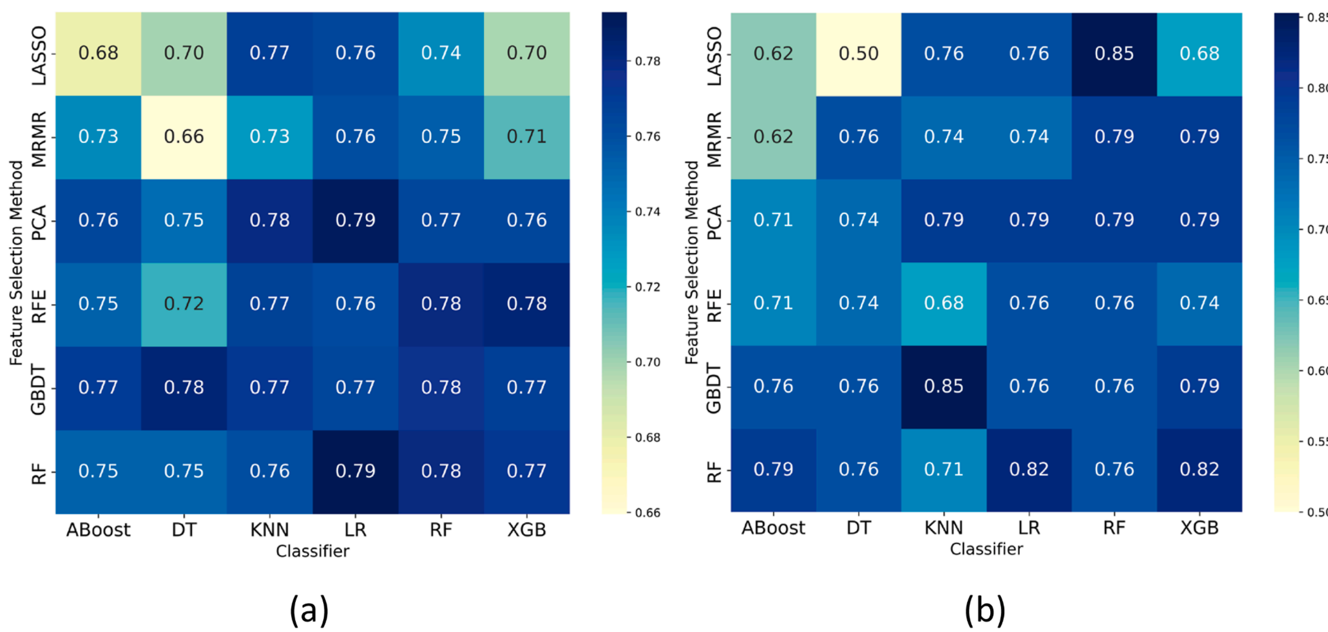

Fig. 7. Heatmaps of the (a) balanced accuracy and (b) sensitivity obtained during the validation phase considering different combinations of feature selection and classifiers for models using all but SS features.

balanced accuracy of 0.85, sensitivity of 0.85, specificity of 0.85 and AUC-ROC of 0.87. Although the combination of features selected by MRMR with the LR classifier and RF with the XGB model demonstrated competitive balanced accuracy (0.84 for MRMR with LR and 0.83 for RF with XGB) and sensitivity (0.88 for MRMR with LR and 0.91 for RF with XGB) values, the achieved specificity was relatively lower (0.79 for MRMR with LR and 0.75 for RF with XGB).

Considering the combination MRMR with ABoost, the results obtained on the test set were 0.78 for balanced accuracy, 0.74 for sensitivity, 0.83 for specificity and 0.82 for AUC-ROC. SHAP analysis on this model identified *sphericity, wavelet_LLH_glcm_MCC* (maximal correlation coefficient (MCC)), and *major_axis_length* as the three most impactful features (Fig. 9). *Sphericity* displayed consistent behavior with the model using only SS features, i.e., higher values of *sphericity* (indicating rounder nodules) were found to be associated to lower likelihood of malignancy. *Wavelet_LLH_glcm_MCC* feature, derived from wavelet transformations involving low-pass (L) and high-pass (H) filters, reflects the complexity of the texture within a ROI region. Lower *wavelet_LLH_glcm_MCC* values, indicating more intricate or complex textures, are associated with a higher probability of malignancy. *Major_axis_length*, which represents the largest axis length of the ROI-enclosing ellipsoid, behaves analogously to the *minor_axis_length* feature explained in models using only SS features, where higher values indicate a greater likelihood of malignancy.

3.4. Comparative analysis

Table 3 summarizes the best results obtained using different feature sets during validation and testing, presented as mean values and their 95% CI. When using three SS features, the model achieved consistent results on both validation and test sets, suggesting robustness with a minimal feature set. When considering all but SS features, performance declined notably on the test set compared to validation. When using all features, the classifier achieved the highest performance during validation with a slightly decrease on the test set. These results suggest that using a small subset of SS features yielded robust performance across validation and testing phases, while excluding these features led to a significant decrease (p-value < 0.05 for most of the metrics) in model's performance on the test set. The superior performance of the only SS feature set can be attributed to the radiologists' reliance on these features for nodule scoring in this dataset. For more details on *p*-values, see Supplementary Material.

Additionally, as the original labels ranged from 1 to 5, with nodules having an average score ≥ 4 considered malignant, we sought to determine how classified and misclassified instances in the test set were spread across malignancy levels. This analysis was performed for the model using only SS features. The histogram in Fig. 10 reveals that a significant portion of misclassified nodules belonged to the category with an average score between 3 and 4, representing the most uncertain

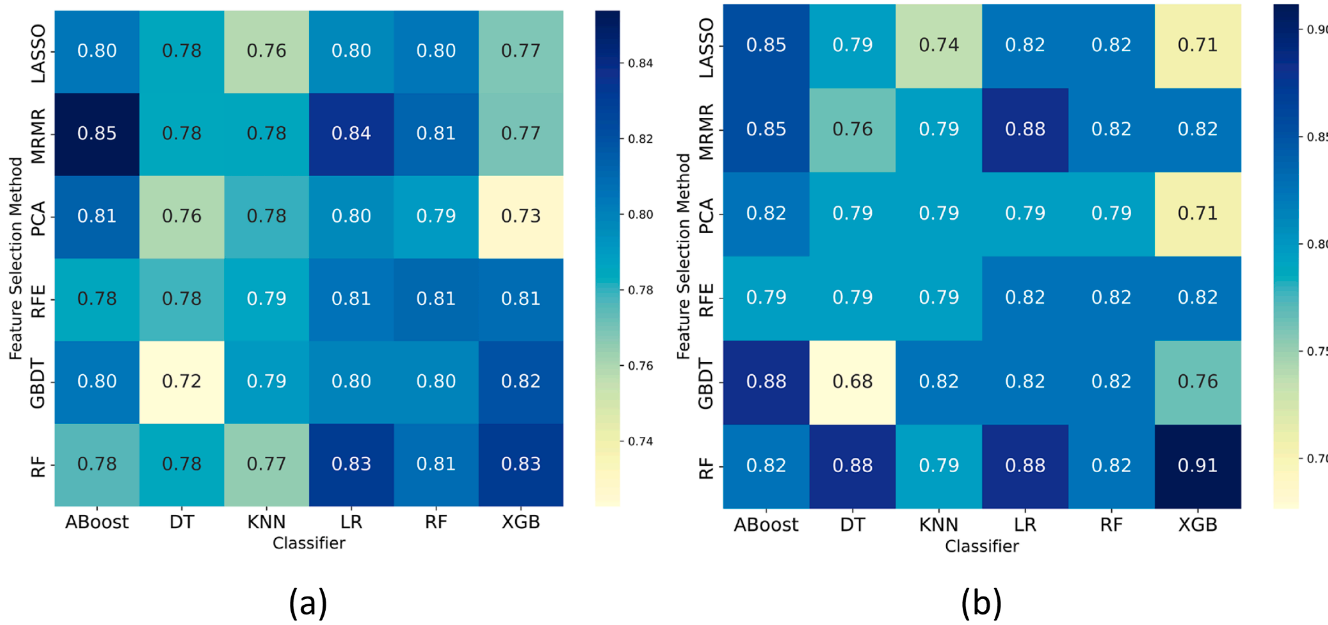

Fig. 8. Heatmaps of the (a) balanced accuracy and (b) sensitivity obtained during validation phase considering different combinations of feature selection and classifiers for models using all features.

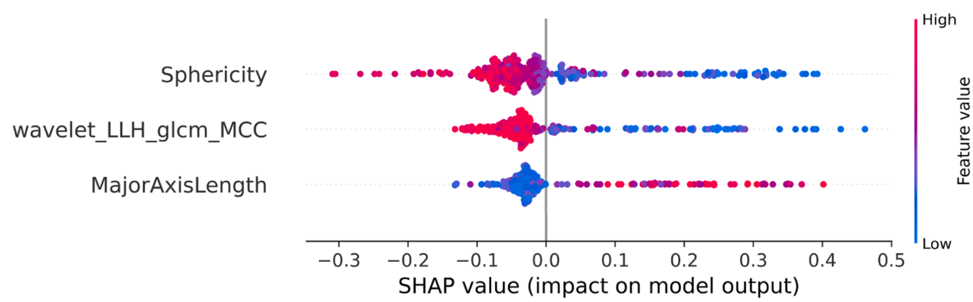

Fig. 9. Beeswarm plot of the predictions made for the instances present in the test set generated from SHAP analysis. It shows the first 3 features selected by MRMR feature selection when used to train an ABoost model.

Table 3
Summary of the best results obtained during validation and the correspondent testing phase for different feature sets using: a) only shape and size (SS) features, b) all features but SS features, and c) including all features. The results, presented as mean (CI), were calculated using 1000 bootstrapping iterations.

Feature set	Bal_acc.	Sen.	Spc.	AUC-ROC	G-mean	#Feat
Only SS features						
validation	0.807 (0.807–0.807)	0.796 (0.793–0.798)	0.819 (0.817- 0.821)	0.874 (0.874–0.875)	0.806 (0.806–0.807)	3
test	0.805 (0.804–0.80)	0.759 (0.757–0.760)	0.850 (0.847 - 0.853)	0.874 (0.874–0.874)	0.802 (0.802–0.803)	
All but SS features						
validation	0.784 (0.783–0.785)	0.796 (0.794–0.799)	0.772 (0.770–0.774)	0.822 (0.821–0.823)	0.783 (0.782–0.784)	20
test	0.664 (0.659–0.669)	0.544 (0.535–0.553)	0.783 (0.780–0.786)	0.822 (0.821–0.823)	0.647 (0.642–0.653)	
All features						
validation	0.826 (0.825 - 0.827)	0.843 (0.841 - 0.845)	0.810 (0.808 - 0.811)	0.865 (0.864 - 0.866)	0.826 (0.825 - 0.826)	5
test	0.769 (0.768 - 0.770)	0.740 (0.738 - 0.742)	0.798 (0.796 - 0.800)	0.865 (0.864 - 0.866)	0.768 (0.767 - 0.769)	

Sen. – sensitivity; Spc. – specificity; AUC-ROC – area under the ROC curve; Bal_acc. – balanced accuracy; G-mean – geometric mean; # Feat. – number of features.

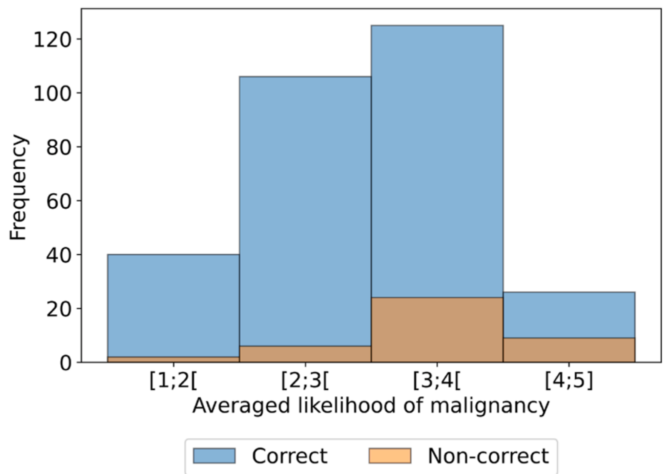

Fig. 10. Histogram representing the original label distribution (average likelihood of malignancy) of the correctly and incorrectly classified instances in the test set.

cases.

4. Discussion and conclusion

In the present study, we developed an explainable radiomics-based model to classify lung nodules using CT images. In particular, we demonstrated the effectiveness of training a classifier from synthetic LDCT, obtained by degrading SDCT images, and testing it on real LDCT images. This strategy enabled us to address the limited availability of annotated LDCT images for training while ensuring the validity of our results through testing on real LDCT images.

Our findings revealed that extracting only three SS features from the lung nodules and using a simple LR classifier, yielded the most consistent results, achieving mean (CI) for balanced accuracy, sensitivity, specificity, AUC-ROC, and geometric-mean scores of 0.805 (0.804–0.805), 0.759 (0.757–0.760), 0.850 (0.847–0.853), 0.874 (0.874–0.874), 0.802 (0.802–0.803), respectively, on the test set. These results were comparable to or exceeded the performance of several studies in the literature, such as Alahmari et al. [11], who used 18 Rider features trained on Random Forest model, obtained sensitivity of 0.49 and specificity of 0.93; and Rundo et al. [12], who used 32 radiomic features trained on logistic regression, obtained a sensitivity of 0.67 and specificity of 0.86. Causey et al. [9] achieved notable results (sensitivity of 0.88 and specificity of 0.88) using the LIDC-IDRI dataset with features obtained from convolutional neural networks. However, when applying only radiomic features the sensitivity dropped to 0.70. Moreover, Causey et al. [9] neglected nodules with an average score for the likelihood of malignancy between 3 and 4 (included in this study), which are the most challenging nodules to classify.

While some models in the literature achieved better performance, they often used smaller sample sizes, for example, Choi et al. [13] considered 72 nodules and developed a support vector machine (SVM) classifier using two radiomic features, Mao et al. [16] extracted 11 radiomic features from each of the 294 nodules and used linear regression model, Garau et al. [14] considered 113 nodules and extracted five radiomic features they trained on SVM classifier, and Liu et al. [17] used 141 nodules to develop a radiomic monogram using three features. We recognize that larger datasets contribute to more stable and non-biased ML models, enhancing reliability and generalizability. Notably, our dataset comprised a larger number of nodules (2625), surpassing by sample sizes all the studies listed in Table 1 (see Fig. 11). Peikert et al. [15] achieved better performance metrics (sensitivity = 0.90, specificity = 0.86) using 8 radiomic features trained with LASSO model with a relatively large dataset, but it is worth noting that their study included larger nodules (≥7 mm) than those analyzed in our study (≥3 mm). To date, radiomic analyses focusing on small lung nodules have not been extensively explored in the literature. We specifically included nodules with a size of ≥ 3 mm, considered optimal for screening and important for achieving earlier diagnosis [31]. In addition, smaller nodules present greater difficulty in classification and are associated with higher rates of false positives [31], emphasizing the need for robust classification models in this context.

Including SHAP analysis in the pipeline allowed us to identify the most impactful features contributing to the classification of the lung nodules. Specifically, *axis_length* and *sphericity* were selected as the top 3 most impactful features both when only SS feature were included in model training and also when all the features were considered. This result is coherent and highlights the importance of SS feature in distinguishing likely benign lung nodules from likely malign ones. Moreover, identifying *minor axis length* as the most important feature in the

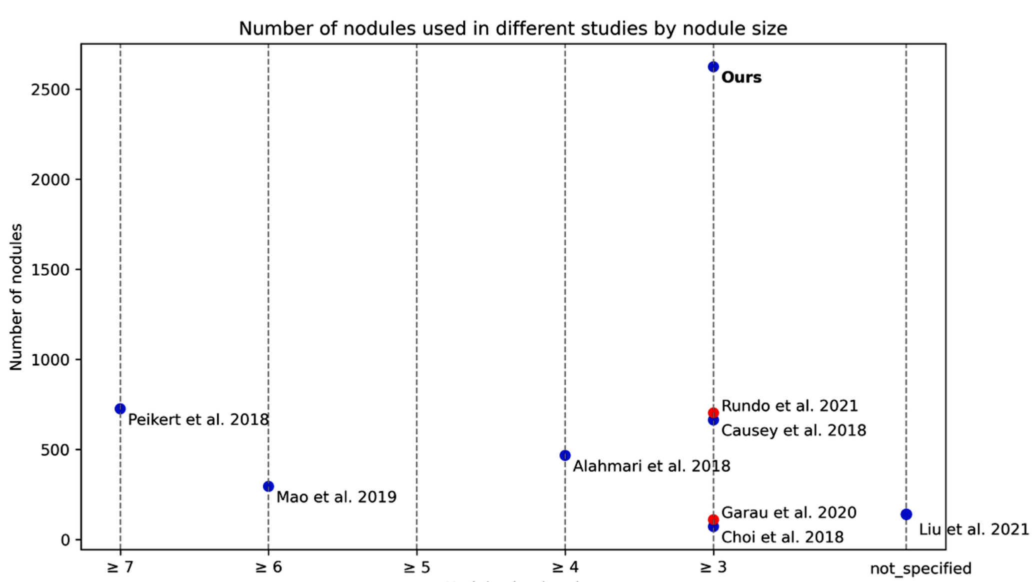

Fig. 11. Distribution of nodule sizes (≥ x mm) and the corresponding number of nodules used across different studies. Each point represents a study, with different color markers used to differentiate studies with similar numbers of nodules analyzed.

best model aligns with one of the criteria used in the Lung Imaging Reporting and Data System (Lung-RADS) developed by the American College of Radiology to standardize the screening of lung cancer on CT images [32].

One limitation of our study was the nature of the dataset, which was not tailored for nodule classification or screening but for developing segmentation methods. Consequently, the nodule labels were based on radiologists' assessments of malignancy likelihood. In this perspective, our study successfully emulated the radiologists' approach to nodule evaluation, proposing a tool to support radiologists' screening activity. In clinical practice, further improvements are needed before it can be considered ready for clinical use. The current study represents a preliminary study aimed at assessing the feasibility of training models using simulated LDCT images for lung nodule classification, overcoming challenges associated with limited annotated LDCT datasets. Additionally, the degradation approach was not further validated, due to the absence of publicly available paired LDCT-SDCT datasets, thus relying on prior validation from Zeng et al. [19]. Although the model's performance on real LDCT data supports the goodness of the degradation method, future studies will include comparison and validation on multiple LDCT simulation techniques.

Finally, although the results of our analysis can be considered robust, being LIDC-IDRI based on a large and multicentric dataset, further validation on external datasets could strengthen our findings.

While an extensive comparison has been made between the proposed model and existing radiomic models for lung nodule classification using screening LDCT images, deep learning approaches were considered out of the scope of this paper. As future work, training deep learning models on degraded SDCT would allow for a more direct comparison with other advanced state-of-the-art methods.

In conclusion, utilizing degraded SDCT images offers a promising avenue for overcoming challenges associated with limited annotated LDCT images. By using both SDCT and synthetic LDCT images for model training, this approach expands the potential for the development of better and more generalizable models, ultimately enhancing patient care by lowering the false positive rate and improving diagnostic accuracy. Additionally, our study underscores the potential of shape and size features to enhance lung nodule classification accuracy.

Data availability

The dataset can be found in: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254.

The best trained model can be found in GitHub: https://github.com/jiayingliu423/LDCT_radiomics.

Funding

AC is funded by the National Plan for NRRP Complementary Investments (PNC, established with the decree-law 6 May 2021, n. 59, converted by law n. 101 of 2021) in the call for the funding of research initiatives for technologies and innovative trajectories in the health and care sectors (Directorial Decree n. 931 of 06-06-2022) - project n. PNC0000003 - AdvaNced Technologies for Human-centrEd Medicine (project acronym: ANTHEM). This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

CRediT authorship contribution statement

Jiaying Liu: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Anna Corti:** Writing – review & editing, Visualization, Investigation, Formal analysis. **Valentina D.A. Corino:** Writing – review & editing, Visualization, Supervision, Investigation, Formal analysis, Conceptualization. **Luca Mainardi:** Writing – review & editing, Visualization, Supervision, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2024.108474.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA Cancer J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.

[2] R. De Angelis, M. Sant, M.P. Coleman, S. Francisci, P. Baili, D. Pierannunzio, A. Trama, O. Visser, H. Brenner, E. Ardanaz, M. Bielska-Lasota, G. Engholm, A. Nennecke, S. Siesling, F. Berrino, R. Capocaccia, H. Amash, C. Amati, F. Di Salvo, S. Bonfarnuzzo, L. Botta, R. Foschi, G. Gatta, C. Margutti, P. Minicozzi, M. Caldora, E. Carrani, S. Mallone, P. Roazzi, S. Rossi, M. Santaquilani, A. Tavilla, M. Hackl, N. Zielonke, W. Oberaigner, K. Henau, E. Van Eycken, N. Dimitrova, Z. Valerianova, M. Sekerija, L. Dušek, M. Zvolský, H. Storm, M. Magi, T. Areleid, N. Malila, K. Seppa, M. Velten, A. Belot, V. Bouvier, X. Troussard, G. Launoy, A. V. Guizard, A.M. Bouvier, J. Faivre, P. Arveux, M. Maynadie, A.S. Woronoff, M. Robaszkiewicz, I. Baldi, A. Monnereau, B. Tretarre, N. Bossard, M. Colonna, F. Molinié, S. Bara, C. Schvartz, J. Clavel, E. Desandes, B. Lacour, B. Lapôtre-Ledoux, P. Grosclaude, M. Grosclaude, R. Stabenow, A. Eberle, S. Luttmann, J. Engel, G. Schubert-Fritschle, P. Kaatsch, B. Holleczek, A. Katalinic, M. Garami, Z. Jakab, J. Jónasson, L. Tryggvadottir, H. Comber, A. Bulatko, G. Mazzoleni, C. Buzzoni, A. Giacomin, P. Mancuso, A. Mazzei, A. Sutera Sardo, A. Caldarella, E. Crocetti, G. Manneschi, L. Dal Maso, D. Serraino, A. Zucchetto, S. Busco, M. Natali, F. Pannozzo, V. Gennaro, P. Ricci, C. Pascucci, M. Auteliano, G. Randi, M. Federico, M. Fusco, M.F. Vitale, M. Usala, R. Cusimano, B. Ravazzolo, F. Vitale, R. Staiti, A. Traina, M. Zarcone, M. Michiara, M. Maule, F. Merletti, R. Tumino, E. Di Felice, P. Giorgi Rossi, M. Vicentini, F. Falcini, L. Bonelli, M. Vercelli, A. Iannelli, R. Cesaraccio, O. Sechi, S. Piffer, M.L. Contrino, A. Madeddu, F. Tisanu, A.C. Fanetti, S. Maspero, S. Rosso, R. Zanetti, P. Candela, T. Scuderi, F. Bianconi, F. La Rosa, F. Stracci, P. Contiero, G. Tagliabue, P. Dei Tos, S. Guzzinati, S. Pildava, G. Smailyte, D. Agius, N. Calleja, R. Micallef, T.B. Johannesen, S. Góźdź, R. Mężyk, J. Rachtan, M. Bębenek, J. Błaszczyk, K. Kępska, L. Antunes, M.J. Bento, C. Castro, A. Mayer-Da-Silva, A. Miranda, D. Coza, C. Safaei Diba, M. Primic-Zakeli, E. Almar, A. Mateos, C. Ramirez, J. Bidaurrazaga, N. Larrañaga, A. Torrella-Ramos, M. Vicente Raneda, J.M. Díaz García, R. Jimenez-Chillaro, A.I. Marcos-Navarro, A. Izquierdo Font, R. Marcos-Gragera, C. Martinez, M.J. Sanchez, C. Navarro, M. D. Chirlaque, C. Moreno-Iribas, S. Felipe García, E. Pardo Romaguera, R. Peris-Bonet, M. Carulla, J. Galceran, S. Khan, M. Lambe, M. Mousavi, C. Bouchardy, M. Usel, H. Frick, C. Herrmann, M. Lorez, S.M. Ess, A. Bordoni, A. Spitale, I. Konzelmann, K. Aben, P. Huijgens, V. Lemmens, R. Verhoeven, R. Otter, C. Allemani, B. Racheten, N. Easey, G. Lawrence, T. Moran, J. Rashbass, M. Roche, J. Verne, J. Wilkinson, C. Stiller, L. Anderson, F. Bannon, A. Gavin, D. Brewster, D. W. Huws, C. White, Cancer survival in Europe 1999–2007 by country and age: results of EUROCARE-5—A population-based study, Lancet Oncol. 15 (2014) 23–34, https://doi.org/10.1016/S1470-2045(13)70546-1.

[3] National Lung Screening Trial Research Team, Reduced lung-cancer mortality with low-dose computed tomographic screening, New Engl. J. Med. 365 (2011) 395–409, https://doi.org/10.1056/NEJMoa1102873.

[4] U. Pastorino, M. Silva, S. Sestini, F. Sabia, M. Boeri, A. Cantarutti, N. Sverzellati, G. Sozzi, G. Corrao, A. Marchianò, Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy, Ann. Oncol. 30 (2019) 1162–1169, https://doi.org/10.1093/annonc/mdz169.

[5] H.J. de Koning, C.M. van der Aalst, P.A. de Jong, E.T. Scholten, K. Nackaerts, M. A. Heuvelmans, J.-W.J. Lammers, C. Weenink, U. Yousaf-Khan, N. Horeweg, S. van 't Westeinde, M. Prokop, W.P. Mali, F.A.A. Mohamed Hoesein, P.M.A. van Ooijen, J.G.J.V. Aerts, M.A. den Bakker, E. Thunnissen, J. Verschakelen, R. Vliegenthart, J. E. Walter, K. ten Haaf, H.J.M. Groen, M. Oudkerk, Reduced lung-cancer mortality with volume CT screening in a randomized trial, New Engl. J. Med. 382 (2020) 503–513, https://doi.org/10.1056/NEJMoa1911793.

[6] A. Khawaja, B.J. Bartholmai, S. Rajagopalan, R.A. Karwoski, C. Varghese, F. Maldonado, T. Peikert, Do we need to see to believe?—Radiomics for lung nodule classification and lung cancer risk stratification, J. Thorac. Dis. 12 (2020) 3303, https://doi.org/10.21037/JTD.2020.03.105.

[7] J.D. Shur, S.J. Doran, S. Kumar, D. Ap Dafydd, K. Downey, J.P.B. O'connor, N. Papanikolaou, C. Messiou, D.M. Koh, M.R. Orton, Radiomics in oncology: a practical guide, Radiographics. 41 (2021) 1717–1732, https://doi.org/10.1148/rg.2021210037.

[8] T. Peikert, B.J. Bartholmai, F. Maldonado, Radiomics-based Management of Indeterminate Lung Nodules? Are We There Yet? Am. J. Respir. Crit. Care Med. 202 (2020) 167–169, https://doi.org/10.1164/rccm.202004-1279ED.
[9] J.L. Causey, J. Zhang, S. Ma, B. Jiang, J.A. Qualls, D.G. Politte, F. Prior, S. Zhang, X. Huang, Highly accurate model for prediction of lung nodule malignancy with CT scans, Sci. Rep. 8 (2018), https://doi.org/10.1038/S41598-018-27569-W.
[10] W. Quanyang, H. Yao, W. Sicong, Q. Linlin, Z. Zewei, H. Donghui, L. Hongjia, Z. Shijun, Artificial intelligence in lung cancer screening: detection, classification, prediction, and prognosis, Cancer Med. 13 (2024), https://doi.org/10.1002/CAM4.7140.
[11] S.S. Alahmari, D. Cherezov, D.B. Goldgof, L.O. Hall, R.J. Gillies, M.B. Schabath, Delta radiomics improves pulmonary nodule malignancy prediction in lung cancer screening, IEEe Access. 6 (2018) 77796–77806, https://doi.org/10.1109/ACCESS.2018.2884126.
[12] L. Rundo, R.E. Ledda, C. Di Noia, E. Sala, G. Mauri, G. Milanese, N. Sverzellati, G. Apolone, M.C. Gilardi, M.C. Messa, I. Castiglioni, U. Pastorino, A low-dose CT-based radiomic model to improve characterization and screening recall intervals of indeterminate prevalent pulmonary nodules, Diagnostics 11 (2021) 1610, https://doi.org/10.3390/diagnostics11091610.
[13] W. Choi, J.H. Oh, S. Riyahi, C.J. Liu, F. Jiang, W. Chen, C. White, A. Rimner, J. G. Mechalakos, J.O. Deasy, W. Lu, Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer, Med. Phys. 45 (2018) 1537–1549, https://doi.org/10.1002/mp.12820.
[14] N. Garau, C. Paganelli, P. Summers, W. Choi, S. Alam, W. Lu, C. Fanciullo, M. Bellomi, G. Baroni, C. Rampinelli, External validation of radiomics-based predictive models in low-dose CT screening for early lung cancer diagnosis, Med. Phys. 47 (2020) 4125–4136, https://doi.org/10.1002/mp.14308.
[15] T. Peikert, F. Duan, S. Rajagopalan, R.A. Karwoski, R. Clay, R.A. Robb, Z. Qin, J. R. Sicks, B.J. Bartholmai, F. Maldonado, Novel high-resolution computed tomography-based radiomic classifier for screen-identified pulmonary nodules in the National Lung Screening Trial, PLoS One 13 (2018), https://doi.org/10.1371/JOURNAL.PONE.0196910.
[16] L. Mao, H. Chen, M. Liang, K. Li, J. Gao, P. Qin, X. Ding, X. Li, X. Liu, Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening, Quant. Imaging Med. Surg. 9 (2019) 263–272, https://doi.org/10.21037/qims.2019.02.02.
[17] J. Liu, H. Xu, H. Qing, Y. Li, X. Yang, C. He, J. Ren, P. Zhou, Comparison of radiomic models based on low-dose and standard-dose CT for prediction of adenocarcinomas and benign lesions in solid pulmonary nodules, Front. Oncol. 10 (2021) 1, https://doi.org/10.3389/fonc.2020.634298.
[18] S.G. Armato, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, B. Zhao, D.R. Aberle, C.I. Henschke, E.A. Hoffman, E.A. Kazerooni, H. MacMahon, E.J.R. Van Beek, D. Yankelevitz, A.M. Biancardi, P.H. Bland, M.S. Brown, R. M. Engelmann, G.E. Laderach, D. Max, R.C. Pais, D.P.Y. Qing, R.Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G.W. Gladish, C.M. Jude, R. F. Munden, I. Petkovska, L.E. Quint, L.H. Schwartz, B. Sundaram, L.E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. Vande Casteele, S. Gupte, M. Sallam, M.D. Heath, M.H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B.Y. Croft, L.P. Clarke, The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a Completed Reference Database of Lung Nodules on CT Scans, Med. Phys. 38 (2011) 915, https://doi.org/10.1118/1.3528204.
[19] D. Zeng, J. Huang, Z. Bian, S. Niu, H. Zhang, Q. Feng, Z. Liang, J. Ma, A Simple Low-dose X-ray CT Simulation from High-dose Scan, IEEE Trans. Nucl. Sci. 62 (2015) 2226, https://doi.org/10.1109/TNS.2015.2467219.
[20] J. Ma, Z. Liang, Y. Fan, Y. Liu, J. Huang, W. Chen, H. Lu, Variance analysis of x-ray CT sinograms in the presence of electronic noise background, Med. Phys. 39 (2012) 4051–4065, https://doi.org/10.1118/1.4722751.
[21] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository, J. Digit. Imaging 26 (2013) 1045–1057, https://doi.org/10.1007/S10278-013-9622-7.
[22] C. Rampinelli, D. Origgi, M. Bellomi, REVIEW Low-dose CT: technique, reading methods and image interpretation, Cancer Imaging 12 (2012) 548–556, https://doi.org/10.1102/1470-7330.2012.0049.
[23] T. Snowsill, H. Yang, E. Griffin, L. Long, J. Varley-Campbell, H. Coelho, S. Robinson, C. Hyde, Low-dose computed tomography for lung cancer screening in high-risk populations: a systematic review and economic evaluation, Health Technol. Assess. (Rockv) 22 (2018) 1–276, https://doi.org/10.3310/HTA22690.
[24] J.J.M. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.C. Fillion-Robin, S. Pieper, H.J.W.L. Aerts, Computational radiomics system to decode the radiographic phenotype, Cancer Res. 77 (2017) e104, https://doi.org/10.1158/0008-5472.CAN-17-0339.
[25] T.K. Koo, M.Y. Li, Cracking the code: providing insight into the fundamentals of research and evidence-based practice a guideline of selecting and reporting intraclass correlation coefficients for reliability research, (2016). https://doi.org/10.1016/j.jcm.2016.02.012.
[26] F. Lo Iacono, R. Maragna, G. Pontone, V.D.A. Corino, A novel data augmentation method for radiomics analysis using image perturbations, J. Imaging Inf. Med. 2024 (2024) 1–14, https://doi.org/10.1007/S10278-024-01013-0.
[27] K.M. Hasib, M.S. Iqbal, F.M. Shah, J. Al Mahmud, M.H. Popel, M.I.H. Showrov, S. Ahmed, O. Rahman, A survey of methods for managing the classification and solution of data imbalance problem, J. Comput. Sci. 16 (2020) 1546–1557, https://doi.org/10.3844/JCSSP.2020.1546.1557.
[28] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, H.J.W.L. Aerts, Machine learning methods for quantitative radiomic biomarkers, Sci. Rep. 5 (1) (2015) 1–11, https://doi.org/10.1038/srep13087, 5 (2015).
[29] F. Pedregosa FABIANPEDREGOSA, V. Michel, O. Grisel OLIVIERGRISEL, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot, Scikit-learn: machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot, J. Mach. Learn. Res. 12 (2011) 2825–2830. http://scikit-learn.sourceforge.net. accessed March 9, 2024.
[30] S.M. Lundberg, P.G. Allen, S.-I. Lee, A Unified approach to interpreting model predictions, in: Conference on Neural Information Processing Systems, 2017. https://github.com/slundberg/shap. accessed March 7, 2024.
[31] M. Silva, G. Milanese, R.E. Ledda, U. Pastorino, N. Sverzellati, Screen-detected solid nodules: from detection of nodule to structured reporting, Transl. Lung Cancer Res. 10 (2021) 2335, https://doi.org/10.21037/TLCR-20-296.
[32] American College of Radiology Committee on Lung-RADS, Lung-RADS Assessment Categories 2022, (2022). https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/Lung-RADS-2022.pdf (accessed May 8, 2024).