QUANTITATIVE TEXTURE ANALYSIS USING MACHINE LEARNING FOR PREDICTING LUNG FUNCTION IN PULMONARY EMBOLISM PATIENTS

LI ZIHAN

MPhil

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

Quantitative Texture Analysis Using Machine Learning for Predicting Lung Function in Pulmonary Embolism Patients

LI Zihan

A thesis submitted in partial fulfilment of the requirements for

the degree of Master of Philosophy

July 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ **(Signed)**

Li Zihan **(Name of student)**

Abstract

Background: Pulmonary embolism (PE) is life-threatening and requires timely and accurate diagnosis, yet current imaging methods, like computed tomography pulmonary angiography, present limitations, particularly for patients with contraindications to iodinated contrast agents.

Purpose: We aimed to develop a quantitative texture analysis pipeline using machine learning (ML) based on non-contrast thoracic computed tomography (CT) scans to discover intensity and textural features correlated with regional lung perfusion (Q) physiology and pathology and synthesize voxel-wise Q surrogates to assist in PE diagnosis.

Methods and Materials: We retrospectively collected ^{99m}Tc-labeled macroaggregated albumin Q-SPECT/CT scans from patients suspected of PE, including an internal dataset of 76 patients (64 for training, 12 for testing) and an external testing dataset of 49 patients. Quantitative CT features were extracted from segmented lung subregions and underwent a two-stage feature selection pipeline. The prior-knowledge-driven preselection stage screened for robust and non-redundant perfusion-correlated features, while the data-driven selection stage further filtered features by fitting ML models for classification. The final classification model, trained with the highest-performing PE-associated feature combination, was evaluated in the testing cohorts based on the Area Under the Curve (AUC) for subregion-level predictability. The voxel-wise Q surrogate was then synthesized using the final selected feature maps (FMs) and model score maps (MSMs) to investigate spatial distributions. The Spearman correlation coefficient (SCC) and Dice similarity coefficient (DSC) were used to assess the spatial consistency between FMs or MSMs and Q-SPECT scans.

Results: The optimal model performance achieved an AUC of 0.863 during internal testing and 0.828 on the external testing cohort. The model identified a combination containing 14 intensity and textural features that were non-redundant, robust, and capable of distinguishing between high- and low-functional lung regions. Spatial consistency assessment in the internal testing cohort showed moderate-to-high agreement between MSMs and reference Q-SPECT scans, with median SCC of 0.66, median DSCs of 0.86 and 0.64 for high- and low-functional regions, respectively.

Conclusions: This study validated the feasibility of using quantitative texture analysis and a data-driven ML pipeline to generate voxel-wise lung perfusion surrogates, providing a radiation-free, widely accessible alternative to functional lung imaging in managing pulmonary vascular diseases.

Acknowledgments

Since the summer of 2023, when I first traveled over 3,000 kilometers from the north to the south across China to begin my postgraduate journey, I have imagined countless times what it would feel like to write this acknowledgment. I have often thought I would be filled with emotions such as gratitude, fulfillment, perhaps even reluctance. Yet as I typed the final word of this thesis, all that remained was an overwhelming sense of appreciation.

First and foremost, I would like to express my sincere gratitude to my chief supervisor, Dr. Ge Ren, and my co-supervisor, Prof. Jing Cai. Thank you for giving me the opportunity to pursue this degree, and for your unwavering support and trust throughout the past two years. I have never considered myself naturally gifted in research, and even after two decades of education, I still find myself needing encouragement and affirmation to keep moving forward. During moments of setbacks in my experiments, your discussions and thoughtful guidance provided both comfort and clarity. Beyond academic supervision, I am deeply grateful for your advice on my future path and for teaching me valuable lessons about life and character.

My heartfelt thanks also go to the staff of the Department of Health Technology and Informatics at The Hong Kong Polytechnic University for their support throughout my studies.

I am truly thankful to Senior Yu-Hua Huang for your patient guidance in my experiments and for helping me through emotional lows. I have always remembered our many heartfelt conversations. I am equally grateful to my team members and peers whose help and companionship reminded me that I was never alone on this long journey.

To my family, thank you for 23 years of unconditional love and support. You once told me that Mom and Dad will always be my rock, and that gave me both a safe haven and the courage to face any challenge. It is because of your efforts that I have been able to witness a world you never had the chance to see. Thank you for allowing me to be my most vulnerable, selfish, flawed, and truest self, and for loving me all the same.

To my dearest friends, Jingyi Zhang and Yuqi Huang, your friendship is one of my most treasured gifts. It has been a source of strength that carried me through difficult times. Thank you for walking alongside me as I slowly and clumsily grow into a better version of myself.

Finally, I want to thank the version of myself who, despite all the hardships and setbacks, never gave up. You have weathered every storm and reached the other side. Well done.

Publications

A. Peer-reviewed Journal Articles

1. **Li, Z.***, Zhao, M.*, Li, Z., Huang, Y. H., Chen, Z., Pu, Y., Zhao, M., Liu, X., Wang, M., Wang, K., Yeung, H. Y., Geng, L., Cai, J., Zhang, W., Yang, R., &Ren, G. (2024). Quantitative texture analysis using machine learning for predicting interpretable pulmonary perfusion from non-contrast computed tomography in pulmonary embolism patients. *Respiratory Research*, *25*(1), 1-15. DOI: 10.1186/s12931-024-03004-9

   **These authors contributed equally to this work.*
2. Huang, Y. H., **Li, Z.**, Xiong, T., Chen, Z., Teng, X., Li, B., Lou, Z., Dong, Y., Ma, Z., Ge, H., Ren, G., &Cai, J. (2025). Constructing Surrogate Lung Ventilation Maps from 4DCT-derived Subregional Respiratory Dynamics. *International Journal of Radiation Oncology* Biology* Physics,* 121(5), 1328-1338. DOI: 10.1016/j.ijrobp.2024.11.074
3. Chen, Z., **Li, Z.**, Huang, Y. H., Teng, X., Zhang, J., Xiong, T., Dong, Y., Song, L., Ren, G., &Cai, J. (2025). Anatomy-wise lung ventilation imaging for precise functional lung avoidance radiation therapy. *Physics in Medicine & Biology* DOI: 10.1088/1361-6560/adb123
4. Li, Z., **Li, Z.**, Lam, S.K., Wang, X., Wang, P., Song, L., Lee, F.K.-H., Yip, C.W.-Y., Cai, J., &Li, T. (2025). A Multi-Modal Deep Learning Approach for Predicting Eligibility for Adaptive Radiation Therapy in Nasopharyngeal Carcinoma Patients. *Cancers*, 17, 2350. DOI: 10.3390/cancers17142350

5. Pu, Y., Zhu, Y., Qu, J., Wang, X., Li, W., Zhao, M., Yu, P., **Li, Z.**, Teng, X., Zhang, X., Zhang, J., Peng, T., Cai, J., & Ren, G. (2025). Revealing New Possibilities for Breast MRI Enhancement: Mamba-Driven Cross-Attention GAN with VMKANet. *International Workshop on Computational Mathematics Modeling in Cancer Analysis*, 150-159. DOI: 10.1007/978-3-032-06624-4_16

B. Peer-revied Conference Abstract and Presentations

1. **Li, Z**., Huang, Y. H., Chen, Z., Li, B., Ge, H., Cai, J., &Ren, G. (2025). Enhanced Lung Function Assessment through Machine Learning Analysis of 4DCT Subregional Respiratory Dynamics. *2025 AAPM 67th Annual Meeting*, Washington DC, United State, presenting author of the oral presentation.

2. **Li, Z.**, Huang, Y. H., Chen, Z., Li, B., Ge, H., Cai, J., &Ren, G. (2025). Temporal analysis of 4DCT subregional respiratory dynamics based on machine learning for lung function assessment. *European Society for Radiotherapy and Oncology (ESTRO) 2025 Annual Meeting, Vienna, Austria*, presenting author of the digital poster.

3. **Li, Z.**, Huang, Y. H., Pu, Y., Zhao, M., Wang, K., Yang, R., Cai, J., &Ren, G. (2024). Quantitative Texture Analysis Using Machine Learning for Predicting Interpretable Pulmonary Perfusion from Non-contrast Computed Tomography in Pulmonary Embolism Patients. *32nd Annual Scientific Meeting of Hong Kong College of Radiologists (HKCR), Hong Kong SAR*, presenting author of the oral presentation.

Table of Contents

List of Tables

List of Figures

List of Acronyms

^{99m}Tc-MAA	^{99m}Tc-labeled macroaggregated albumin
AUC	Area Under the Curve
CI	Confidence intervals
CTPA	Computed tomography pulmonary angiography
CTPI	Computed tomography-based perfusion imaging
CTVI/PI	CT-based ventilation and perfusion imaging
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DIR	Deformable image registration
DL	Deep learning
DSC	Dice similarity coefficient
DSC_{HI}	DSC for high functional region
DSC_{LO}	DSC for low functional region
DVF	Displacement vector field
ERS	European Respiratory Society
ES	Effect size
FM	Feature map
GLCM	Grey Level Co-occurrence Matrix
GLDM	Grey Level Dependence Matrix
GLRLM	Grey Level Run Length Matrix
GLSZM	Grey Level Size Zone Matrix

HU	Hounsfield Unit
IBSI	Imaging Biomarker Standardization Initiative
ICC	Intraclass correlation coefficient
IJF	Integrated Jacobian Formulation
IQR	Interquartile range
IRB	Institutional Review Board
LoG	Laplacian of Gaussian
LR	Logistic regression
maskSLIC	Masked simple linear iterative clustering
MCVC	Mass Conserving Volume Change
ML	Machine learning
MSM	Model score map
NGTDM	Neighbouring Grey Tone Difference Matrix
NM	Nuclear medicine
PCC	Pearson correlation coefficient
PE	Pulmonary embolism
PUTH	Peking University Third Hospital
Q1	First quartile
Q3	Third quartile
QMH	Queen Mary Hospital
RFE	Recursive feature elimination
ROC	Receiver operating characteristic

ROI	Region-of-interest
SCC	Spearman correlation coefficient
SHAP	SHapley Additive exPlanations
SVM	Support vector machine
V/Q-SPECT	Ventilation/perfusion single-photon emission computed tomography
V/Q-SPECT/CT	Ventilation/perfusion single-photon emission computed tomography with computed tomography

Chapter 1
Introduction

1.1 Pulmonary Embolism

Pulmonary embolism (PE), a major differential diagnosis in patients presenting with chest pain or dyspnea, ranks as the third most common cause of vascular-related mortality globally, preceded only by myocardial infarction and stroke [1]. In PE, the obstruction of the pulmonary artery is typically caused by the accumulation of blood clots (thrombi), air bubbles, fat tissue, or amniotic fluid. This often occurs in individuals with pre-existing conditions such as heart disease, cancer, severe fractures, or during pregnancy [2, 3]. PE is associated with significant mortality, particularly in high-risk patients presenting with shock or cardiac arrest, where the 30-day mortality rate can reach 52-65% [4]. It is estimated that PE affects 300,000 to 600,000 people annually in the United States, resulting in at least 100,000 fatalities [5, 6]. The complex etiology and high mortality rate underscores the critical importance of early detection and the timely implementation of effective treatment, including anticoagulation and mechanical thrombectomy, as delays in the clinical management of PE significantly worsen patient outcomes.

1.2 Clinical Workflow for the Diagnosis and Management of PE

The diagnostic approach to PE follows a structured clinical workflow, as outlined in the 2019 European Respiratory Society (ERS) guidelines for the diagnosis and

management of pulmonary embolism, developed in collaboration with the ERS [7]. The process typically begins with an assessment of the clinical pretest probability, which integrates factors such as physiological parameters (e.g., age, heart rate and oxygen saturation level), clinical symptoms (e.g., hemoptysis and unilateral leg swelling) and relevant medical history (e.g., recent surgery or trauma, venous thromboembolism event and estrogen use). However, diagnosing PE cannot rely solely on clinical evaluation, as the associated signs, symptoms, and laboratory findings are often deceptively nonspecific [8]. Common clinical manifestations such as unexplained dyspnea, pleuritic chest pain, syncope, palpitations, cough, or hemoptysis lack specificity and frequently overlap with those of other cardiopulmonary conditions, thereby complicating accurate clinical evaluation [9]. In suspected cases, clinical probability is typically assessed before thoracic imaging are performed to confirm the presence of PE, as imaging remains a key step in the diagnostic process.

1.3 Imaging Modalities for PE Diagnosis

Among the currently available PE diagnosis, computed tomography pulmonary angiography (CTPA) is regarded as the gold standard, offering detailed visualization of pulmonary vasculature and accurate identification of emboli. However, CTPA has several limitations that restrict its use in certain patient populations. The requirement for iodinated contrast agents renders it unsuitable for patients with impaired renal function or contrast allergies, while its high radiation exposure poses potential risks, particularly for females of childbearing age. Studies have shown that repeated use of ICAs increases the risk of acute

adverse reactions [10]. Ventilation/perfusion single-photon emission computed tomography (V/Q-SPECT) is one of the most widely used imaging modality that provides high diagnostic accuracy, with reported sensitivity and specificity of 94% and 95%, respectively [11]. Previous studies have demonstrated that combining V/Q-SPECT with CT (V/Q-SPECT/CT) enhances diagnostic performance by improving the anatomical localization of perfusion defects [12]. However, V/Q-SPECT is invasive, expensive, and less widely available than other imaging modalities like CT, often leading to longer waiting times for testing, especially in resource-limited hospitals or regions without nuclear medicine (NM) facilities [13].

In addition to mainstream pulmonary function examinations based on contrast agents, several studies have attempted to obtain perfusion information indirectly from more commonly available clinical thoracic scans, with computed tomography-based perfusion imaging (CTPI) gaining the most attention. A category of CTPI techniques employs deformable image registration (DIR) to physically model surrogates of pulmonary blood supply and circulation from non-contrast respiratory-related CT scans (i.e., four-dimensional CT, 4DCT), for restoring static perfusion distribution at the voxel level [14]. These methods typically have a strict and interpretable modeling process, but the prediction is severely affected by image quality (e.g. CT noise and motion artifacts) and DIR results. These limitations can lead to significant variability in perfusion assessments, potentially compromising diagnostic accuracy. Additionally, some deep-learning (DL) approaches have been proposed to directly synthesize pulmonary perfusion images from a single CT image via deep neural networks [15-17]. The application of end-to-end DL techniques,

although innovative and efficient, faces challenges associated with their complex decision-making processes (often referred to as the "black box" problem), and lack of consideration of clinically relevant pathological mechanisms. Overall, although still in the early stages of development, the emergence of CTPI has the potential to reduce the misuse of contrast agents and improve clinical technicians' efficiency.

1.4 Research Aim and Objectives

To address the research gaps of current perfusion imaging techniques, this thesis extends CTPI explorations by harnessing the capabilities of quantitative texture analysis. Texture analysis is capable of extracting high-dimensional features from routine non-contrast CT imaging that are typically not discernible by the human eyes and quantitatively transforming them into imaging biomarkers with explicit mathematical definitions [18, 19]. The primary aim of this thesis is to capture perfusion information from non-contrast CT scans and synthesize voxel-wise perfusion surrogates to assist in PE detection. The specific objectives of the thesis are:

1. To conduct a quantitative texture analysis and data-driven machine learning (ML) pipeline on non-contrast thoracic CT scans to discover both intensity and textural features that are correlated to the underlying regional lung perfusion physiology and pathology.
2. To develop a reliable and interpretable method for predicting voxel-wise lung perfusion surrogates and study the spatial distribution of the perfusion-correlated features and fitted-model outputs.

The proposed method represents a novel perspective on CTPI, providing potentially improved diagnostic reference on PE and other perfusion-related lung diseases.

1.5 Thesis Overview

This thesis is structured to systematically address the aforementioned research objectives. **Chapter 2** provides a comprehensive literature review on existing CT-based ventilation and perfusion imaging (CTVI/PI) techniques, establishing the foundational background for the proposed research. **Chapter 3** addresses the core objectives of the thesis by introducing a novel methodology for extracting perfusion information and generating CTPI as a diagnostic reference for PE. **Chapter 4** details the experimental results, including the visualizations of the predicted perfusion surrogates and the quantitative evaluation of spatial agreement with reference Q-SPECT images. Finally, **Chapter 5** discusses and concludes the thesis by summarizing the key findings, highlighting current limitations, and outlining potential directions for future research in PE diagnosis.

Chapter 2
Literature Review on Existing CT-based Ventilation/Perfusion Imaging Techniques

In clinical practice, CTVI/PI techniques aim to generate surrogate maps that provide functional information essential for diagnosis and treatment planning. Over the past two decades, a broad spectrum of methodologies has been developed to derive such functional maps, with varying degrees of accuracy, robustness, and clinical applicability. This chapter discusses and reviews the evolution of CTVI/PI algorithms, starting from conventional DIR-based methods based on the estimation of displacement vector fields (DVFs), and then transitioning to more recent developments, including deep learning-based and other emerging methods.

2.1 Conventional DIR-based Method

2.1.1 Classical Algorithms

Classical DIR-based CTVI algorithms are primarily divided into two categories: intensity-based methods and Jacobian-based methods. Both CTVI approaches necessitate image segmentation to define the lung volume and utilize DIR to establish a spatial transformation between the geometries of the lungs during inhalation and exhalation. Intensity or Hounsfield Unit (HU) methods estimate volume changes based on the

variations in HU values between spatially corresponding voxels during inhalation and exhalation [20, 21].HU represents material density on a linear scale, where −1000 HU corresponds to air and 0 HU corresponds to water or tissue. Intensity-based ventilation estimation methods are mathematically based on the assumption that lung tissue can be physically modeled as a linear combination of air and tissue components, a method initially proposed by Simon when calculating specific lung compliance [22].

Simon's method estimates the air fraction within a specific lung region of a CT image using voxel-wise HU values. The air fraction F_{air} at a given voxel can be calculated as

$$F_{air} = -\frac{HU}{1000} \tag{2.1}$$

The fractional change in air content within a specified volume is then

$$\frac{\Delta V}{V} = \frac{F_2 - F_1}{F_1(1 - F_2)} \tag{2.2}$$

where ΔV denotes the regional air volume change induced by inhalation, and V represents the air volume at end-exhalation. $F1$ and $F2$ correspond to the air fractions in the exhale and inhale phases, respectively [20]. Using DVF derived from the DIR algorithm between the end-inhalation and end-exhalation 4DCT phases, a voxel-wise ventilation estimate based on HU variation is given by

$$Vent_{HU}(\boldsymbol{x}) = \frac{-1000 \times (\overline{HU}_{in}(\boldsymbol{x}') - HU_{ex}(\boldsymbol{x}))}{HU_{ex}(\boldsymbol{x}) \times (1000 + \overline{HU}_{in}(\boldsymbol{x}'))} \tag{2.3}$$

where $HU_{ex}(\boldsymbol{x})$ is the HU value at voxel $\boldsymbol{x}$ in the end-exhale CT image, and $\overline{HU}_{in}(\boldsymbol{x}')$ denotes the average HU of end-inhale phase voxels that are mapped to $\boldsymbol{x}$ under the DVF.

The accuracy of the resulting ventilation map calculated using this method depends on the accuracy of the HU value in the CT images. The uncertainty caused by image noise and artifacts will significantly affect the performance of this intensity-based metric.

In contrast, Jacobian-based methods directly recover volume changes from the Jacobian factor of the DIR solution [23, 24]. These methods are grounded in the assumption that local ventilation is proportional to volume changes induced by air inflow. A classical CTVI technique was proposed by Reinhardt et al. in 2008 [23], where regional lung tissue expansion and contraction are inferred from DVF between two 4DCT phase images. Mathematically, the local volume change is computed from the determinant of the Jacobian matrix of the transformation T. Assuming a unit voxel volume at baseline, a Jacobian determinant equal to 1 implies no regional deformation. Values greater than 1 indicate regional expansion, while values less than 1 denote regional contraction. Moreover, the local volume change at each inhale-phase voxel serves as the basis for computing ventilation, defined by the deviation of the Jacobian determinant from unity

$$Vent_{Jac}(\boldsymbol{x}) = \begin{vmatrix} 1+\frac{\partial T_1(\vec{\boldsymbol{x}})}{\partial x_1} & \frac{\partial T_1(\vec{\boldsymbol{x}})}{\partial x_2} & \frac{\partial T_1(\vec{\boldsymbol{x}})}{\partial x_3} \\ \frac{\partial T_2(\vec{\boldsymbol{x}})}{\partial x_1} & 1+\frac{\partial T_2(\overrightarrow{\boldsymbol{x_{in}}})}{\partial x_2} & \frac{\partial T_2(\vec{\boldsymbol{x}})}{\partial x_3} \\ \frac{\partial T_3(\vec{\boldsymbol{x}})}{\partial x_1} & \frac{\partial T_3(\vec{\boldsymbol{x}})}{\partial x_2} & 1+\frac{\partial T_3(\vec{\boldsymbol{x}})}{\partial x_3} \end{vmatrix} - 1 \qquad (2.4)$$

An alternative geometric strategy was proposed by Zhang et al. [25], in which regional lung volume change is determined by calculating the volume of the tetrahedra that results from displacing the eight voxel vertices from their initial cubic configuration according to

the calculated DVF. This method has been shown to be mathematically equivalent to a finite difference approximation of the analytical Jacobian determinant [26].

However, these methods are calculated at the voxel level, and the image artifacts and the DIR accuracy will seriously affect their results. To improve the robustness of the CTVI/PI algorithms, alternative methods focusing on subregional analysis have been developed [21, 24, 27-29]. To reduce the uncertainty caused by the DIR, subregional based methods were implemented to calculate the volume or density changes. For instance, Castillo et al. introduced the Integrated Jacobian Formulation (IJF) and Mass Conserving Volume Change (MCVC) numerical methods, utilizing sub-regional volume change measurements that satisfy a certain uncertainty tolerance [21, 24].

2.1.2 Improved Algorithms

Over nearly two decades of development, numerous improved CTVI/PI algorithms have been proposed, and each has its advantages and disadvantages.

A. Hybrid model using joint density and volume information

A hybrid method has been proposed as an extension of the original HU-based model [30], incorporating both variations in HU values and regional volume variations to correct for voxel-level density alterations due to tissue compression. The resulting hybrid CTVI is computed as

$$CTVI_{Hy} = \frac{HU^*_{ex}(x) - HU^*_{in}(x+v)}{HU^*_{in}(x+v) + 1000} \times \frac{HU_{ex}(x) + 1000}{1000} \qquad (2.5)$$

where

$$HU_{ex}^{*}(x) = \frac{HU_{ex}(x)}{Jac(x, v)} \tag{2.6}$$

B. Biomechanical lung modeling

Morfeus is a finite-element-based DIR algorithm designed for lung CT registration [31]. It enables the specification of spatially varying elastic properties within the lung and constrains local deformation based on biomechanical characteristics. This approach has been shown to yield more accurate DVF and provides access to strain and stress distributions within lung tissue. Building on Morfeus, Cazoulat et al. [32] introduced a method that employs lung mechanical stress as a surrogate for ventilation estimation. This technique uses strain maps derived from DIR between the end-inhalation and end-exhalation 4DCT phases, in conjunction with tissue elasticity parameters, to compute the spatial distribution of mechanical stress, which is then used to infer ventilation. The resulting ventilation maps demonstrated significantly stronger and more consistent correlations with NM-based ground truth compared to conventional DIR-based CTVI metrics.

C. Robust CTVI mapping concepts

The classical Jacobian metric, as a mathematical formulation of the volume-change model, estimates voxel-level volume variation by computing the first derivative of the DVF. However, CT images are inherently discretized into a voxel matrix, making direct computation of continuous derivatives infeasible. Since the Jacobian is defined for smooth

and continuous functions, numerical differentiation must instead be approximated using a finite difference scheme, typically expressed as

$$\frac{\partial \emptyset_i(x_k)}{\partial x_j} \approx \emptyset_i(x_k + e_j) - \emptyset_i(x_k) \tag{2.7}$$

This approximation introduces an uncertainty on the order of O (1), which can be as large as the voxel grid spacing. Such discretization-related error renders the original Jacobian-based method susceptible to both DIR accuracy and CT image quality.

As an enhanced formulation of volume-based ventilation estimation, the IJF proposed by Castillo et al. [24] aims to provide a more robust numerical approach for computing voxel-level volume changes from spatial transformations obtained via DIR. In this method, the deformation field between the end-inhalation and end-exhalation 4DCT phases is first estimated using DIR. The lung region on the end-exhale image is then partitioned into multiple subregions, defined by randomly distributed, equally spaced point clouds. Using a hit-or-miss algorithm, the corresponding subregions in the end-inhale phase are identified and iteratively expanded to satisfy Gaussian distribution assumptions and enable standard error analysis. Volume change ratios for each subregion are subsequently calculated. Finally, voxel-wise ventilation is reconstructed by solving a constrained linear least-squares problem. Compared to the conventional Jacobian-based CTVI, the IJF method demonstrates improved repeatability and stability, with significantly enhanced robustness to variations in DIR accuracy and image quality.

Similarly, the "coarse-to-fine" strategy employed in the IJF framework has also been adapted to enhance the robustness of HU-based CTVI methods. Building on this idea,

Castillo et al. [21] introduced the MCVC approach, a robust variant of the HU-based model. This method estimates voxel-level ventilation by leveraging subregional volume change measurements, defined through mean density ratios within localized regions. The CTVIs produced using this approach have shown strong correlation with those derived from the IJF method, indicating that incorporating robustness into ventilation estimation can yield more consistent results across different algorithmic paradigms (i.e., HU-based vs. volume-change-based).

D. Multi-layer supervoxels estimation

As an implementation of patch-based strategies, Szmul et al. [27] proposed a robust CTVIapproach that estimates ventilation by analyzing intensity changes within supervoxels. Using DVF derived from DIR, both end-inhalation and end-exhalation 4DCT phases are converted into multi-layer supervoxel representations. For each supervoxel, the mean intensity difference is computed to generate supervoxel-level ventilation estimates. These estimates are then averaged across all layers to produce a voxel-wise ventilation distribution. Compared to conventional HU-based models, this method may offer greater physiological alignment with lung anatomical structures.

E. Mass-conserving perfusion modeling

Castillo et al. developed CTPI based on a mass conservation model which describes the unknown mass change as a linear combination of spatially corresponding inhale and exhale HU estimated voxel densities [14]. CTPI requires a DIR between the inhale/exhale lung CT pair, a preprocessing lung volume segmentation, and an estimate for the Jacobian

of the DIR transformation. Given this information, the CTPI image, which provides the magnitude mass change for each voxel within the lung volume, is formulated as the solution to a constrained linear least squares problem defined by a series of subregional mean magnitude mass change measurements. The amount of uncertainty in a subregional sample mean measurement is related to measurement resolution and can be characterized with respect to a tolerance parameter τ.

2.2 Deep Learning-based Method

In recent years, the widespread adoption of DL in computer vision has led to remarkable advancements, and related techniques have increasingly been applied to medical imaging tasks such as diagnosis, segmentation, and synthesis. Based on the characteristics that CT images commonly contain high-resolution information, various lung diseases are usually manifested as intensity alterations. And multilayer convolutional networks show good performance in various medical image synthesis tasks. In the lung function imaging field, some experimental DL based CT-to-ventilation mapping methods were proposed. These methods synthesize lung functional images by extracting physiologic and biologic information from CT images directly without performing DIR.

The first application of DL to CTVI was reported by Zhong et al. in 2019 [33]. In this preliminary study, two respiratory phases from the 4DCT series, specifically the end-inhalation and end-exhalation phases, were used as input to a nine-layer fully convolutional network, with ventilation images generated by the HU-based method serving as predicting label. Although no “ground-truth” functional images from NM imaging were available, the

study demonstrated the feasibility of generating ventilation maps from 4DCT using deep CNNs. Liu et al. [34] proposed an improved approach based on the widely used 2D U-Net architecture. Their method was trained and validated on a cohort of 50 patients with paired 4DCT and SPECT-V scans, and was shown to outperform conventional DIR-based algorithms with statistically significant improvements. Ren's studies employed input data containing a single conventional CT image acquired at the same posture as the CTPI image [16, 17]. These methods are based on the fact that the HU value is influenced by the air/tissue ratio, which in turn determines the attenuation coefficient in a specific location. In the lung parenchyma, the fractional air/tissue can act as a surrogate for the blood-gas exchange rate. Consequently, pulmonary diseases involving abnormal blood-gas exchange often exhibit textural alterations in the lung parenchyma.

While deep learning methods have shown promising results, they do have certain limitations. One such limitation is their requirement for a large amount of training data. Data from other sources may not generalize well and could lack interpretability. Additionally, the complexity of deep learning models can make it challenging to understand the reasoning behind their predictions.

2.3 Other Emerging Method

2.3.1 Air-tissue Product Attenuation Method

The air-tissue product attenuation model [35] represents the first non-DIR CTVI approach for computing CTVI, enabling ventilation estimation directly from 4DCT without

relying on DIR algorithms. This method is built upon the physiological assumption that regional ventilation correlates with the product of local air and tissue densities, known as air-tissue products. The corresponding ventilation distribution is computed as

$$\mathrm{CTVI}_{\overline{\mathrm{HU}}}(\boldsymbol{x}) = \sum_{\phi=1}^{N} V_{\phi}(\boldsymbol{x})/N \tag{2.8}$$

where

$$V_{\phi}(\boldsymbol{x}) = \begin{cases} f_{\phi}^{\mathrm{Air}}(\boldsymbol{x}) \times f_{\phi}^{\mathrm{Tissue}}(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in L(\phi) \\ 0, & \text{if } \boldsymbol{x} \notin L(\phi) \end{cases} \tag{2.9}$$

$$= \begin{cases} \dfrac{\mathrm{HU}_{\phi}(\boldsymbol{x})}{-1000} \times \dfrac{\mathrm{HU}_{\phi}(\boldsymbol{x}) + 1000}{1000}, & \text{if } \boldsymbol{x} \in L(\phi) \\ 0, & \text{if } \boldsymbol{x} \notin L(\phi) \end{cases}$$

In these equations, $HU_{\phi}(\mathbf{x})$ denotes the HU value at voxel $\mathbf{x}$ in the ϕ-th phase of the 4DCT series. For HU values between −1000 and 0, the terms $f_{\phi}^{\mathrm{Air}}(\boldsymbol{x})$ and $f_{\phi}^{\mathrm{Tissue}}(\boldsymbol{x})$ correspond to the estimated air and tissue fractions at voxel $\mathbf{x}$, respectively. $L(\phi)$ indicates the binary mask for phase ϕ that defines the lung region in the corresponding 4DCT phase. The final output, $\mathrm{CTVI}_{\overline{\mathrm{HU}}}(\boldsymbol{x})$, represents a ventilation surrogate mapped onto the time-averaged 4DCT geometry. Although this attenuation-based method is computationally straightforward and reproducible, the absence of lung tissue motion information generally results in limited spatial accuracy [30].

Building on this work, Tian et al. [36] proposed a simplified version of the air-tissue attenuation model in 2019, which requires only a single phase-averaged 4DCT image as

input, thereby eliminating the need for the full series of 4DCT phase images. The corresponding ventilation surrogate is given by

$$\mathrm{CTVI_{AVG}}(\boldsymbol{x}) = \begin{cases} \dfrac{\mathrm{HU}_{AVG}(\boldsymbol{x})}{-1000} \times \dfrac{\mathrm{HU}_{AVG}(\boldsymbol{x}) + 1000}{1000} & , \text{if } \boldsymbol{x} \in L_{AVG} \\ 0 & , \text{if } \boldsymbol{x} \notin L_{AVG} \end{cases} \tag{2.10}$$

where $\mathrm{HU}_{AVG}(\boldsymbol{x})$ refers to the HU value at $\boldsymbol{x}$ position (voxel), and L_{AVG} is the lung mask defined on the phase-averaged 4DCT image. Compared to the original air-tissue product attenuation metric, this approach has less input requirement and improved computation efficiency.

2.3.2 Texture Feature-based Method

In Lafata's study [37], thirty-nine radiomic features were extracted from the lungs of 64 patients to explore their potential as imaging biomarkers for pulmonary function. These features collectively captured the lung's morphology, intensity variations, fine-texture, and coarse-texture of the pulmonary tissue. Comparing the extracted lung radiomics data to conventional pulmonary function tests, they found that patients with larger lungs of homogeneous, low attenuating pulmonary tissue (as measured via radiomics) were associated with poor spirometry performance and a lower diffusing capacity for carbon monoxide.

Chapter 3 Materials and Methods

This chapter aims to present an interpretable framework that identifies and evaluates voxel-wise pulmonary perfusion surrogates from thoracic CT through quantitative subregional texture analysis and data-driven ML. By integrating multi-stage feature selection, classification modeling, and spatial consistency validation against Q-SPECT, the proposed method demonstrates the potential to generate functional imaging from routinely acquired CT scans. Portions of this study have previously been published in Respiratory Research as a peer-reviewed journal article [38], and permission to reuse the manuscript in the thesis was granted by the publisher.

3.1 Study Design

Figure 3-1 depicts the overall workflow of this study. In the feature selection stage, lung regions in CT images from all patients were segmented into multiple subregions. Based on a subject-specific threshold, each CT lung subregion was binarily labeled as either high- or low-functioning according to its spatial-averaged Q-SPECT signal. Within each high or low functional subregion, features were extracted for intensity and texture information and underwent a two-stage feature selection pipeline. The prior-knowledge-driven preselection stage initially screened for robust and non-redundant perfusion-correlated features, while the data-driven selection stage further filtered features by fitting ML models for a classification task.

The final classification model was trained with the highest-performing PE-associated feature combination and evaluated for the subregion-level predictability. Feature maps (FMs) and model score maps (MSMs) of the final selected features were generated to study their spatial distributions and as potential surrogate perfusion maps. The voxel-wise Spearman correlation coefficient (SCC) and Dice similarity coefficient (DSC) for functional lung volumes were used to assess the spatial consistency between FMs or MSMs and Q-SPECT scans. The key stages in this workflow are further specified below.

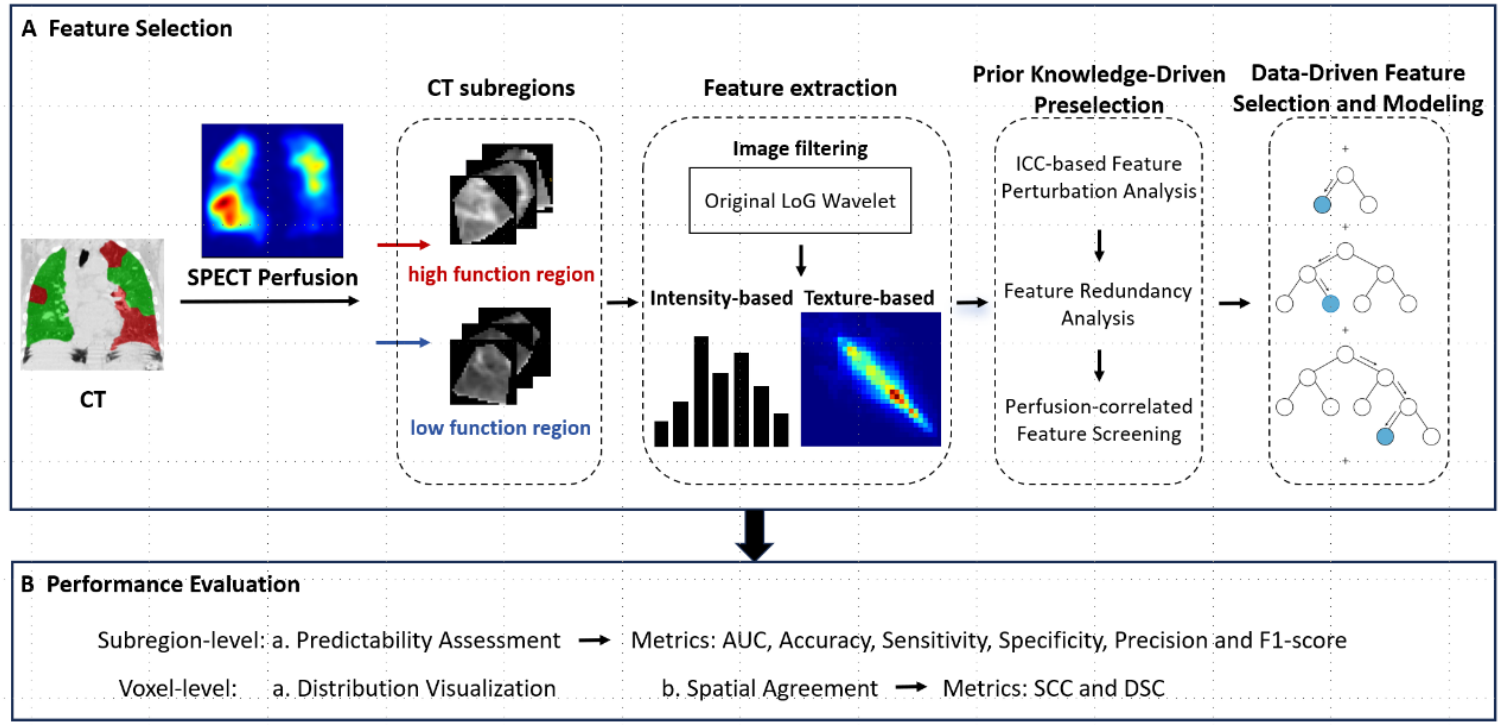

Figure 3-1 The overall framework of proposed study.

Function-correlated features were extracted and underwent a two-stage feature selection pipeline. The PE-associated selected features were subsequently mapped to perfusion surrogates (A). Quantitative subregion-wise predictability and voxel-wise spatial consistency assessment were applied to evaluate the performance (B).

Abbreviations: SCC, Spearman correlation coefficient; DSC, Dice similarity coefficient.

3.2 Patient Dataset

3.2.1 Internal QMH Dataset

The internal discovery cohort was retrospectively collected from a database previously utilized in our research [16], with approval from the Institutional Review Board (IRB) of the affiliated institution. The raw database includes 173 patients who underwent pulmonary ^{99m}Tc-labeled macroaggregated albumin (^{99m}Tc-MAA) Q-SPECT/CT scanning at Queen Mary Hospital (QMH), between 2019 and 2023, for suspected lung diseases. Subject exclusion criteria for this study were as follows: (a) presence of other lung conditions, such as pulmonary hypertension, lung cancer, or congestive heart failure, and (b) incomplete pulmonary imaging. Patient characteristics are summarized in **Table 3-1**.

Patients' Q-SPECT/CT images for the internal discovery cohort were acquired using a GE Discovery 670 SPECT/CT scanner at a frame rate of 30 s/frame, totaling 60 frames. Prior to imaging, each patient was injected with 3 mCi of ^{99m}Tc-MAA in a supine position to prevent sedimentation of macroaggregates into the lung bases. The CT scans were acquired at 120 kVp and 80 to 120 mAs, and reconstructed into a 512 × 512 matrix with a resolution of 0.97 × 0.97 × 1.25 mm^3. Q-SPECT scans were acquired in a standard 128 × 128 matrix with a voxel size of 4.42 × 4.42 × 4.42 mm^3.

Table 3-1 Patient characteristics (n=173).

Characteristics		Value	Percent
Sex	Male	68	39.3%
	Female	105	60.7%
Age (years: mean ± SD)		67±15	
Disease	Pulmonary embolism	71	41.0%
	Pulmonary hypertension	42	24.3%
	Lung cancer	16	9.2%
	congestive heart failure	3	1.7%
	Others *	34	19.7%
Nondisease	Healthy	7	4.0%

Abbreviation: SD, standard deviation.

*Others: Acute on chronic renal failure, aortic dissection, bronchiectasis, cancer carcinoma, chest pain, chronic obstructive airway disease, deep venous thrombosis, fever, hepatopulmonary syndrome, hypoplastic right lung, hypoxia, lung infection, lung metastases, lung transplant, lymphoma, pulmonary hypertension, pleural metastases, pneumonia, post-COVID-19, pulmonary angiosarcoma, pulmonary arteriovenous malformation, pulmonary fibrosis, pulmonary leiomyosarcoma, shortness of breath, systemic lupus erythematosus, tuberculosis.

3.2.2 External PUTH Dataset

An external testing dataset of Q-SPECT/CT scans collected from 2020 to 2023 at Peking University Third Hospital (PUTH) was included (**Figure 3-2**) and was approved by the Medical Science Research Ethics Committee of the affiliated institution. Q-SPECT/CT images for the external testing cohort were collected using the Symbia Intevo SPECT/CT scanner by Siemens Healthineers. The CT scans were acquired at 130 kVp and 70 to 200 mAs, and reconstructed into a 512 × 512 matrix with a resolution of 0.97 × 0.97 × 5.00 mm^3. Q-SPECT scans were acquired in a standard 128 × 128 matrix with a voxel size of 4.80 × 4.80 × 4.80 mm^3.

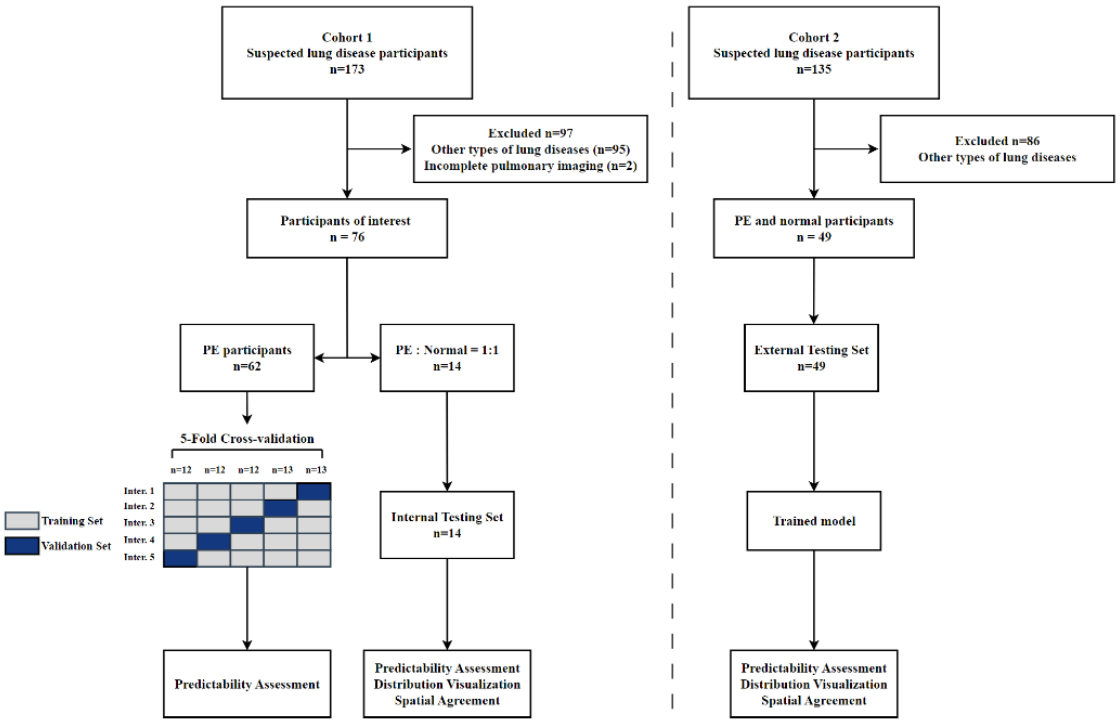

Figure 3-2 The flowchart of Q-SPECT/CT scans included in the study shows the number of participants of interest in the pipeline development, the number of participants excluded from the study due to other types of lung diseases or incomplete lung imaging, and the number of participants used for external testing.

Among them, the training cohort (cohort 1) was divided into five groups. In each fold, the classifier was trained in four groups and validated in the remaining group.

3.3 Image Preprocessing

Q-SPECT images were rigidly registered with CT images to align their spatial positions. All images, including CT scans and SPECT-based perfusion scans, were resampled to an isotropic resolution of 1.0 × 1.0 × 1.0 mm^3 using linear interpolation to ensure geometric consistency across modalities and subjects. To isolate the lung regions for focused analysis and reduce noise from surrounding tissues, lung masks were generated from the CT images for each patient using an open-source U-net-based model [39]. Each generated lung mask was manually checked, and the trachea regions were excluded from the mask. To reduce subsequent computational costs, all images were cropped to include only the lung regions. To suppress non-lung regions and anatomical structures irrelevant to PE while enhancing contrast, only lung voxels within the intensity range of -1000 Hounsfield unit (HU) to 200 HU were utilized from the CT images and normalized to the range of 0 to 1.

3.4 Subregion Partition and Functional Label Identification

Figure 3-3 shows the lung mask partitioning pipeline we developed for defect identification based on thoracic CT data. For each patient, the lung mask was segmented into roughly equal-sized subregions using the masked simple linear iterative clustering (maskSLIC) algorithm [40], with each subregion set to a volume of 21 × 21 × 21 mm^3. These subregions, characterized by compactness and a balance between intensity homogeneity and spatial proximity, were applied to the corresponding CT and SPECT-

based perfusion images. Subsequently, based on the subject-specific signal threshold calculated in corresponding SPECT images, these subregions were categorized into high-functional subregions and low-functional subregions. Specifically, the signal threshold, employed in previous NM-based and CT-based lung function studies, was defined as deviating by 15% from a normal lung function distribution, with careful mitigation of hotspots' impact [41, 42].

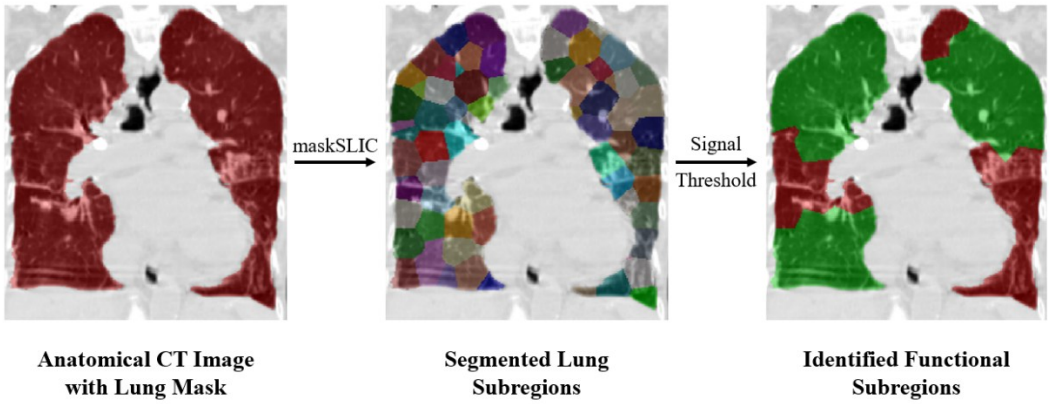

Figure 3-3 Schematic of the lung mask partitioning steps for identifying defects from thoracic CT in a 92-year-old woman with pulmonary embolism.

3.5 Thoracic CT Feature Extraction

The gray-level intensity of CT images was discretized into 64 bins ranging from -1000 to 200 HU for feature calculation. A total of 1116 features were extracted from each subregion using a radiomics software package (PyRadiomics; Harvard Medical School, Boston, Massachusetts, USA), which transforms images into quantitative data and is compliant with the Imaging Biomarker Standardization Initiative (IBSI) [43]. There were six types of features: first-order features, grey level co-occurrence matrix (GLCM) features, grey level size zone matrix (GLSZM) features, grey level run length matrix (GLRLM) features, grey level dependence matrix (GLDM) features, and neighboring gray-tone difference matrix (NGTDM) features. The latter five types are collectively referred to as textural features. In addition to the original images, features were also calculated from images processed through Laplacian of Gaussian (LoG) filtering and wavelet filtering. All features were extracted repeatedly from LoG-filtered images with sigma values of 0.5mm, 1mm, and 2mm, and from wavelet-filtered images at eight different frequencies. **Table 3-2** presents the radiomics features used in this study.

Table 3-2 List of radiomics features extracted from CT images.
Abbreviations: GLCM, Grey Level Co-occurrence Matrix; GLSZM, Grey Level Size Zone Matrix; GLRLM, Grey Level Run Length Matrix; GLDM, Grey Level Dependence Matrix; NGTDM, Neighbouring Grey Tone Difference Matrix;

Category	Feature count	Feature name
first-order (intensity-based)	18	10Percentile, 90Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean, Mean Absolute Deviation, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy, Uniformity, Variance
GLCM (texture-based)	24	Autocorrelation, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Inverse Difference(Id), Inverse Difference Moment (Idm), Inverse Difference Moment Normalized (Idmn), Inverse Difference Normalized (Idn), Informational Measure of Correlation 1 (Imc1), Informational Measure of Correlation 2 (Imc2), Inverse Variance, Joint Average, Joint Energy, Joint Entropy, Matthews Correlation Coefficient (MCC), Maximum Probability, Sum Average, Sum Entropy, Sum Squares
GLSZM	16	Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray

(texture-based)		Level Zone Emphasis, Large Area Emphasis, Large Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Low Gray Level Zone Emphasis, Size Zone Non-Uniformity, Size Zone Non-Uniformity Normalized, Small Area Emphasis, Small Area High Gray Level Emphasis, Small Area Low Gray Level Emphasis, Zone Entropy, Zone Percentage, Zone Variance
GLRLM (texture-based)	16	Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Run Emphasis, Long Run Emphasis, Long Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Entropy, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Run Variance, Short Run Emphasis, Short Run High Gray Level Emphasis, Short Run Low Gray Level Emphasis
GLDM (texture-based)	14	Dependence Entropy, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Dependence Variance, Gray Level Non-Uniformity, Gray Level Variance, High Gray Level Emphasis, Large Dependence Emphasis, Large Dependence High Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Low Gray Level Emphasis, Small Dependence Emphasis, Small Dependence High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis
NGTDM (texture-based)	5	Busyness, Coarseness, Complexity, Contrast, Strength

3.6 Prior Knowledge-driven Feature Screening

3.6.1 Feature Robustness Analysis

We conducted a perturbation analysis to evaluate the robustness of the extracted radiomics features. To simulate patient movement during imaging, variations in noise levels across different imaging devices, and uncertainties in region-of-interest (ROI) contouring, we randomly applied four fundamental image perturbation techniques: rotation (R), translation (T), noise addition (N), and contour randomization (C) [44]. The specific parameters used for each perturbation method are detailed in **Table 3-3**. For each subregion, 10 images were independently generated with random perturbations, and radiomics features were re-extracted from them. Intraclass correlation coefficient (ICC) [45] was calculated for each feature to quantify its robustness against perturbations. The ICC threshold of 0.75 for identifying features with relatively high robustness was commonly used in previous radiomics literature [44]. Features that met or exceeded the threshold were considered reliable and retained for further analysis.

Table 3-3 The perturbation parameters.

Abbreviations: AP, Anteroposterior; SI, Superoinferior; LM, Left-right lateral.

Perturbation methods	Perturbation range	Reference axis	Total number
Rotation Angles (degree)	-5, -3, 0, 3, 5	AP, SI, LM	300
Translation Distance (mm)	0, 0.1, 0.3, 0.5	SI	
Noise addition level	0, 1, 2	-	
Contour Randomization	3	-	

3.6.2 Feature Redundancy Analysis

To analyze feature redundancy within the dataset, we utilized the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [46]. The Pearson correlation coefficient (PCC) was employed as the metric to measure pairwise distances between features. A threshold of 0.95 for the PCC was set to identify and group highly correlated features into distinct redundant clusters. The threshold was chosen based on a published study [47]. In that paper, a PCC threshold of 0.9 was set to identify redundant feature clusters from 79 radiomic features. As for this study, we slightly increase the PCC threshold to 0.95, to retain more diverse features from different categories in a larger initial feature pool and to balance the number of features.

3.6.3 Perfusion-correlated Feature Preselection

In this step, the features that best represented the function differences were filtered out. For all feature clusters, a non-parametric two-sided Mann-Whitney U test was conducted on the feature values extracted from high-function subregions and low-function subregions to calculate the effect size (ES), which measures the degree of rank differences between the two groups [48]. For each redundant cluster, we retained the feature with the highest absolute ES value, provided that the absolute ES value was greater than 0.33, to ensure the capability to distinguish between high-functional and low-functional regions of the lung. The ES threshold of 0.33 was set based on the statistical interpretations of the nonparametric test-derived effect size (rank-biserial correlation). A rank-biserial

correlation ES > 0.33 indicates that the feature might have a medium-to-large ability to differentiate perfusion regions in lung CT. The filtered features served as the candidate feature set for subsequent data-driven feature selection.

3.7 Machine Learning-based Feature Selection

We randomly selected an equal number of participants from the eligible PE participants to match the normal participants, together forming an internal testing cohort, while the remaining patients constituted the training cohort. Z-score normalization was applied to ensure that feature values were within a comparable range, preventing any particular feature from dominating the modeling process. Specifically, the feature values in the candidate feature set were scaled to have a mean of 0 and a standard deviation of 1. As shown in **Figure 3-2**, the training cohort was divided into five groups, and a five-fold cross-validation with the CatBoost classifier was employed to avoid overfitting [49]. In each fold, one group was selected for validation, while the remaining four groups were used for classifier training. Feature selection in each fold was performed using the recursive feature elimination (RFE) algorithm based on importance, eliminating the features with the lowest importance ranking until the number of remaining features reached the predetermined count. As a result, five fold-specific feature combinations with perfusion-correlated predictive capabilities were obtained. These five feature combinations were combined to create an integrated feature set, where each feature was ranked based on its frequency of occurrence during cross-validation. We retained the features that appeared

more than three times as the final selected feature combination for training a final model, which was applied to the testing cohorts for evaluating subregion-level predictability. The predictabilities of the fold-wise models and the final model were evaluated by the receiver operating characteristic (ROC) curves, with the fold-averaged the Area Under the Curve (AUC) for the internal validation cohort. 1000 bootstrap resamples were performed on all cohorts to calculate the 95% confidence intervals (CI) to assess the stability of the model performance. Additionally, the final model with the final feature combination was then applied to the external testing cohort to assess the accuracy and generalizability of the model. As performance comparisons, once the final feature combination was selected, logistic regression (LR) and support vector machine (SVM) were also employed to fit the training cohort and validate AUCs with the internal testing cohort.

3.8 Voxel-level Spatial Distribution Study

For the features within the final combination, we calculated their FMs for visualizing the spatial distribution of features on the CT image. In order to enhance computational efficiency and obtain relatively smooth FMs that retain the trend of the spatial distribution of features and eliminate artifacts, we employed a coarse-to-fine approximation method that transitioned from subregion-wise feature measurement to voxel-wise dense FM, instead of using the sliding-window filtering method [47]. For each patient, we first constructed a weighted matrix for each voxel with respect to all subregions based on Shepard's class of moving least squares approximation [50, 51]. Using the constructed

weighted matrix, the feature values measured from each subregion can be transferred to each lung voxel that forms the FM.

3.9 Quantitative Analysis

We quantitatively evaluated the spatial similarities between FMs and reference Q-SPECT scan images using the SCC and DSC. The SCC, a non-parametric metric that captures the degree of the voxel-wise intensity correlation, is defined as:

$$\mathrm{SCC} = \frac{\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)\cdot\left(R(y_i) - \overline{R(y)}\right)}{\sqrt{\left(\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)^2 \cdot \sum_{i=1}^{n}\left(R(y_i) - \overline{R(y)}\right)^2\right)}} \tag{3.1}$$

where $R(x_i)$ and $R(y_i)$ denote the ranks of the FM and Q-SPECT signal intensities for the i-th voxel, and $\overline{R(x)}$ and $\overline{R(y)}$ represent the mean ranks over all n non-zero voxels. Additionally, we computed the DSC for low functional region (DSC_{LO}) and high functional region (DSC_{HI}) to quantify functional region overlap, as defined by:

$$\mathrm{DSC} = \frac{2 * |X \cap Y|}{|X| + |Y|} \tag{3.2}$$

where X and Y are binary masks representing the same functional regions (low-functional or high-functional regions) derived from the FM and the reference Q-SPECT images, respectively. The same approach as used for subregion functional label identification in Section 3.4 was applied to define subject-specific signal thresholds for dividing high- and low-functional regions from FMs and corresponding reference Q-SPECT images. Similarly,

we generated MSMs based on the model's predicted probabilities for subregions and assessed the spatial consistency with the reference Q-SPECT scans.

Chapter 4 Results

4.1 Study Population

Of the 173 participants from QMH enrolled in the study, 97 were excluded due to other types of lung diseases (n = 95) or incomplete pulmonary imaging (n = 2). To ensure the homogeneity of the disease discovery and to simulate a realistic population, 49 patients diagnosed with PE and normal conditions out of 135 participants from the PUTH cohort were involved. The selected participants of interest (n = 76) had a mean age of 67±17 years, with a female proportion of 61.4%. Patient demographic data are presented in **Table 4-1**, which indicates that there was no statistically significant difference between the patients from the two centers ($p > 0.05$).

Table 4-1 Study Population.

Characteristics	QMH								PUTH	
	Training Cohort (n=62)		Testing Cohort (n=14)		P Value	All Participants (n=76)			External Validation Cohort (n=49)	P Value
Age (y)	67±17		66±18		0.91	67±17			61±18	0.15
Gender					0.92					0.3
Male (Value, %)	23	37.1	6	42.9		46	38.7	11	22	
Female (Value, %)	39	62.9	8	57.1		29	61.4	39	78	

Note: Unless otherwise stated, the data are presented as mean ± standard deviation.

4.2 Feature Preselection

In the process of prior knowledge-driven preselection, **Figure 4-1** illustrates the results of the feature robustness analysis. By setting the threshold of ICC value to 0.75, a total of 587 robust features were filtered out from all the considered features. As shown in **Figure 4-2**, feature redundancy analysis on the robust features identified 329 independent non-redundant clusters. After removing features with ES values lower than 0.33, the candidate feature set included a total of 151 features, each selected as the highest ES value feature from each redundant cluster, for data-driven feature selection.

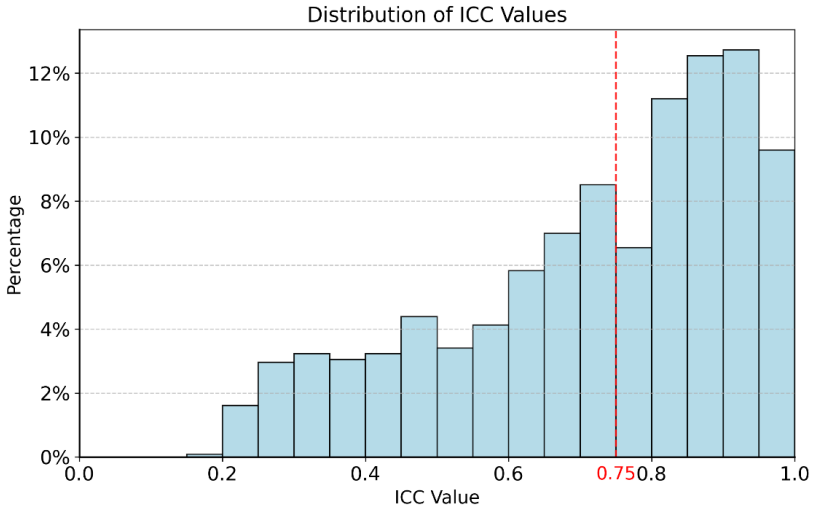

Figure 4-1 Histogram of the distribution of the ICC values for all considered features.

The vertical axis represents the percentage of features within the corresponding ICC value range out of the total number of features. The red line indicates the threshold of 0.75, above which robust features are retained.

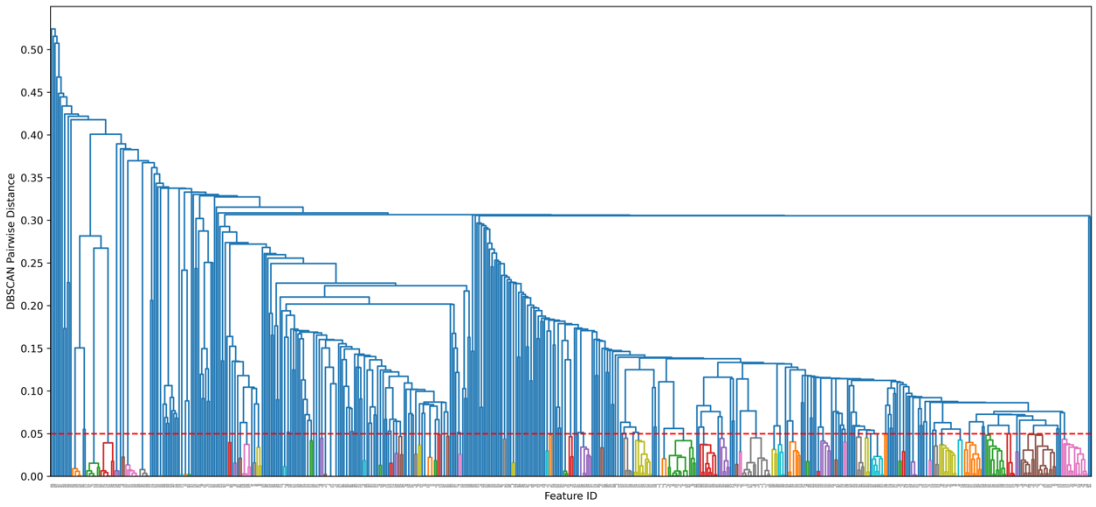

Figure 4-2 Hierarchical clustering of feature redundancy analysis.

Clusters formed below the red dashed line indicate highly correlated features, using a correlation threshold of 0.95.

4.3 Subregion-wise Classification Probabilities

Figure 4-3 depicts the change in the AUC of the CatBoost classifier with the number of selected features. The optimal model performance was determined when the number of selected features reached 16, achieving the highest predictability during the cross-validation. At this point, the AUC for the training cohort was 0.928 (95% CI: 0.922, 0.935), and the fold-averaged AUC for the internal validation cohort was 0.823 (95% CI: 0.814, 0.831). The evaluation of the classifier on predictability assessment, including AUC, accuracy, sensitivity, specificity, precision and F1-score, in the internal discovery and external testing dataset are summarized in **Table 4-2**. **Figure 4-4.A** presents the ROC curves for the evaluation of subregion-level predictability in the training, internal validation, and internal testing cohorts. Fourteen features, which were selected more than three times in the five-fold cross-validation, were retained as the final feature combination, including features (1) original First-order 10Percentile, (2) original First-order Skewness, (3) log-sigma-2-0-mm NGTDM Busyness, (4) log-sigma-1-0-mm First-order Skewness, (5) log-sigma-2-0-mm First-order 90Percentile, (6) wavelet-LLH First-order Median, (7) wavelet-LLH First-order Mean, (8) wavelet-HLH First-order 90Percentile, (9) log-sigma-2-0-mm First-order 10Percentile, (10) wavelet-LHL First-order Minimum, (11) wavelet-LHH GLCM Imc1, (12) wavelet-LLH First-order 10Percentile, (13) wavelet-LLH First-order Minimum, (14) wavelet-HLH First-order RootMeanSquared.

As shown in **Figure 4-4.B**, we compared the performance of different classifiers with the final feature combination. The CatBoost model achieved an AUC of 0.863 (95% CI:

0.849, 0.877) during internal testing. Using the identical set of features, the LR and SVM classifier yielded AUCs of 0.797 (95% CI: 0.779, 0.814) and 0.847 (95% CI: 0.832, 0.861), respectively, indicating that the CatBoost classifier model outperformed the other classifiers in predictability. Additionally, the final CatBoost model also had a generalizable subregion-level predictability on the external testing cohort with an AUC of 0.828 (95% CI: 0.820, 0.838). **Figure 4-5** displays the SHapley Additive exPlanations (SHAP) beeswarm summary plot with features and their contributions color-coded, where each dot corresponds to a subregion as a sample. Features in the final combination are ranked in descending order based on their contribution to the model's predictability, specifically by the mean absolute SHAP value. High feature values aligned with high contribution levels indicate a positive contribution to the model's predictability.

Table 4-2 Subregion-level agreement in predictability metrics across the training, internal validation, and internal testing cohorts.

	Training	Internal Testing			External Testing
		CatBoost	LR	SVM	
AUC	0.93 (0.92, 0.93)	0.86 (0.86, 0.87)	0.80 (0.80, 0.81)	0.85 (0.85, 0.86)	0.82 (0.81, 0.83)
Accuracy (%)	85 (84, 86)	78 (78, 79)	71 (71, 72)	75 (75, 76)	75 (74, 76)
Sensitivity (%)	92 (92, 93)	67 (65, 69)	70 (68, 72)	77 (76, 78)	71 (70, 72)
Specificity (%)	85 (84, 86)	84 (84, 85)	77 (76, 77)	82 (82, 83)	82 (81, 83)
Precision	0.84 (0.83, 0.85)	0.86 (0.85, 0.87)	0.84 (0.84, 0.85)	0.87 (0.85, 0.89)	0.94 (0.92, 0.96)
F1-score	0.88 (0.87, 0.89)	0.82 (0.82, 0.83)	0.74 (0.73, 0.75)	0.78 (0.76, 0.80)	0.83 (0.83, 0.84)

Note: All Data in parentheses are 95% confidence intervals.

Abbreviations: LR, Logistic Regression; SVM, Support Vector Machine; AUC, Area under the curve.

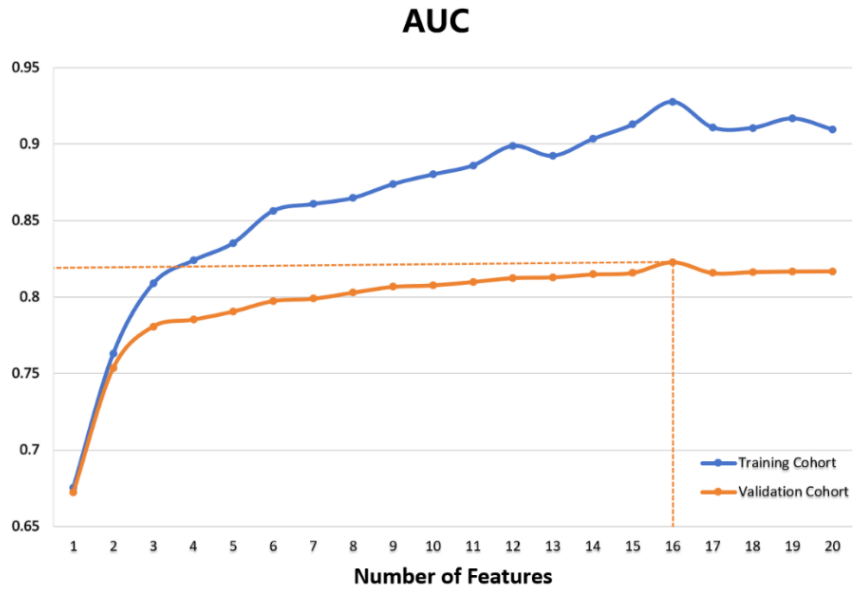

Figure 4-3 The change in ROC AUC of the CatBoost model in training and internal validation cohorts from the internal discovery dataset with the number of selected features.

The optimal model performance was achieved when 16 features were selected, resulting in the highest predictability (the fold-averaged AUC) in the internal validation cohort.

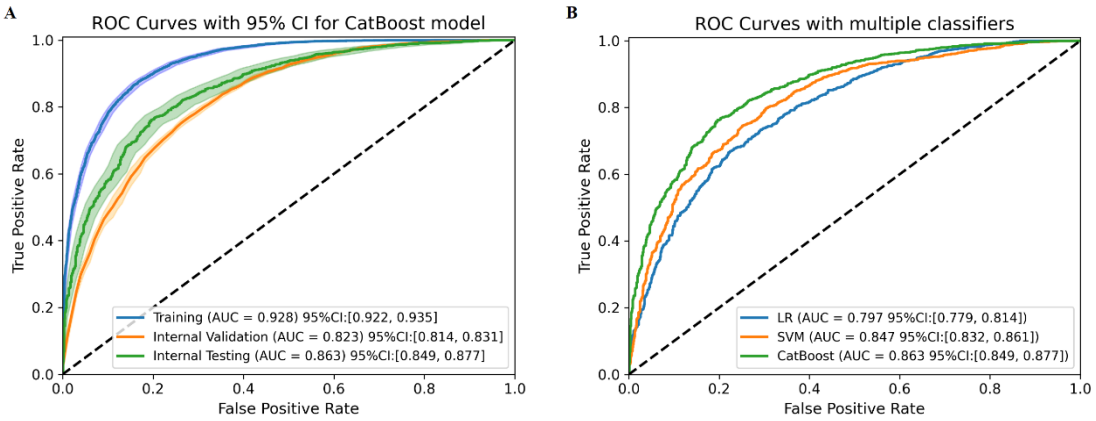

Figure 4-4 ROC curves show the result of subregion-level evaluation on agreement in the training, internal cross-validation and internal testing cohorts (A) and for multiple classifiers in the internal testing cohort (B).

The shaded areas represent the 95% CI for the ROC curves, indicating the range within which the model's true performance is expected to fall with 95% confidence.

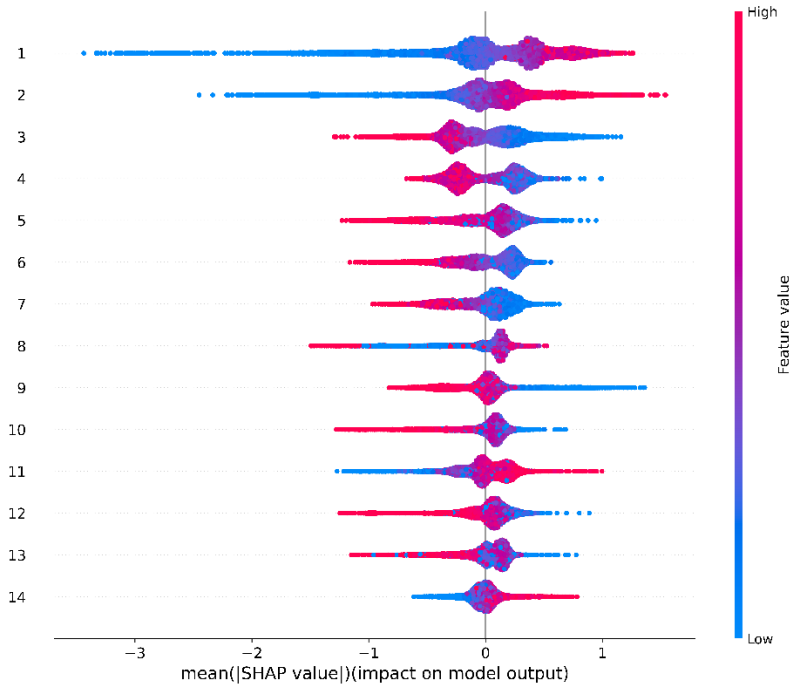

Figure 4-5 SHAP summary plot for the final CatBoost classifier.

4.4 FMs and MSMs Visualization

Figure 4-6 visualizes the FMs and MSMs of a representative patient with PE, using the Q-SPECT scan as a reference. Major defects of perfusion can be identified with the Q-SPECT scan in the left lower and right middle lobes of the lungs. Notably, the relationship between these FMs and the perfusion defects aligns with their previously calculated ES. Several positively correlated FMs (with positive ES) show low signal regions that broadly correspond to the perfusion defects. Conversely, negatively correlated FMs (with negative ES) display high signal regions in areas of reduced perfusion. The MSM demonstrates a positive correlation with perfusion distribution and exhibits greater visual similarity to the Q-SPECT scan than any single FM.

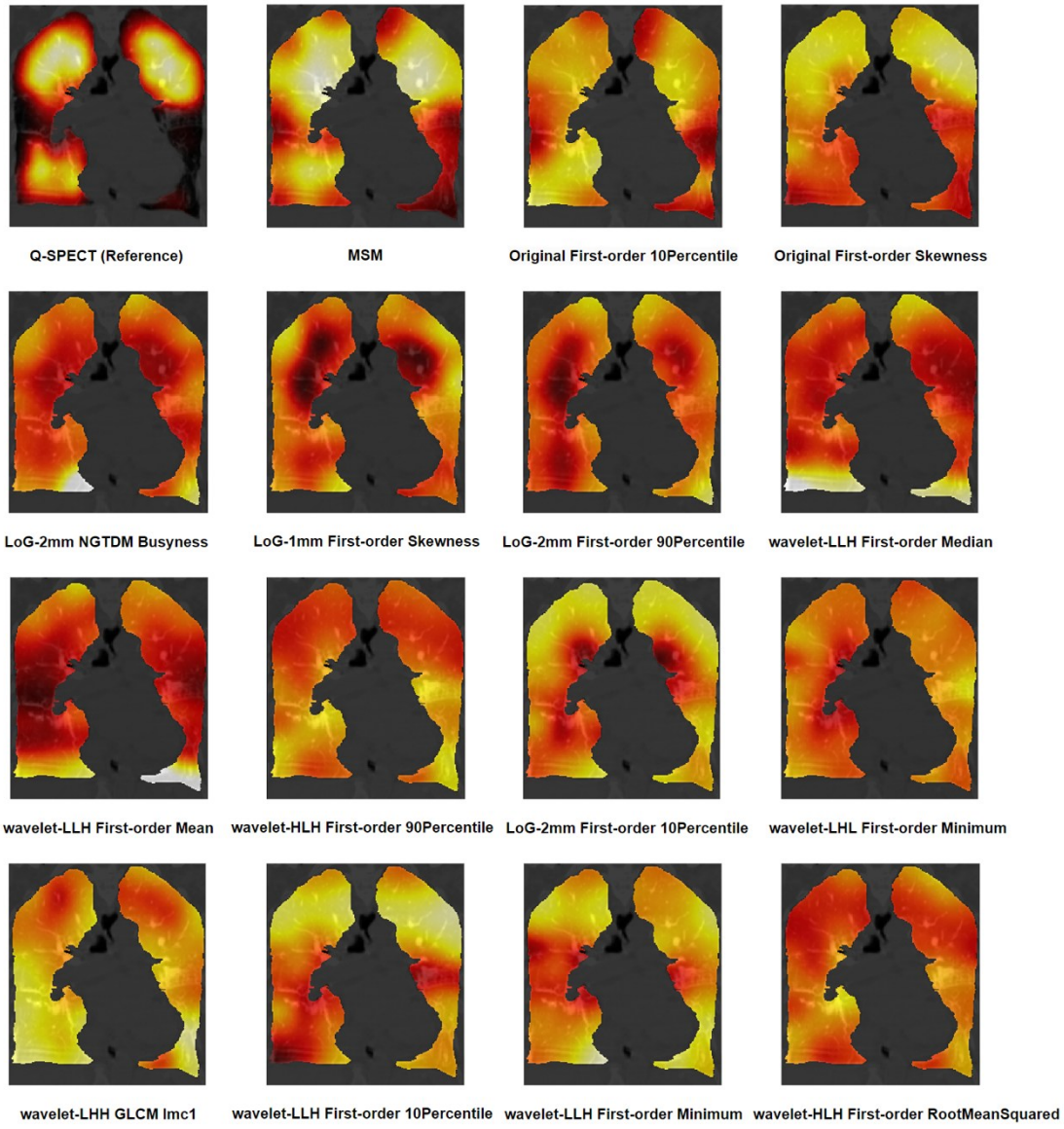

Figure 4-6 Reference Q-SPECT scan, feature maps of the 14 selected features from the final feature combination and MSM of a representative subject with pulmonary embolism disease.

4.5 Voxel-wise Spatial Agreement

Figure 4-7 and **Table 4-3** summarize the quantitative evaluation results on spatial consistency among internal and external testing cohorts. In box plots, the central points represent the mean, reflecting the central tendency of the dataset. The central horizontal line denotes the median. The top and bottom edges of the box indicate the first quartile (Q1) and the third quartile (Q3), respectively, illustrating the interquartile range (IQR). The whiskers extend to the smallest and largest values within 1.5 times the IQR from Q1 and Q3, respectively, with values beyond this range, marked by blue points, are considered outliers. Evaluations in the internal testing cohort are shown MSMs reached the median SCC of 0.66 and the median DSC of 0.64 and 0.86 for high and low functional regions, respectively. For the external testing cohort, MSMs achieved the median SCC of 0.60 as well as the median DSC of 0.49 and 0.85 for high and low functional regions, respectively.

At the same time, we conducted subgroup analyses to evaluate the model performance across different clinical risk profiles. We divided the testing datasets into central and peripheral embolism groups at the patient level to analyze the spatial agreement of the surrogate perfusion maps against the reference perfusion scans. The results, as illustrated in **Figure 4-8**, showed no significant differences in SCC and DSC metrics across these subgroups, suggesting that our method performs consistently regardless of embolism location. Furthermore, we performed similar subgroup evaluations on the external testing cohort based on sPESI score and ESC risk groups, as shown in **Figures 4-9** and **4-10**. While most comparisons showed no significant differences, we observed a statistically significant

difference in DSC High for patients with different sPESI scores in **Figure 4-9**. This finding warrants further investigation and may provide insights into the method's sensitivity to disease severity.

.

Table 4-3 Evaluation of feature and model-based spatial consistency across internal and external testing cohorts.

Feature Number/ Model	Internal Testing Cohort			External Testing Cohort		
	Median SCC	Median DSC_{LO}	Median DSC_{HI}	Median SCC	Median DSC_{LO}	Median DSC_{HI}
Model	0.66	0.64	0.86	0.60	0.49	0.85
1	0.40	0.06	0.78	0.48	0.03	0.95
2	0.04	0.38	0.70	-0.04	0.23	0.79
3	-0.50	0.15	0.62	-0.44	0.11	0.79
4	-0.51	0.61	0.74	-0.41	0.25	0.71
5	-0.59	0.09	0.62	-0.52	0.09	0.78
6	-0.38	0.28	0.47	-0.42	0.21	0.58
7	-0.46	0.28	0.37	-0.54	0.07	0.57
8	0.15	0.31	0.77	0.13	0.22	0.87
9	-0.51	0.55	0.71	-0.55	0.31	0.70
10	-0.23	0.43	0.77	-0.26	0.33	0.82
11	0.19	0.18	0.70	-0.09	0.12	0.95
12	-0.30	0.44	0.72	-0.47	0.37	0.79
13	-0.13	0.36	0.75	-0.35	0.36	0.86
14	0.08	0.29	0.78	0.03	0.23	0.85

Abbreviations: SCC, Spearman correlation coefficient; DSC, Dice similarity coefficient.

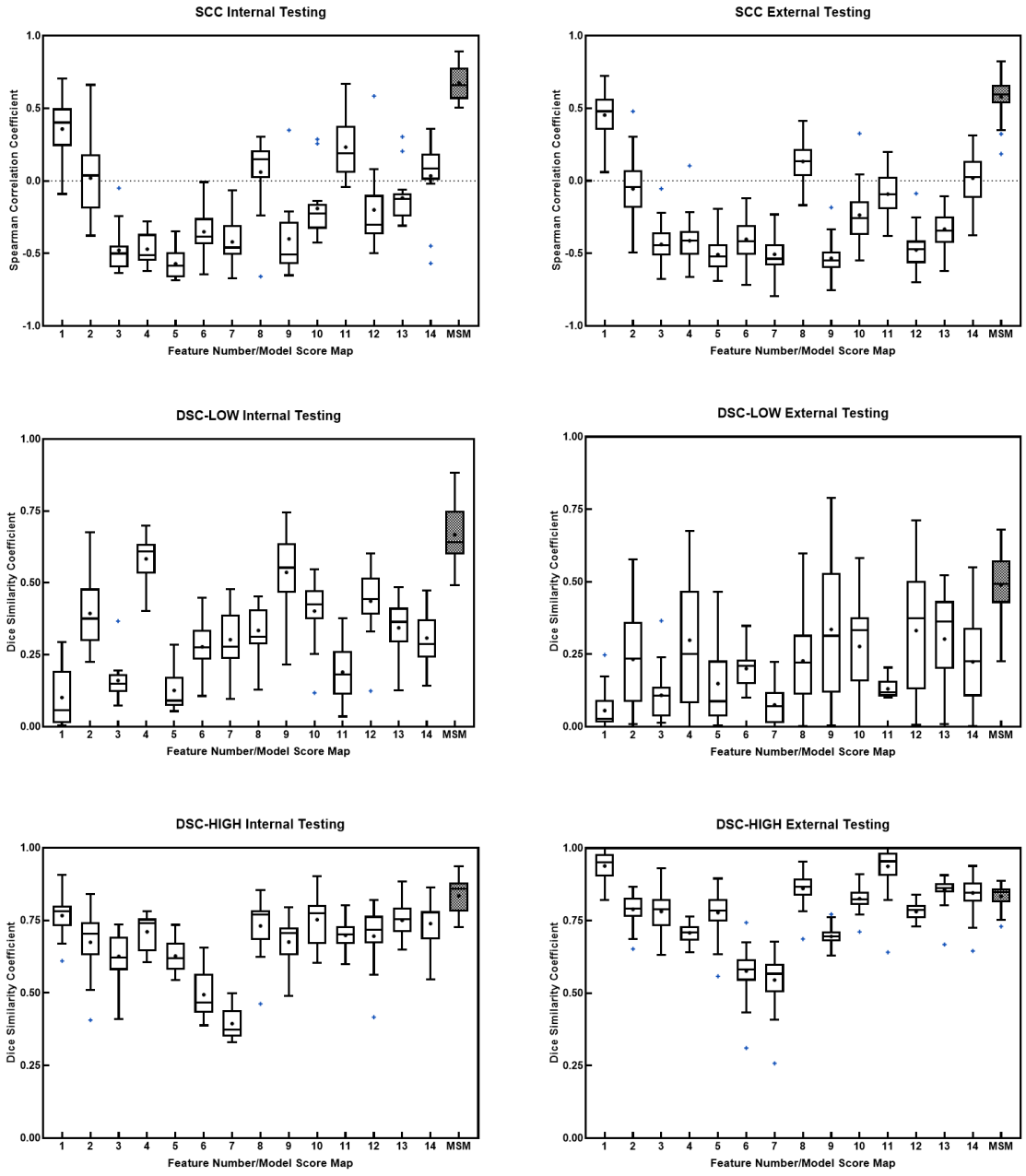

Figure 4-7 Box plots of SCC and DSC for internal and external testing cohorts.

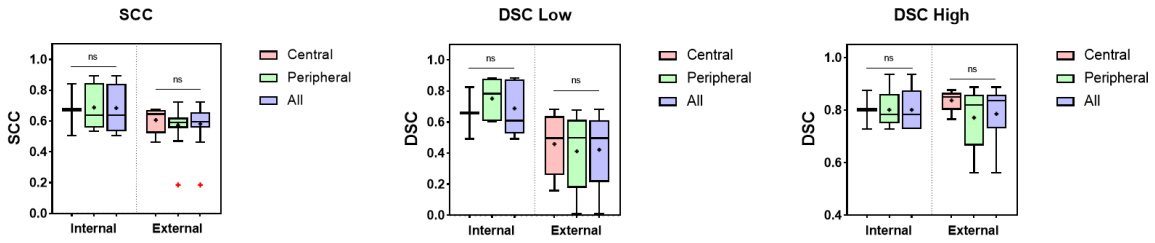

Figure 4-8 Subgroup analysis of the MSM results based on SCC and DSC for central and peripheral embolism in the internal and external testing cohorts.

Abbreviation: ns, not significant.

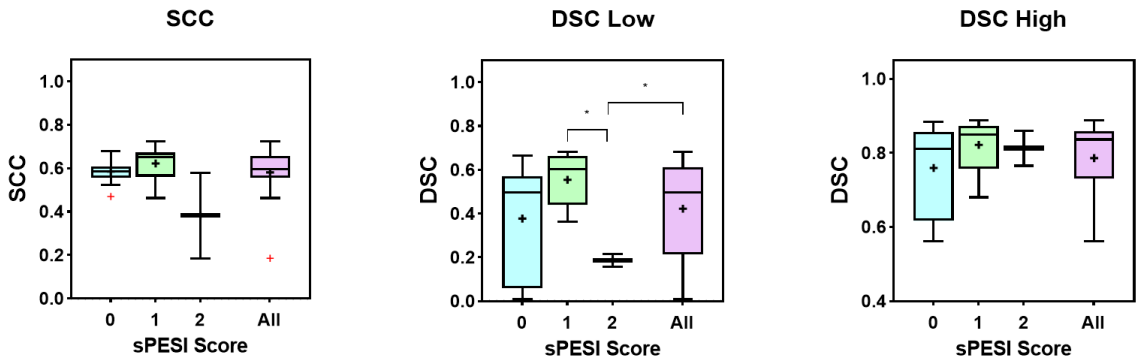

Figure 4-9 Subgroup analysis of the MSM results based on SCC and DSC for sPESI Score in the external testing cohorts.

Asterisk (*) indicates statistically significant differences (p-val < 0.05).

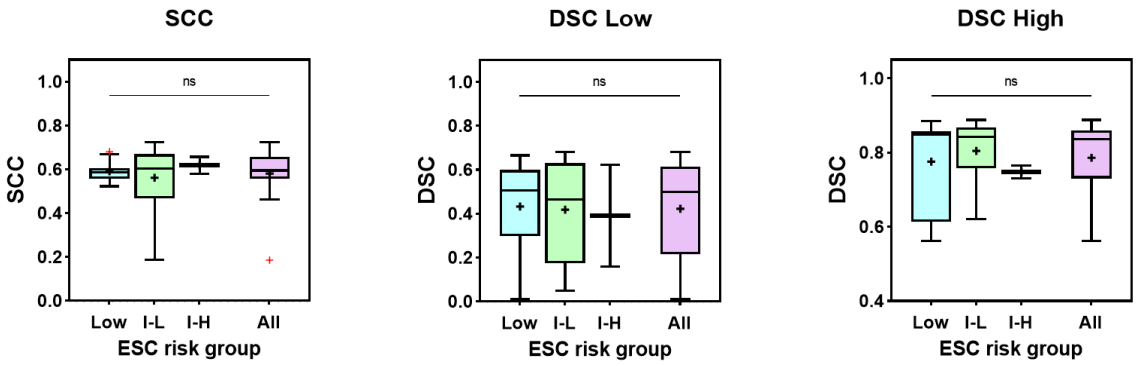

Figure 4-10 Subgroup analysis of the MSM results based on SCC and DSC for ESC risk groups in the external testing cohorts.

Abbreviations: I-L, Intermediate–Low; I-H, Intermediate–High; ns, not significant.

Chapter 5 Discussion and Conclusion

5.1 Key Findings

In this study, a cohort of 125 research participants underwent robustness analysis, redundancy analysis, and subregion-level perfusion-correlated feature screening on a set containing 1116 features. Subsequently, a data-driven feature selection and modeling pipeline was established. The model we developed identified 14 density-based and texture-based features, all of which were non-redundant, robust and capable of distinguishing between high and low functional regions of the lung (ES>0.33). The optimal model, which performed best during the cross-validation, was applied to the internal and external testing cohorts, yielding an AUC of 0.86 and 0.82, respectively, initially indicating effective predictability and cross-institutional generalizability.

The optimal model developed using the CatBoost algorithm retained 14 key features reflecting spatial heterogeneity of lung perfusion in PE cases. Most features were first-order features, including 10Percentile, 90Percentile, Minimum, Median, and Mean, which quantitatively assess image intensity values. Skewness and Root mean squared measure the asymmetry and variability of intensity distribution, respectively, indicating the density distribution and internal heterogeneity of defect subregions. The remaining features included two high-dimensional textural features: NGTDM Busyness, indicating the degree of gray-level variation between image voxels and their neighborhoods, suggesting higher irregularity in local textures, potentially containing edges; and GLCM-Imc1, indicating

gray-level correlation within the image. The latter was extracted using a wavelet-LHH filter, applying high-pass filters in the y and z dimensions to enhance spatial heterogeneity and detect solid components of defect subregions.

Within the selected feature combination, the median SCC of the FMs for 7 features indicated moderate-to-high spatial similarity with reference Q-SPECT images and the MSMs showed significant spatial consistency with reference Q-SPECT images. Both visual comparisons and quantitative evaluations in the spatial distribution study of the selected features' FMs and MSMs suggest that while individual FMs capture various aspects of perfusion distribution, the fitted MSM has the potential to provide more comprehensive and accurate representations of lung perfusion. The improved spatial correlation between the MSM and Q-SPECT scans underscores the value of the proposed ML approach in integrating multiple intensity and texture features to predict pulmonary function.

In terms of methodology, this study proposes a new perspective on CTPI for assisting PE diagnosis, distinguishing itself from DIR-based and purely deep learning-based methods, which are susceptible to image quality issues or lack interpretability [52, 53].In our framework, alongside intensity and texture features from raw images, we utilized high-dimensional features extracted from LoG- and wavelet-filtered images. Although high-dimensional features are more susceptible to motion artifacts [54, 55], our study conducted perturbation analysis to screen for robust features. Furthermore, acknowledges the nonlinear relationship between texture features and local lung function, while LR has a relatively simple parametric structure and provides a linear model to predict outcomes [56],

CatBoost and SVM are more efficient and robust in addressing nonlinear problems. Prior studies have shown CatBoost's superior effectiveness in handling nonlinear data relationships compared to other ML techniques [57]. From a clinical standpoint, providing an alternative method of assessing lung perfusion through routine CT scanning may reduce the need for more invasive or radiation-intensive procedures. Our method generates voxel-level perfusion maps that help localize perfusion defects precisely, thereby enhancing the diagnosis and management of PE and other lung diseases associated with pulmonary perfusion. To minimize any bias that the algorithm might introduce, we applied a five-fold cross-validation for the training dataset from the prior experience [58]. In addition to this, feature robustness analysis selected features that appeared more than three times during cross-validation, avoiding reliance on a single feature selection method. When dealing with a relatively small sample size, we adjusted the model parameters based on prior experience to ensure that the AUC difference between the training and testing datasets remained below 0.1, which helps to prevent overfitting.

5.2 Limitations and Future Research Directions

Despite the promising results, this study presents several limitations and points for improvement that are worth considering. The dataset’s relatively homogeneous characteristics may limit the model's generalizability to more varied populations. The age distribution of the participants in this study is limited, with an average of 65±18 years, as age-related physiological changes and the higher prevalence of lung disease in the elderly

may affect the model's performance. Different racial and ethnic groups may vary in lung structure, function, and disease susceptibility, which could influence the generalizability of the model. The participants in this study lacked sufficient racial and ethnic diversity, and future studies should consider a wider range of groups to ensure the applicability of the model. In addition to the demographic limitations, geographic diversity was not adequately addressed in this study. Environmental factors such as air pollution, altitude, and other location-specific elements can affect lung health to varying degrees. Conducting evaluations across a wider range of geographic locations would enhance the model's robustness. Overall, future research should focus on these aspects of diversity to ensure the model's effectiveness and generalizability across diverse populations.

Variations in CT scanning equipment, scanning protocols, and operator skill during image acquisition can lead to differences in image quality, resolution, and segmentation accuracy, which in turn may impact the study results. The relatively small sample size, while mitigated by cross-validation and external testing, suggests that evaluation on larger, diverse (i.e., across different scanning devices and protocols) cohorts would be beneficial. Additionally, when we attempted to replace the volume parameter settings for subregions generation with $26 \times 26 \times 26$ mm^3 (originally $21 \times 21 \times 21$ mm^3), the predictability on the internal testing cohort decreased to 0.79. The impact of the parameters on model performance indicates a need for further optimization of these settings.

Future research directions also explore the applicability of the proposed method to pulmonary vascular diseases that affect perfusion other than PE discussed in this study. To further enhance clinical relevance, an extended DL-based framework will be developed

that integrates virtual CT-derived functional imaging with radiomics and clinical information. This integrated approach is expected to enhance pre-screening efficiency in suspected PE cases while minimizing unnecessary imaging, especially in settings where V/Q-SPECT availability is limited. Its performance will be evaluated through prospective multi-center studies using specificity and AUC as key metrics.

Regarding the practical challenges of applying our classifier tool in clinical practice, we recognize that for any machine learning tool to be effectively implemented, it must be compatible with existing hospital information systems. This often requires relying on robust and secure APIs to interact with the clinical environment, ensuring seamless data transmission and system integration [59]. It also emphasizes the importance of designing user-friendly interfaces and providing comprehensive training to clinical staff, including mastering the technical operation of the tool and interpreting the model's outputs in a clinical setting to ensure the tool's effective use.

Finally, the model needs to be validated on new datasets and continuously updated. External validation using data from diverse patient populations or institutions can further enhance the model's generalizability. Furthermore, investigating the potential of the proposed method for longitudinal monitoring of perfusion changes across serial CT scans could provide valuable information for characterizing disease progression and response to therapy.

5.3 Conclusion

This thesis validates the feasibility of using quantitative texture analysis and a data-driven ML pipeline to generate voxel-wise lung perfusion surrogates across different institutions. We selected potential perfusion-correlated features to integrate into the model and found that the spatial distribution of the features and model output visualizations showed moderate to strong consistency with pulmonary functional imaging (i.e., perfusion SPECT imaging). Future collaboration within multidisciplinary teams could facilitate the non-invasive screening of PE and other perfusion-related lung diseases. Integrating our classifier into routine clinical practice could accelerate and optimize clinical decision-making.

References

[1] S. Z. Goldhaber and H. Bounameaux, "Pulmonary embolism and deep vein thrombosis," *The Lancet,* vol. 379, no. 9828, pp. 1835-1846, 2012.

[2] M. Di Nisio, N. van Es, and H. R. Büller, "Deep vein thrombosis and pulmonary embolism," *The Lancet,* vol. 388, no. 10063, pp. 3060-3073, 2016/12/17/ 2016.

[3] G. Pantaleo, N. Luigi, T. Federica, S. Paola, N. Margherita, and M. Tahir, "Amniotic Fluid Embolism: Review," *Current Pharmaceutical Biotechnology,* vol. 14, no. 14, pp. 1163-1167, 2013.

[4] N. Kucher, E. Rossi, M. De Rosa, and S. Z. Goldhaber, "Massive pulmonary embolism," (in eng), *Circulation,* vol. 113, no. 4, pp. 577-82, Jan 31 2006.

[5] A. Rahimtoola and J. D. Bergin, "Acute pulmonary embolism: An update on diagnosis and management," *Current Problems in Cardiology,* vol. 30, no. 2, pp. 61-114, 2005/02/01/ 2005.

[6] Y. K. Sung and J. A. Kline, "Unchanging Mortality from Pulmonary Embolism in the United States," *Ann Am Thorac Soc,* vol. 20, no. 11, pp. 1554-1556, Nov 2023.

[7] M. V. Huisman *et al.*, "2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS)," *European Heart Journal,* vol. 41, no. 4, p. 543À603, 2020.

[8] A. Squizzato, L. Galli, and V. E. A. Gerdes, "Point-of-care ultrasound in the diagnosis of pulmonary embolism," *Critical Ultrasound Journal,* vol. 7, no. 1, p. 7, 2015/05/27 2015.

[9] J. A. Kline, "Diagnosis and exclusion of pulmonary embolism," *Thrombosis research,* vol. 163, pp. 207-220, 2018.

[10] K. R. Beckett, A. K. Moriarity, and J. M. Langer, "Safe Use of Contrast Media: What the Radiologist Needs to Know," *RadioGraphics,* vol. 35, no. 6, pp. 1738-1750, 2015.

[11] I. H. Iftikhar, N. H. Iftikhar, M. Naeem, and A. BaHammam, "SPECT ventilation/perfusion imaging for acute pulmonary embolism: meta-analysis of diagnostic test accuracy," *Academic Radiology,* vol. 31, no. 2, pp. 706-717, 2024.

[12] T. Grüning, B. E. Drake, S. L. Farrell, and T. Nokes, "Three-year clinical experience with VQ SPECT for diagnosing pulmonary embolism: diagnostic performance," *Clinical Imaging,* vol. 38, no. 6, pp. 831-835, 2014.

[13] S. L. Hansen, R. de Nijs, J. Mortensen, and R. M. G. Berg, "Ventilation–perfusion SPECT versus CTPA in young adult females with suspected pulmonary embolism," *European Respiratory Journal,* vol. 55, no. 6, p. 2000448, 2020.

[14] E. Castillo *et al.*, "Quantifying pulmonary perfusion from noncontrast computed tomography," *Medical Physics,* vol. 48, no. 4, pp. 1804-1814, 2021.

[15] E. M. Porter *et al.*, "Synthetic pulmonary perfusion images from 4DCT for functional avoidance using deep learning," *Physics in Medicine & Biology,* vol. 66, no. 17, p. 175005, 2021/08/23 2021.

[16] G. Ren *et al.*, "Investigation of a Novel Deep Learning-Based Computed Tomography Perfusion Mapping Framework for Functional Lung Avoidance Radiotherapy," *Front Oncol,* vol. 11, p. 644703, 2021.

[17] G. Ren *et al.*, "Deep Learning-Based Computed Tomography Perfusion Mapping (DL-CTPM) for Pulmonary CT-to-Perfusion Translation," *Int J Radiat Oncol Biol Phys,* vol. 110, no. 5, pp. 1508-1518, Aug 1 2021.

[18] M. E. Mayerhoefer *et al.*, "Introduction to Radiomics," *Journal of Nuclear Medicine,* vol. 61, no. 4, pp. 488-495, 2020.

[19] B. Kocak, B. Baessler, R. Cuocolo, N. Mercaldo, and D. Pinto dos Santos, "Trends and statistics of artificial intelligence and radiomics research in Radiology, Nuclear Medicine, and Medical Imaging: bibliometric analysis," *European Radiology,* vol. 33, no. 11, pp. 7542-7555, 2023/11/01 2023.

[20] T. Guerrero *et al.*, "Quantification of regional ventilation from treatment planning CT," *International Journal of Radiation Oncology, Biology, Physics,* vol. 62, no. 3, pp. 630-634, 2005.

[21] E. Castillo, Y. Vinogradskiy, and R. Castillo, "Robust HU-based CT ventilation from an integrated mass conservation formulation," *Medical Physics,* vol. 46, no. 11, pp. 5036-5046, 2019.

[22] B. A. Simon, "Non-Invasive Imaging of Regional Lung Function using X-Ray Computed Tomography," *Journal of Clinical Monitoring and Computing,* vol. 16, no. 5, pp. 433-442, 2000/07/01 2000.

[23] J. M. Reinhardt, K. Ding, K. Cao, G. E. Christensen, E. A. Hoffman, and S. V. Bodas, "Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation," *Medical Image Analysis,* vol. 12, no. 6, pp. 752-763, 2008/12/01/ 2008.

[24] E. Castillo *et al.*, "Robust CT ventilation from the integral formulation of the Jacobian," *Medical Physics,* vol. 46, no. 5, pp. 2115-2125, 2019.

[25] G. Zhang, T. Huang, T. Dilling, C. Stevens, and K. Forster, "Derivation of high-resolution pulmonary ventilation using local volume change in four-dimensional CT data," in *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany: Vol. 25/4 Image Processing, Biosignal Processing, Modelling and Simulation, Biomechanics*, 2010: Springer, pp. 1834-1837.

[26] R. Castillo, E. Castillo, J. Martinez, and T. Guerrero, "Ventilation from four-dimensional computed tomography: density versus Jacobian methods," *Physics in Medicine & Biology,* vol. 55, no. 16, p. 4661, 2010.

[27] A. Szmul, T. Matin, F. V. Gleeson, J. A. Schnabel, V. Grau, and B. W. Papież, "Patch-based lung ventilation estimation using multi-layer supervoxels," *Computerized Medical Imaging and Graphics,* vol. 74, pp. 49-60, 2019.

[28] E. Castillo *et al.*, "On the spatial correlation between robust CT‐ventilation methods and SPECT ventilation," *Medical physics,* vol. 47, no. 11, pp. 5731-5738, 2020.

[29] P. Xue *et al.*, "Effective lung ventilation estimation based on 4D CT image registration and supervoxels," *Biomedical Signal Processing and Control,* vol. 79, p. 104074, 2023.

[30] J. Kipritidis *et al.*, "The VAMPIRE challenge: a multi‐institutional validation study of CT ventilation imaging," *Medical physics,* vol. 46, no. 3, pp. 1198-1217, 2019.

[31] G. Cazoulat, D. Owen, M. M. Matuszak, J. M. Balter, and K. K. Brock, "Biomechanical deformable image registration of longitudinal lung CT images using vessel information," *Physics in Medicine & Biology,* vol. 61, no. 13, p. 4826, 2016.

[32] G. Cazoulat, J. M. Balter, M. M. Matuszak, S. Jolly, D. Owen, and K. K. Brock, "Mapping lung ventilation through stress maps derived from biomechanical models of the lung," *Medical physics,* vol. 48, no. 2, pp. 715-723, 2021.

[33] Y. Zhong *et al.*, "Deriving ventilation imaging from 4 DCT by deep convolutional neural network," *Medical physics,* vol. 46, no. 5, pp. 2323-2329, 2019.

[34] Z. Liu *et al.*, "A deep learning method for producing ventilation images from 4DCT: First comparison with technegas SPECT ventilation," *Medical Physics,* vol. 47, no. 3, pp. 1249-1257, 2020.

[35] J. Kipritidis *et al.*, "Estimating lung ventilation directly from 4D CT Hounsfield unit values," *Med Phys,* vol. 43, no. 1, p. 33, Jan 2016.

[36] Y. Tian *et al.*, "Availability of a simplified lung ventilation imaging algorithm based on four-dimensional computed tomography," *Physica Medica,* vol. 65, pp. 53-58, 2019.

[37] K. J. Lafata, Z. Zhou, J.-G. Liu, J. Hong, C. R. Kelsey, and F.-F. Yin, "An Exploratory Radiomics Approach to Quantifying Pulmonary Function in CT Images," *Scientific Reports,* vol. 9, no. 1, p. 11509, 2019/08/08 2019.

[38] Z. Li *et al.*, "Quantitative texture analysis using machine learning for predicting interpretable pulmonary perfusion from non-contrast computed tomography in pulmonary embolism patients," *Respiratory Research,* vol. 25, no. 1, pp. 1-15, 2024.

[39] J. Hofmanninger, F. Prayer, J. Pan, S. Rohrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *Eur Radiol Exp,* vol. 4, no. 1, p. 50, Aug 20 2020.

[40] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans Pattern Anal Mach Intell,* vol. 34, no. 11, pp. 2274-82, Nov 2012.

[41] J. A. Parker *et al.*, "SNM practice guideline for lung scintigraphy 4.0," *J Nucl Med Technol,* vol. 40, no. 1, pp. 57-65, Mar 2012.

[42] A. M. Faught *et al.*, "Evaluating the Toxicity Reduction With Computed Tomographic Ventilation Functional Avoidance Radiation Therapy," *Int J Radiat Oncol Biol Phys,* vol. 99, no. 2, pp. 325-333, Oct 1 2017.

[43] A. Zwanenburg *et al.*, "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping," *Radiology,* vol. 295, no. 2, pp. 328-338, May 2020.

[44] X. Teng *et al.*, "Building reliable radiomic models using image perturbation," *Scientific Reports,* vol. 12, no. 1, p. 10035, 2022/06/16 2022.

[45] A. Zwanenburg *et al.*, "Assessing robustness of radiomic features by image perturbation," *Sci Rep,* vol. 9, no. 1, p. 614, Jan 24 2019.

[46] A. Ram, S. Jalal, A. S. Jalal, and M. Kumar, "A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases," *International Journal of Computer Applications,* vol. 3, no. 6, pp. 1-4, 2010.

[47] Y. H. Huang *et al.*, "Respiratory Invariant Textures From Static Computed Tomography Scans for Explainable Lung Function Characterization," *J Thorac Imaging,* vol. 38, no. 5, pp. 286-296, Sep 1 2023.

[48] D. S. Kerby, "The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation," *Comprehensive Psychology,* vol. 3, 2014.

[49] L. Ostroumova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Neural Information Processing Systems*, 2017,

[50] D. S. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," *Proceedings of the 1968 23rd ACM national conference,* pp. 517-524, 1968.

[51] D. Levin, "The approximation power of moving least-squares," *Mathematics of computation,* vol. 67, no. 224, pp. 1517-1531, 1998.

[52] F. Kiessling, "The changing face of cancer diagnosis: From computational image analysis to systems biology," *European Radiology,* vol. 28, no. 8, pp. 3160-3164, 2018/08/01 2018.

[53] E. Capobianco and J. Deng, "Radiomics at a Glance: A Few Lessons Learned from Learning Approaches," *Cancers,* vol. 12, no. 9, p. 2453, 2020.

[54] Y. Huang *et al.*, "Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non—small cell lung cancer," *Radiology,* vol. 281, no. 3, pp. 947-957, 2016.

[55] G.-W. Ji *et al.*, "Biliary tract cancer at CT: a radiomics-based model to predict lymph node metastasis and survival outcomes," *Radiology,* vol. 290, no. 1, pp. 90-98, 2019.

[56] Y. Cui *et al.*, "Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer," *European radiology,* vol. 29, pp. 1211-1220, 2019.

[57] Z. Zhang, K. M. Ho, and Y. Hong, "Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care," *Critical Care,* vol. 23, pp. 1-10, 2019.

[58] A. Hakkak Moghadam Torbati, S. Pellegrino, R. Fonti, R. Morra, S. De Placido, and S. Del Vecchio, "Machine Learning and Texture Analysis of [18F] FDG PET/CT Images for the Prediction of Distant Metastases in Non-Small-Cell Lung Cancer Patients," *Biomedicines,* vol. 12, no. 3, p. 472, 2024.

[59] J. A. Balch *et al.*, "Machine learning–enabled clinical information systems using fast healthcare interoperability resources data standards: scoping review," *JMIR Medical Informatics,* vol. 11, p. e48297, 2023.