

PAPER • OPEN ACCESS


Non-contrasted computed tomography (NCCT) based chronic thromboembolic pulmonary hypertension (CTEPH) automatic diagnosis using cascaded network with multiple instance learning

To cite this article: Mayang Zhao *et al* 2024 *Phys. Med. Biol.* **69** 185011


View the [article online](#) for updates and enhancements.

You may also like

- [Lightweight CNN architecture design for rolling bearing fault diagnosis](#)
Lingli Jiang, Changzhi Shi, Heshan Sheng *et al.*
- [Systematic Product Development of Control and Diagnosis Functionalities](#)
R Stetter and A Simundsson
- [Acquisition of missile fault diagnosis knowledge based on incomplete information of flow graph](#)
Zhaozheng Liu, Mingqing Xiao, Haizhen Zhu *et al.*



Joining forces:
One complete
QA solution for
Dosimetry with
myQA[®], QUASAR[™]
and Radcal[®]!



The diagram is a circular graphic with a dark background and a pattern of small, colorful dots. It consists of four colored segments arranged in a circle: a dark blue segment at the top left labeled 'Machine QA', a green segment at the top right labeled 'Patient Specific QA', a light blue segment at the bottom labeled 'Medical Imaging QA', and a pink segment at the center labeled 'Risk Management'. In the center of the pink segment is a white circle containing a green wireframe human head.



PAPER

OPEN ACCESS

RECEIVED
29 April 2024REVISED
6 August 2024ACCEPTED FOR PUBLICATION
27 August 2024PUBLISHED
13 September 2024

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Non-contrasted computed tomography (NCCT) based chronic thromboembolic pulmonary hypertension (CTEPH) automatic diagnosis using cascaded network with multiple instance learning

Mayang Zhao¹ , Liming Song¹, Jiarui Zhu¹, Ta Zhou¹, Yuanpeng Zhang¹, Shu-Cheng Chen², Haojiang Li³, Di Cao³, Yi-Quan Jiang⁴, Waiyin Ho⁵, Jing Cai^{1,*} and Ren Ge^{1,*}

¹ Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China, People's Republic of China

² School of Nursing, Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China, People's Republic of China

³ Department of Radiology, State Key Laboratory of Oncology in South China, Collaborative Innovation Centre for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Centre, Guangzhou, People's Republic of China

⁴ Department of Minimally Invasive Interventional Therapy, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, People's Republic of China

⁵ Department of Nuclear Medicine, Queen Mary Hospital, Hong Kong Special Administrative Region of China, People's Republic of China

* Authors to whom any correspondence should be addressed.

E-mail: jing.cai@polyu.edu.hk and gary-ge.ren@polyu.edu.hk

Keywords: chronic thromboembolic pulmonary hypertension, non-contrasted computed tomography, multiple instance learning, cascade network, automatic diagnosis

Abstract

Objective. The diagnosis of chronic thromboembolic pulmonary hypertension (CTEPH) is challenging due to nonspecific early symptoms, complex diagnostic processes, and small lesion sizes. This study aims to develop an automatic diagnosis method for CTEPH using non-contrasted computed tomography (NCCT) scans, enabling automated diagnosis without precise lesion annotation. **Approach.** A novel cascade network (CN) with multiple instance learning (CNMIL) framework was developed to improve the diagnosis of CTEPH. This method uses a CN architecture combining two Resnet-18 CNN networks to progressively distinguish between normal and CTEPH cases. Multiple instance learning (MIL) is employed to treat each 3D CT case as a 'bag' of image slices, using attention scoring to identify the most important slices. An attention module helps the model focus on diagnostically relevant regions within each slice. The dataset comprised NCCT scans from 300 subjects, including 117 males and 183 females, with an average age of 52.5 ± 20.9 years, consisting of 132 normal cases and 168 cases of lung diseases, including 88 cases of CTEPH. The CNMIL framework was evaluated using sensitivity, specificity, and the area under the curve (AUC) metrics, and compared with common 3D supervised classification networks and existing CTEPH automatic diagnosis networks. **Main results.** The CNMIL framework demonstrated high diagnostic performance, achieving an AUC of 0.807, accuracy of 0.833, sensitivity of 0.795, and specificity of 0.849 in distinguishing CTEPH cases. Ablation studies revealed that integrating MIL and the CN significantly enhanced performance, with the model achieving an AUC of 0.978 and perfect sensitivity (1.000) in normal classification. Comparisons with other 3D network architectures confirmed that the integrated model outperformed others, achieving the highest AUC of 0.8419. **Significance.** The CNMIL network requires no additional scans or annotations, relying solely on NCCT. This approach can improve timely and accurate CTEPH detection, resulting in better patient outcomes.

List of abbreviations

Abbreviation	Definition
CTEPH	Chronic thromboembolic pulmonary hypertension
Non-CTEPH	Other lung diseases apart from CTEPH
NCCT	Non-contrasted computed tomography
CNMIL	Cascaded network with multiple instance learning
CN	Cascade network
MIL	Multiple instance learning
CTA	Computed tomography angiography
V/Q	Ventilation-perfusion scans
SPECT	Single-photon emission computed tomography
CAD	Computer-aided detection
SOTA	State-of-the-art
AUC	The area under the curve
SD	Standard deviation
Resnet-I	Resnet-18 2D model used in the first stage without incorporating MIL
Resnet-I_MIL	Resnet-18 2D model used in the first stage incorporating MIL
Resnet	Resnet-18 2D model without using CN or MIL
Resnet_CN	Resnet-18 2D model using CN but not using MIL
Resnet-I_MIL	Resnet-18 2D model using CN but not using MIL
Resnet_CN	Resnet-18 2D model not using CN but using MIL
Resnet_CNMIL	Resnet-18 2D model using both CN and MIL

1. Introduction

CTEPH is a progressive pulmonary vascular disease characterized by the obstruction of the pulmonary artery and arterioles, attributed to near-end thrombus embolism or distal vascular remodeling (Fedullo *et al* 2011). This obstruction elevates right heart load and pulmonary vascular resistance, culminating in right heart failure and posing a severe threat to the patient's life (Price *et al* 2010). However, its low occurrence and limited research on CTEPH pathophysiology contribute to delayed diagnoses and frequent misdiagnoses (Price *et al* 2010). However, the early symptoms of CTEPH, such as dyspnea, hemoptysis, and chest pain, are not considered nonspecific (Narechania *et al* 2020). Additionally, conventional CTEPH diagnostic methods, such as ventilation/perfusion scans or CTA, are not commonly used in routine evaluations (Gopalan *et al* 2017). These diagnostic techniques often lead to delays and misdiagnoses. Delays in diagnosis may result in the development of more distal vasculopathy and right heart dysfunction due to elevated pulmonary artery pressure (D'Armini 2015). Left untreated, CTEPH has a mean life expectancy of less than 3 years (Delcroix *et al* 2016). Diagnostic errors not only cause patients to miss the optimal treatment window but also pose a direct threat to their health through inappropriate treatment plans (Deano *et al* 2013). Therefore, a timely and accurate diagnosis of CTEPH patients is of utmost significance.

Confirming early-stage CTEPH presents a significant challenge, as it requires specialized equipment such as V/Q or CTA. However, the high cost and additional radiation associated with these methods hinder their routine use in early-stage CTEPH cases where obvious disease characteristics are lacking. When individuals first experience symptoms, they commonly seek medical care due to discomfort. However, due to the aforementioned challenges, doctors often fail to consider the possibility of CTEPH and misdiagnose it as other conditions. Research indicates that more than 50% of cases result in misdiagnosis or missed diagnosis (Lang 2015, Kim *et al* 2019). After a period of treatment, patients may develop more obvious symptoms, prompting doctors to consider the possibility of CTEPH and conduct the V/Q scan or computed CTA to confirm. However, by the time of CTEPH diagnosis, an average of 14.1 months has passed since the patient's initial discomfort. A study involving 151 patients revealed that 66 had already exhibited symptoms of heart failure, indicating a poor prognosis (Galiè *et al* 2016). These statistics underscore the severity of the condition and emphasize the urgent need for improved diagnostic methods.

CAD has emerged as a valuable tool in the medical field (Pepke-Zaba *et al* 2011, Lee *et al* 2017, Nam *et al* 2020, Ren *et al* 2021). It utilizes statistical modeling, machine learning, or deep learning techniques to assist in the detection and analysis of abnormalities or markers of disease (LeCun *et al* 2015, Li *et al* 2019, Ribeiro *et al* 2020, Wang *et al* 2020). In the realm of CTEPH, there is a paucity of literature on automatic diagnosis of CTEPH, focused on employing quantitative analysis techniques for CTEPH abnormal signal detection in V/Q or CTA images. For instance, one study utilized MATLAB to compare mean volume difference values between ventilation and perfusion images (Seiffert *et al* 2020). Another investigation applied a graph-cut

method filter to explore the differences in pulmonary vascular morphology and densitometry between arteries and veins in CTA images, aiming to discern CTEPH cases (Zhai *et al* 2023). Additionally, a separate study employed the Lung Density Analysis tool to investigate the associations between mosaic perfusion patterns and CTEPH (Cerny *et al* 2023). Furthermore, a lung graph deep learning model was utilized to quantify the texture of lung parenchyma for achieving fully automated classification of CTEPH. Despite these advancements, existing CTEPH automatic diagnosis methods exhibit certain limitations: (1) the current automatic diagnosis methods primarily rely on V/Q scans or CTA. However, the accessibility of V/Q scans is often limited due to their cost and availability in hospitals. Additionally, CTA carries risks associated with radiation exposure and potential allergic reactions, making it less desirable as a primary diagnostic tool. (2) Both semi-automatic and fully automatic CAD methods mentioned earlier require meticulous delineation of the affected areas. For example, the set of CTA images from a single patient necessary for diagnosis includes an average of 600 submillimeter slices in the axial plane. In order to detect CTEPH, it is necessary to view most of them carefully and to pay attention to the presence of small formations inside the vessels (stenoses and occlusions).

Addressing these limitations is indeed crucial to ensure wider accessibility, minimize risks, and improve the efficiency of automatic diagnosis methods for CTEPH. In our research project, we aim to develop a CNMIL based on NCCT images for CTEPH automatic diagnosis. This innovative approach offers several key contributions:

- (1) Automation without precise annotations: The CNMIL framework enables automated CTEPH diagnosis without the need for precise lesion annotation, relying instead on case-level reports. This significantly reduces the burden and time required for manual labeling by medical professionals.
- (2) Utilization of common NCCT scans: The framework leverages commonly acquired NCCT images, eliminating the need for additional scans such as CTA, which involves radiation exposure, or the use of specialized equipment like V/Q scans. This enhances patient safety and makes the method more practical for widespread clinical use.
- (3) Enhanced diagnostic accuracy with focused attention: By integrating MIL with a CN architecture and incorporating an attention module, our method achieves superior diagnostic performance compared to traditional 3D supervised classification networks and existing CTEPH diagnosis models. This combination allows the model to focus on the most diagnostically relevant regions within each image slice, resulting in high sensitivity, specificity, and AUC metrics, particularly excelling in distinguishing CTEPH cases.

2. Materials and method

2.1. Study design

The overall design of our framework is illustrated in figure 1. We used a CN with five-fold cross-validation for CTEPH prediction. Initially, data was input into the model case by case and processed through ResNet18 CNN networks to extract features. These features were processed by an attention mechanism, calculating weights for each slice, and fully connected layers, yielding scores. These scores were summed for the final prediction, distinguishing between normal and abnormal cases in the first stage. In the second stage, the model further classified between CTEPH and non-CTEPH cases. Detailed steps are provided below.

2.2. Study cohort

In our study, we analyzed clinical NCCT and Q-SPECT from two cohorts. The first cohort included 176 cases from Queen Mary Hospital, consisting of 88 patients with CTEPH, 8 health cases, and 80 patients with other lung diseases such as asthma, carcinoma, and others (see table 1 and figure 2). The second cohort consisted of 124 health cases obtained from The First Affiliated Hospital of Nanchang University.

2.3. Image acquisition

For the first cohort from Queen Mary Hospital:

- NCCT was performed using a General Electric BrightSpeed Elite CT Scanner with 16 detector rows. The scanning parameters were set as follows: the tube voltage was 120 kV, and the tube current ranged from 30 to 397 mA. The slice thickness was 1.25 mm, pixel spacing was 0.98 mm, the slice interval was 5 mm, and the image matrix size was 512×512 .

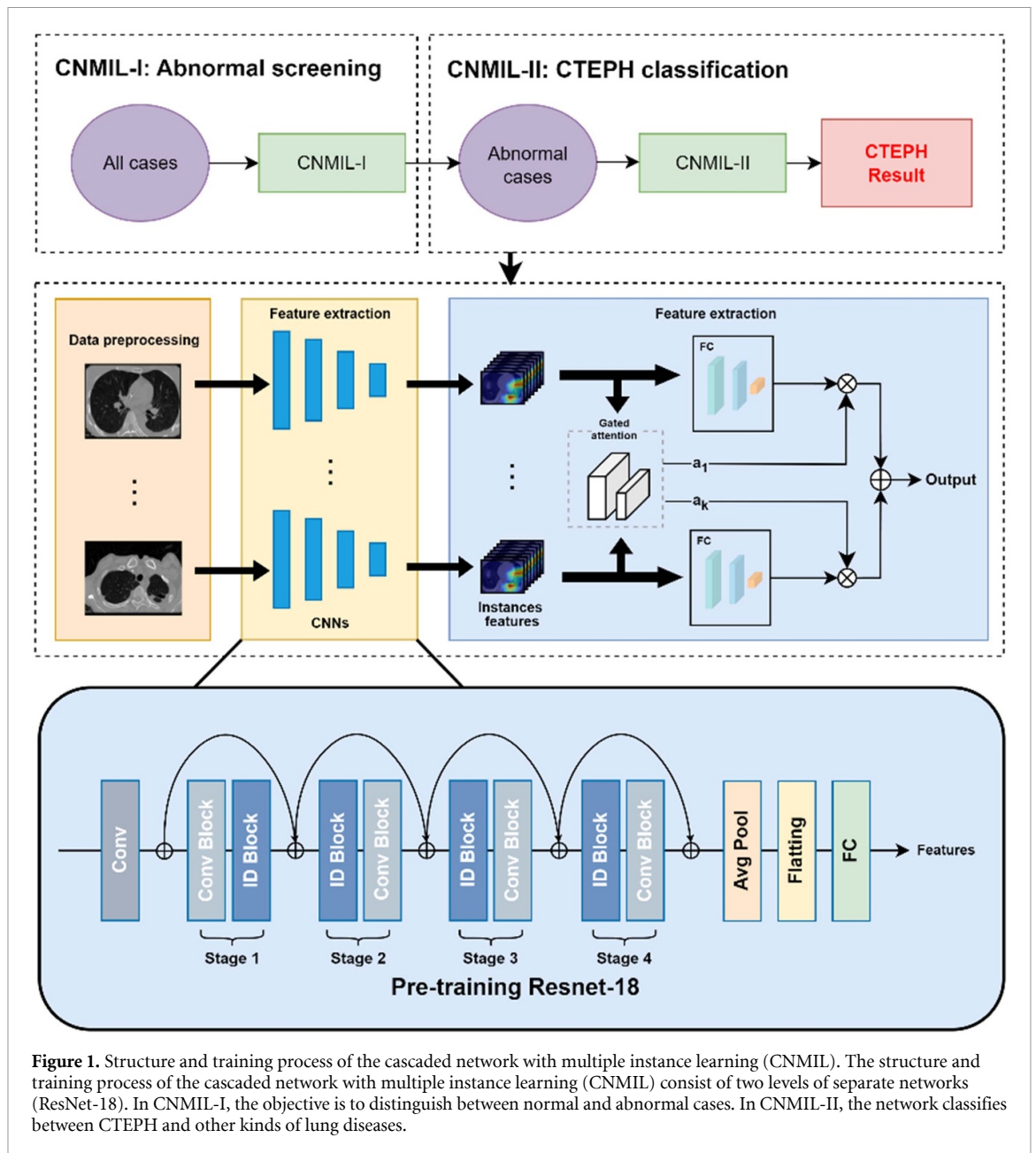


Figure 1. Structure and training process of the cascaded network with multiple instance learning (CNMIL). The structure and training process of the cascaded network with multiple instance learning (CNMIL) consist of two levels of separate networks (ResNet-18). In CNMIL-I, the objective is to distinguish between normal and abnormal cases. In CNMIL-II, the network classifies between CTEPH and other kinds of lung diseases.

Table 1. Patient characteristics.

Category	Number
Number of images	300
Sex	
Male	117
Female	183
Age	
Mean \pm SD of age (years)	52.5 \pm 20.9
Diagnosis	
Healthy	132
CTEPH	88
Non-CTEPH	80

SD = standard deviation, CTEPH = chronic thromboembolic pulmonary hypertension, Non-CTEPH = patients with pulmonary disease but without CTEPH.

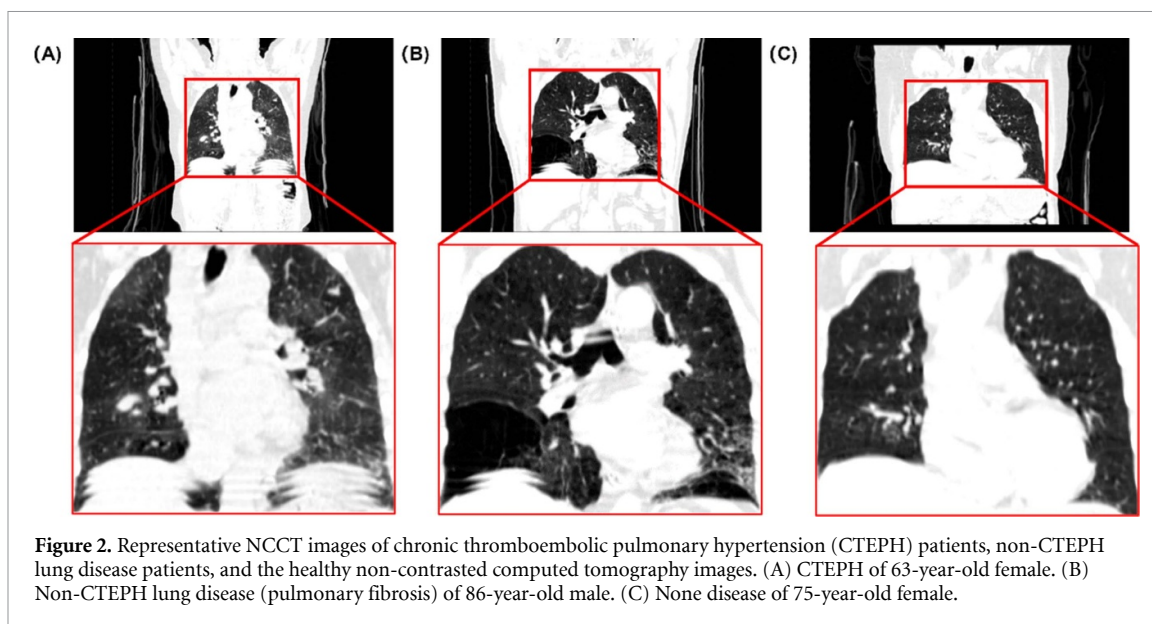


Figure 2. Representative NCCT images of chronic thromboembolic pulmonary hypertension (CTEPH) patients, non-CTEPH lung disease patients, and the healthy non-contrasted computed tomography images. (A) CTEPH of 63-year-old female. (B) Non-CTEPH lung disease (pulmonary fibrosis) of 86-year-old male. (C) None disease of 75-year-old female.

- Q-SPECT images were acquired using a GE Discovery 670 SPECT/CT scanner at a frame rate of 30 s/frame; the patient was injected with 3 mCi (111 MBq) of the radiopharmaceutical macro-aggregated albumin before imaging.

For the second cohort from The First Affiliated Hospital of Nanchang University:

- NCCT was performed using a General Electric LightSpeed VCT CT Scanner with 64 detector rows. The scanning parameters were as follows: the tube voltage was 120 kV, and the tube current ranged from 30 to 397 mA. The slice thickness was 1.25 mm, pixel spacing was 0.82 mm, the slice interval was 1.25 mm, and the image matrix size was 512×512 .

2.4. Image interpretation

All images were reviewed by an experienced nuclear medicine physician. In this study, the MSKCC Q-SPECT/CT and Modified PLOPED II criteria were used to determine the presence of CTEPH in Q-SPECT/CT scans. According to the MSKCC Q-SPECT/CT criteria, any wedge-shaped peripheral perfusion defect occupying more than 50% of a lung segment without corresponding pulmonary parenchymal or pleural disease was considered positive (Lu *et al* 2014). In the Modified PLOPED II criteria, two or more large mismatched segmental perfusion defects or the arithmetic equivalent of moderate and/or large defects were considered positive (Sostman *et al* 2008). Perfusion defects that did not fulfill the above criteria were considered negative for CTEPH.

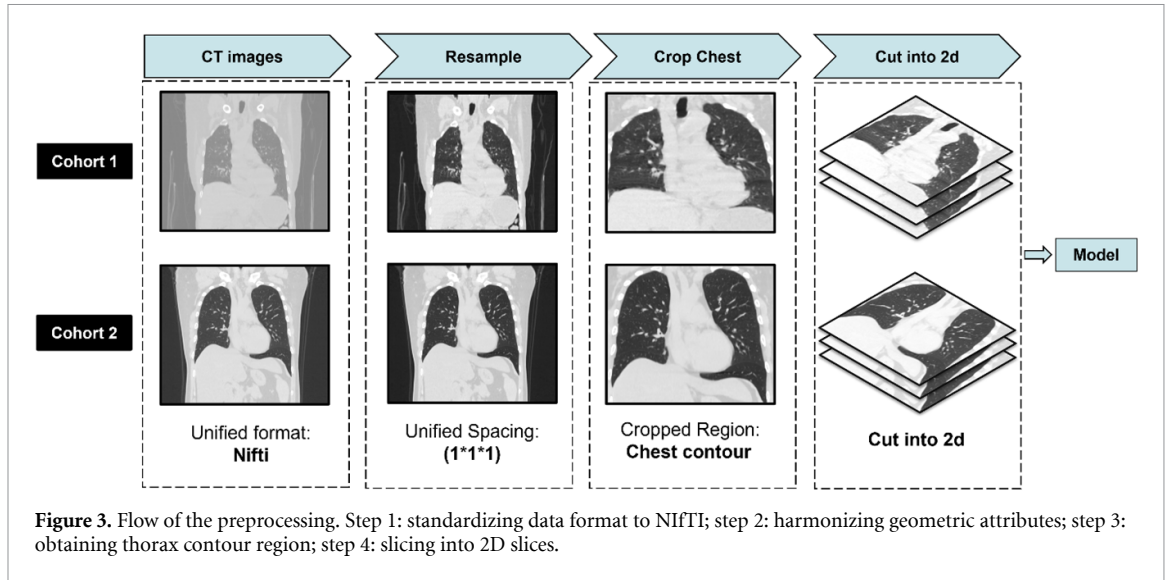
2.5. Data preprocessing

Due to our NCCT data being sourced from two different cohorts, the data preprocessing outline is shown in figure 3. The preprocessing pipeline for CT images involves four key steps applied to both cohorts. First, raw 3D CT images are acquired. These images are then resampled to standardize spatial resolution, ensuring consistency across all scans. Next, the images are cropped to focus on the chest region, isolating the thoracic cavity and removing irrelevant anatomical areas. Following this, the cropped 3D images are sliced into 2D images since our model is based on a 2D classification model with MIL. Finally, these 2D slices are aggregated and prepared as input for the 2D MIL model. This pipeline standardizes and transforms the CT images, enhancing the model's accuracy and efficiency.

2.6. Cascaded network

To address the complexity and potential pitfalls of directly performing a three-class classification for identifying CTEPH (Wang *et al* 2023), we employed the CNMIL, trained with a cascaded learning strategy, as illustrated in figure 1. This strategy involves an initial screening and exclusion of healthy cases, followed by a focused classification of the remaining abnormal cases into CTEPH and non-CTEPH categories.

The cascaded network consists of two levels of convolutional networks with identical structures, each operating independently. The first stage, CNMIL-I, is designed to differentiate lung disease cases from normal cases, mimicking real-life clinical practice. The second stage, CNMIL-II, is used to distinguish



CTEPH cases from other abnormal cases. By dividing the CTEPH diagnosis task into two levels, each neural network can function more effectively and accurately. By isolating the task of identifying specific disease conditions to this stage, the model can focus on the finer distinctions between CTEPH and other abnormal cases. This hierarchical approach allows each model to operate within a more narrowly defined scope, enhancing overall classification performance and reducing the risk of overfitting.

2.7. MIL with attention module

In training 2D supervised models for disease classification, mainstream deep learning approaches are heavily reliant on expert-level manual labeling (Sarker 2021). Each slice in the dataset is annotated as either ‘positive’ or ‘negative’ based on these manual segmentations (Diogo *et al* 2023). However, due to time constraints, physicians may not always have the capacity for extensive segmentation tasks (Zhou *et al* 2023). Additionally, 2D networks process input layer by layer, often ignoring the inter-layer relationships, which can lead to suboptimal results (Avesta *et al* 2023). In contrast, 3D networks consider these relationships, providing a more holistic view (Kulkarni *et al* 2021). However, for diseases like CTEPH, where only a few slices out of a case’s 200 slices may show abnormalities, these small lesions often do not provide sufficient visual cues to distinguish them from the background or other similar categories, making training 3D models challenging (Shahin *et al* 2022). These models are prone to overfitting and require large amounts of data to converge effectively. Consequently, alternative methods are needed.

MIL is a weakly supervised method well-suited to this scenario. In MIL, each patient’s CT image is regarded as a ‘bag,’ with each slice considered an ‘instance,’ which aligns well with the structure of CT cases and slices. Each case is inputted slice-by-slice, facilitating more efficient and practical training. These slices are processed through networks that extract their corresponding features, denoted as $H = \{h_1, \dots, h_k\}$, where h_k represents each slice, and H is the embedding features of the case. Then we use the following MIL pooling:

$$z = \sum_{k=1}^K a_k h_k . \quad (1)$$

Additionally, it is necessary to assign different attention scores to different slices, rather than treating each slice equally as in 3D classification. We introduce the following attention module. The attention weight a_k for each instance k are calculated as follows:

$$\alpha_k = \frac{\exp\{\omega^\top \tanh(\mathcal{V}h_k^\top)\}}{\sum_{j=1}^k \exp\{\omega^\top \tanh(\mathcal{V}h_j^\top)\}} \quad (2)$$

where:

- $\omega \in \mathbb{R}^{L \times 1}$ and $\mathcal{V} \in \mathbb{R}^{L \times M}$ are represent the parameters of the convolutional network.
- j is the index that runs over all slices, from 1 to K , in a case.
- The superscript \top denotes the transpose of a matrix or vector.

In this formulation, acts as a weight vector that helps to determine the importance of each hidden representation h_k , while \mathcal{V} is a transformation matrix that projects h_k . Within parameter \mathcal{V} , the values of L

and M are hyperparameters that have been set to 500 and 128, respectively, with the reasons for these choices to be discussed later. The output of the formulation is a predicted value ranging from 0 to 1. By combining multiple α values from the attention mechanism with the predicted values from the network, we can obtain scores for each slice, following the same process for each slice. These scores are then summed to obtain the final prediction for the entire case in Resnet_CN-I, distinguishing between normal and abnormal. In Resnet_CN-I_MIL, the network classifies between CTEPH and non-CTEPH cases. Finally, based on the Binary Cross Entropy Loss, the loss function is calculated using the following formula:

$$\text{Loss}_{\text{BCE}}(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) . \quad (3)$$

Here, \hat{y} is the final prediction score.

2.8. Evaluation metrics

For each fold, we randomly divided the dataset into new training and testing sets, enabling us to retrain the model and evaluate its performance. We utilized various metrics, including AUC, accuracy, sensitivity, and specificity. Sensitivity represents the proportion of true positives that the test accurately detects. Mathematically, it can be calculated as:

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} . \quad (4)$$

Specificity indicates its ability to accurately identify individuals who do not have the condition. It represents the proportion of true negatives that are correctly recognized by the test:

$$\text{Specificity} = \frac{\text{True negatives}}{\text{False positives} + \text{True negatives}} . \quad (5)$$

Accuracy is calculated by dividing the sum of true positives and true negatives by the total number of cases:

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{Positive} + \text{Negative}} . \quad (6)$$

ResNet-18 (He *et al* 2016) served as the deep learning model for the proposed method, with transfer learning applied to classify NCCT images of CTEPH, non-CTEPH lung disease, and healthy individuals. To mitigate overfitting and performance degradation, we utilized transfer learning (Raghu *et al* 2019) using publicly available weights pre-trained on MedicalNet (Chen *et al* 2019) and applied data augmentation techniques (Chlap *et al* 2021), including $\pm 5^\circ$ rotation, horizontal flipping, and Gaussian noise. The training process was conducted using the SGD optimizer, starting with a learning rate of 0.005, which was progressively adjusted over 200 epochs using a StepLR scheduler that reduced the rate by a factor of 0.1 every 40 epochs. To mitigate potential biases from dataset partitioning, we conducted three separate tests for all of our experiments, which was particularly important due to the relatively small size of our dataset. The experiments were performed using Python 3.10 in the PyCharm IDE, with PyTorch 2.3.1 as the machine learning framework, on a system equipped with an NVIDIA GeForce GTX 3090 with 24GB VRAM. This setup ensured efficient model training and evaluation, providing reliable results across all experiments.

3. Results

Our proposed CNMIL architecture is flexible and does not restrict the choice of CNN backbone. For comparative purposes, we evaluated CNMIL with backbones having the smallest number of parameters, specifically ResNet-18, EfficientNet-B0, Inception v1, and Xception, chosen for their efficient resource use, reduced training time, and lower risk of overfitting. Referred to as ResNet, EfficientNet, Inception, and Xception, these configurations were compared with their corresponding 3D classification networks in section 3.1. Ultimately, ResNet was selected for further experiments. In section 3.2, we benchmarked CNMIL with ResNet-18 against SOTA methods for CTEPH classification, explaining our choice and presenting the results. Since CNMIL uses 2D inputs, we compared outcomes under different cutting perspectives in section 3.3. Section 3.4 displays Grad-CAM activation maps and corresponding SPECT perfusion images from the best test results. We discussed hyperparameter settings for MIL with attention mechanisms in section 3.5 and presented the ablation study results in section 3.6.

Table 2. Comparison of results using different 2D networks with CNMIL and their corresponding 3D networks. The bolded CNMIL parts indicate the best performance across the metrics compared to their non-CNMIL counterparts.

Modality	AUC	Accuracy	Sensitivity	Specificity
ResNET_3d	0.7633	0.7900	0.6250	0.8585
ResNET_CNMIL	0.8069	0.8333	0.7955	0.8491
EfficientNET_3d	0.7648	0.7500	0.6390	0.7989
EfficientNET_CNMIL	0.8148	0.8233	0.7727	0.8443
Inception_3d	0.7995	0.7667	0.7386	0.7783
Inception_CNMIL	0.8305	0.8300	0.7841	0.8491
Xception_3d	0.8235	0.7700	0.7500	0.7783
Xception_CNMIL	0.8419	0.8300	0.8182	0.8349

3.1. Comparative performance analysis of 2D classification networks using CNMIL vs. their 3D counterparts

Table 2 presents a comparison of performance metrics for various common 2D classification networks using the CNMIL architecture as the backbone against their corresponding 3D networks, evaluating AUC, Accuracy, Sensitivity, and Specificity. Referring to the confusion matrix results in figure 4, the results show that the proposed CNMIL method consistently surpasses the 3D networks across all metrics for each backbone network. Specifically, for ResNET, CNMIL increased AUC from 0.7633 to 0.8069, Accuracy from 0.7900 (237/300) to 0.8333 (250/300), and Sensitivity from 0.6250 (55/88) to 0.7955 (70/88), with a slight decrease in Specificity from 0.8585 (182/212) to 0.8491 (180/212). For EfficientNET, CNMIL enhanced AUC from 0.7648 to 0.8148, Accuracy from 0.7500 (225/300) to 0.8233 (247/300), Sensitivity from 0.6390 (56/88) to 0.7727 (68/88), and Specificity from 0.7989 (169/212) to 0.8443 (179/212). For Inception, CNMIL raised AUC from 0.7995 to 0.8305, Accuracy from 0.7667 (230/300) to 0.8300 (249/300), Sensitivity from 0.7386 (65/88) to 0.7841 (69/88), and Specificity from 0.7783 (165/212) to 0.8491 (180/212). Similarly, for Xception, CNMIL improved AUC from 0.8235 to 0.8419, Accuracy from 0.7700 (231/300) to 0.8300 (249/300), Sensitivity from 0.7500 (66/88) to 0.8182 (72/88), and Specificity from 0.7783 (165/212) to 0.8349 (177/212). Additionally, according to figure 4(B), CNMIL achieves equal or better AUC results in almost all experiments compared to the 3D networks, except for fold 1 where the AUC for ResNET_3D decreased from 0.78 to 0.75 with ResNET_CNMIL. These findings clearly demonstrate that the CNMIL method yields superior performance compared to the corresponding 3D networks, underscoring its effectiveness in enhancing classification performance for CTEPH detection.

3.2. Comparative analysis of different CTEPH classification models

We chose ResNET as the backbone for our CNMIL architecture primarily due to its significant performance improvements in Sensitivity and Accuracy, which are critical metrics for the automatic diagnosis of CTEPH. According to table 2, ResNET achieved the highest Accuracy among all backbones, improving from 0.7900 to 0.8333 with CNMIL. More importantly, the Sensitivity of ResNET showed a remarkable increase from 0.6250 to 0.7955, representing an improvement of 27.28%. This substantial enhancement in Sensitivity is crucial for accurately identifying true positive cases of CTEPH. Therefore, given the superior performance in these key metrics, all subsequent experiments were conducted using ResNET-18 as the backbone for our CNMIL architecture.

In table 3, we present a comparison of the classification performance of our framework (Resnet_CNMIL) and SOTA automatic detection methods for CTEPH within CTA, V/Q or NCCT images. Our framework, boasting an accuracy of 0.833, sensitivity of 0.795, and specificity of 0.849, exhibits superior performance in CTEPH classification.

The performance of our framework significantly surpasses the AUC of 0.786 reported in research (Zhou *et al* 2019) based on CTA, as well as the Accuracy of 0.670, Sensitivity of 0.670, and Specificity of 0.800 reported in research (Seiffert *et al* 2020) based on V/Q. Additionally, our framework outperforms the Accuracy of 0.760, Sensitivity of 0.780, and Specificity of 0.750 reported in research (Jimenez-del-Toro *et al* 2020) based on NCCT as well.

3.3. Comparative analysis of different views in obtaining 2D slices

To ensure that the results were not influenced by the orientation of the slices, we conducted experiments using three different views: sagittal, coronal, and transverse. As shown in table 4, for the sagittal view, the model achieved an AUC of 0.791, an accuracy of 0.797, a sensitivity of 0.726, and a specificity of 0.825. In the coronal view, the performance improved, with an AUC of 0.807, an accuracy of 0.823, a sensitivity of 0.784, and a specificity of 0.839. The transverse view yielded the highest performance metrics, with an AUC of

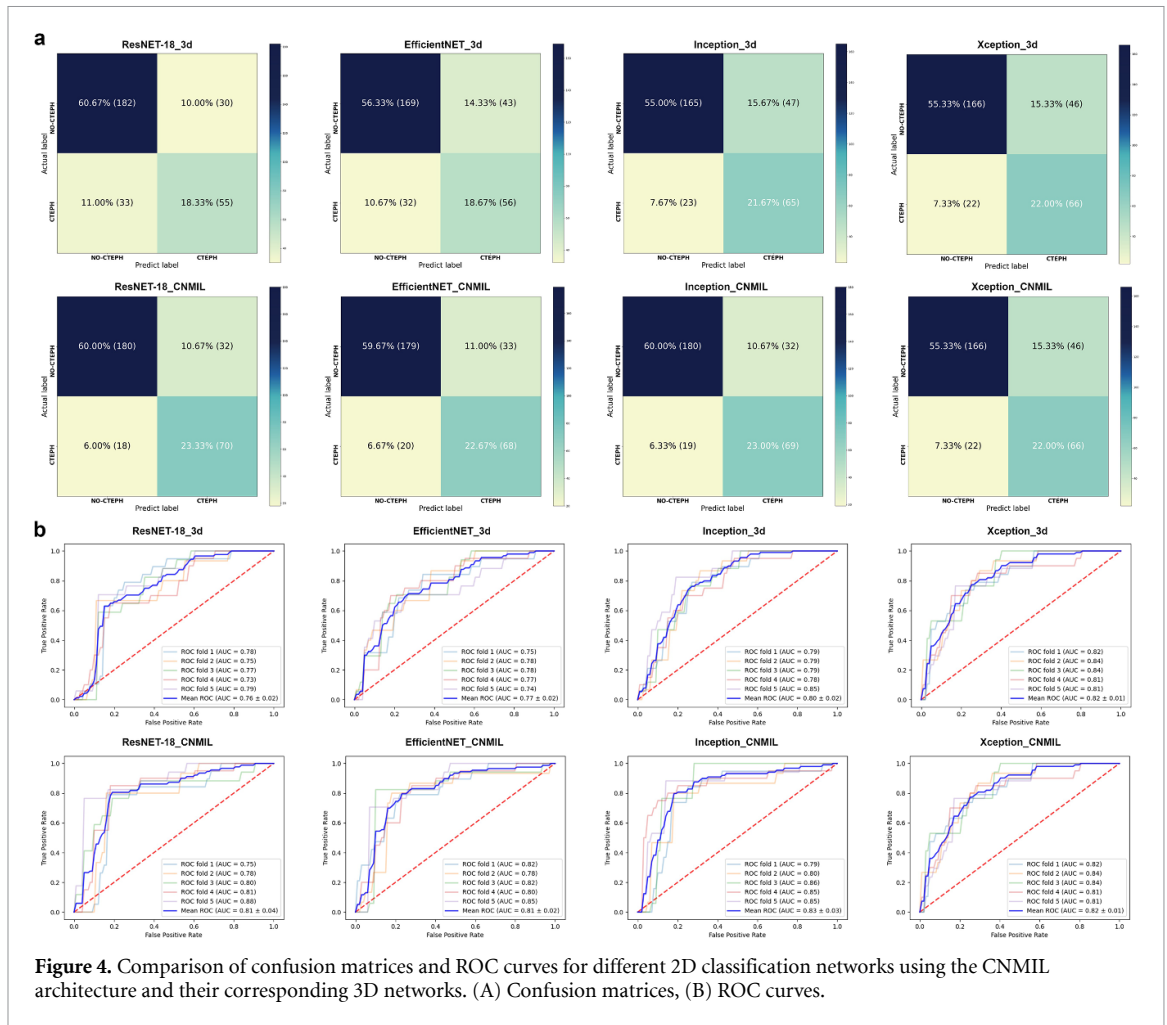


Figure 4. Comparison of confusion matrices and ROC curves for different 2D classification networks using the CNMIL architecture and their corresponding 3D networks. (A) Confusion matrices, (B) ROC curves.

Table 3. Comparison with state-of-the-art automatic detection methods for CTEPH.

Benchmark	AUC	Accuracy	Sensitivity	Specificity
Zhou et al (CTA)	0.786			
Seiffert et al (V/Q)	—	0.670	0.670	0.800
Oscar et al (NCCT)	—	0.760	0.780	0.750
Resnet_CNML (Ours)	0.807	0.833	0.795	0.849

Table 4. Comparison of different views.

View	AUC	Accuracy	Sensitivity	Specificity
Sagittal	0.791	0.797	0.726	0.825
Coronal	0.807	0.823	0.784	0.839
Transverse	0.807	0.833	0.795	0.849

V/Q = ventilation/perfusion scan, CTA = computed tomography angiography, NCCT = non-contrasted computed tomography.

0.807, an accuracy of 0.833, a sensitivity of 0.795, and a specificity of 0.849. These results indicate that while all views show strong performance, the transverse view achieves the highest accuracy and specificity, demonstrating that our model maintains robust and consistent performance across different anatomical slice orientations.

3.4. Interpreting the model: comparing CAM images with SPECT perfusion images

To further understand the model’s decision-making process and ensure its focus on clinically relevant areas, we utilized Gradient-weighted Class Activation Mapping to generate attention maps for both CTEPH and

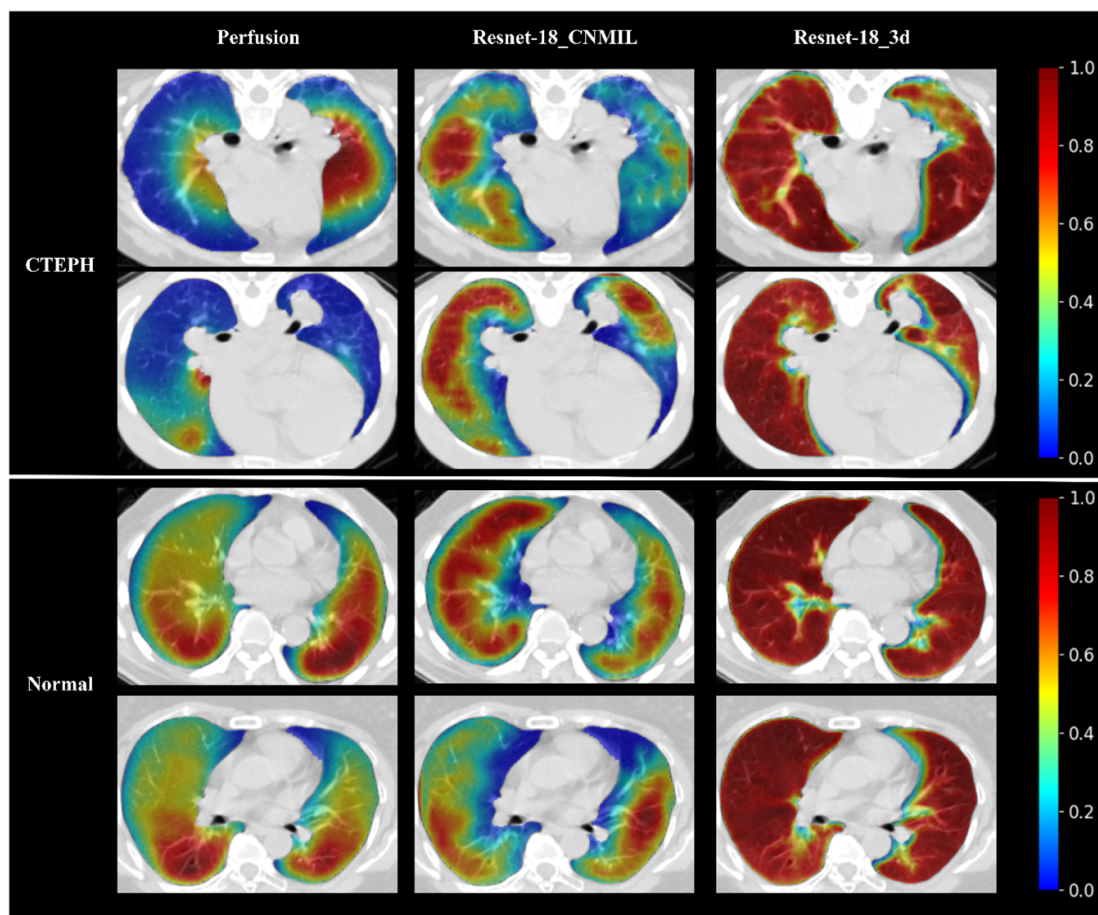


Figure 5. Grad-CAM activation maps from SPECT perfusion images, ResNET-18_3d, and ResNET-18_CN MIL for CTEPH and normal cases. The first and second rows show CTEPH patients, while the third and fourth rows show normal cases. All images are overlaid with paired CT images. In perfusion images, red indicates higher functional values, and blue indicates lower values. In Grad-CAM maps, red indicates higher model attention.

normal cases. The Grad-CAM attention maps provide visual explanations of the model's predictions by highlighting the regions of the input image that contribute most to the decision (Selvaraju *et al* 2020). Figure 5 displays the Grad-CAM activation maps from ResNET_3d, ResNET_CN MIL, and the corresponding SPECT perfusion images from the best test results of our experiments, distinguishing between 2 cases of CTEPH and 2 normal cases. For both CTEPH and normal cases, ResNET_3d focused on the lung parenchyma region. In contrast, ResNET_CN MIL's attention maps were closely aligned with the functional images shown in the perfusion scans. Specifically, ResNET_CN MIL concentrated on the low-function regions in CTEPH patients, which are indicative of the disease, thereby correlating better with the actual perfusion deficits seen in SPECT images. For normal cases, ResNET_CN MIL accurately highlighted the entire functional region of the lungs, reflecting a healthy perfusion pattern.

3.5. Hyperparameter configuration experiments

The experimental results for configurations of the hyperparameters L and M are presented in table 5. The results show that the configuration with $L = 500$ consistently outperformed those with $L = 1000$ in both AUC and sensitivity. Among the configurations with $L = 500$, the configuration with $M = 128$ achieved the highest AUC value of 0.833 and a sensitivity of 0.795. This suggests that this particular combination of hyperparameters provided the best balance between model accuracy and the ability to correctly identify positive cases.

3.6. Ablation study

We conducted an ablation study to evaluate the contribution of different components in our CN model for both normal and CTEPH classification tasks. The performance metrics for each configuration are summarized in table 6 and figure 6.

For the classification of normal cases, which is the first stage of our CN architecture used to screen out normal cases, we tested two versions of our Resnet-18 2D model: Resnet-I, which does not incorporate MIL,

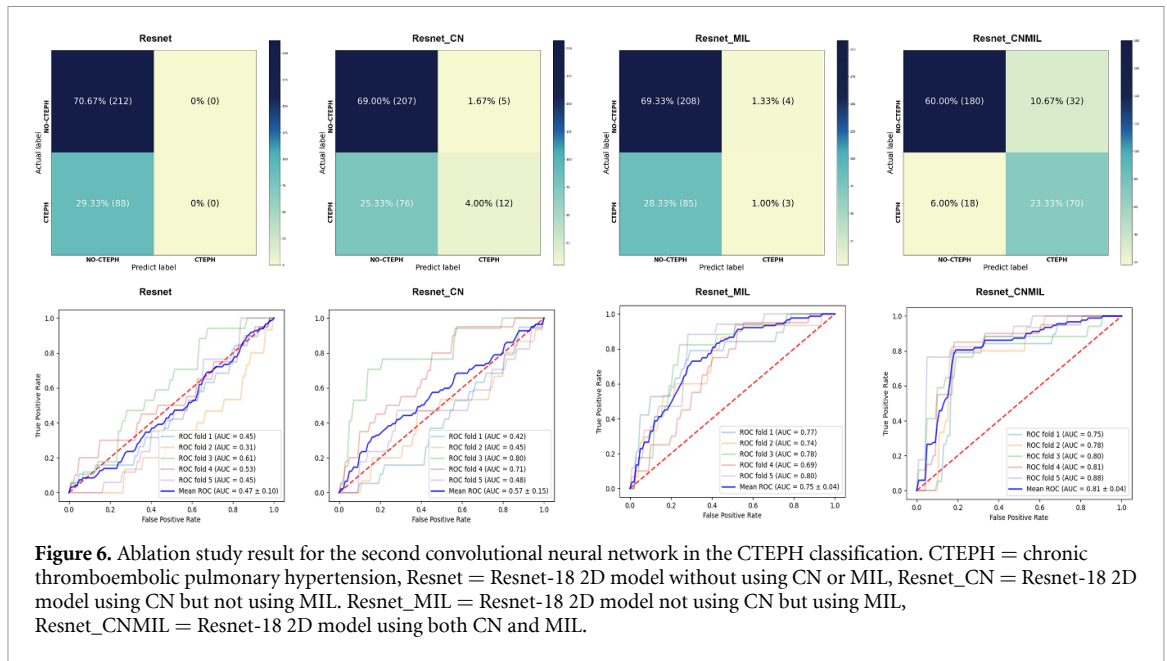


Table 5. Performance metrics for different values of L and M .

Value of L	Values of M	AUC	Sensitivity
1000	256	0.799	0.783
	128	0.783	0.784
	64	0.824	0.740
500	256	0.822	0.769
	128	0.833	0.795
	64	0.829	0.764

Table 6. The quantization results of the ablation study.

Module	AUC	Accuracy	Sensitivity	Specificity
Ablation study for normal classification				
Resnet-I	0.974	0.923	0.951	0.895
Resnet-I_MIL	0.978	0.943	1.000	0.895
Ablation study for CTEPH classification				
Resnet	0.469	0.707	0.000	1.000
Resnet_CN	0.571	0.730	0.136	0.976
Resnet_MIL	0.754	0.700	0.758	0.673
Resnet_CNMIL	0.807	0.833	0.795	0.849

CTEPH = chronic thromboembolic pulmonary hypertension, CN = cascade network, MIL = multiple instance learning, Resnet-I = Resnet-18 2D model used in the first stage without incorporating MIL, Resnet-I_MIL = Resnet-18 2D model used in the first stage incorporating MIL, Resnet = Resnet-18 2D model without using CN or MIL, Resnet_CN = Resnet-18 2D model using CN but not using MIL, Resnet_MIL = Resnet-18 2D model not using CN but using MIL, Resnet_CNMIL = Resnet-18 2D model using both CN and MIL.

and Resnet-I_MIL, which integrates MIL into the first stage of the model. The Resnet-I model achieved an AUC of 0.974, an accuracy of 0.923, a sensitivity of 0.951, and a specificity of 0.895. When MIL was incorporated into the model (Resnet-I_MIL), the performance metrics improved, particularly in sensitivity, which reached a perfect score of 1.000. The AUC slightly increased to 0.978, and the accuracy improved to 0.943, while the specificity remained consistent at 0.895.

For the classification of CTEPH cases, we tested several configurations: Resnet, Resnet_CN, Resnet_MIL, and Resnet_CNMIL. The plain Resnet model, which uses Resnet-18 2D without CN or MIL, achieved an

AUC of 0.469, an accuracy of 0.707, a sensitivity of 0.000, and a specificity of 1.000. Adding the CN component to create Resnet_CN improved the AUC to 0.571, accuracy to 0.730, sensitivity to 0.136, while maintaining a high specificity of 0.976. The Resnet_MIL model, incorporating MIL but not using CN, achieved an AUC of 0.754, an accuracy of 0.700, a sensitivity of 0.758, and a specificity of 0.673. The full model, Resnet_CN MIL, combining both CN and MIL components, yielded the highest performance with an AUC of 0.807, an accuracy of 0.833, a sensitivity of 0.795, and a specificity of 0.849.

4. Discussion

In this study, we explored the feasibility and effectiveness of using NCCT for the automatic diagnosis of CTEPH through a CNMIL. Our findings demonstrate the potential of this approach to enhance early diagnosis, eliminating the need for more invasive and costly diagnostic methods such as V/Q scans or CTA.

Our CNMIL model achieved a sensitivity of 0.795 and an AUC of 0.807, outperforming existing methods for automatic CTEPH identification. These results underscore the capability of our model to effectively distinguish between CTEPH and other lung diseases using NCCT. Sensitivity and AUC metrics notably declined when omitting annotations, highlighting the importance of incorporating MIL to improve model performance.

The implications of our findings are significant for clinical practice. By leveraging NCCT scans, which are routinely performed and widely available, our method offers a non-invasive alternative that can facilitate earlier and more accurate CTEPH diagnosis. This is particularly crucial given the often-nonspecific symptoms of early-stage CTEPH, which lead to frequent misdiagnoses and delays in appropriate treatment. Our approach could reduce the diagnostic timeline, enabling timely intervention and potentially improving patient outcomes.

Existing diagnostic methods for CTEPH, including CTA and V/Q scans, present several limitations such as higher costs, additional radiation exposure, and limited accessibility. Our CNMIL approach addresses these challenges by utilizing NCCT scans, which are safer and more readily available. Additionally, traditional CAD methods require extensive annotations and manual intervention, whereas our model operates efficiently with minimal case-level reports, thus reducing the burden on medical professionals.

Despite the promising results, our study has several limitations. The dataset size, while substantial, is limited to specific cohorts, which may affect the generalizability of our findings. Future work should aim to validate the CNMIL model across larger and more diverse datasets. Moreover, while our model demonstrates strong performance, further refinement of the network architecture and preprocessing steps could enhance its accuracy and robustness. Additionally, the integration of other imaging modalities, such as perfusion scans, could be explored to further improve diagnostic accuracy. Longitudinal studies assessing the impact of early CTEPH diagnosis using our method on patient outcomes would provide valuable insights into its clinical utility.

5. Conclusion

Our developed CNMIL network offers a groundbreaking solution for CTEPH diagnosis by leveraging the most commonly available CT images without the need for additional scans or annotations from professionals. With its enhanced accuracy and efficiency, this approach has the potential to revolutionize the timely and accurate detection of CTEPH, leading to improved patient outcomes and more effective disease management. By eliminating the reliance on costly and invasive procedures, our model provides a cost-effective and patient-friendly alternative for CTEPH detection. Future studies should further validate and explore the full potential of this innovative approach in larger cohorts to solidify its clinical applicability and impact.

Data availability statement

The datasets generated or analyzed during the study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Acknowledgments

The authors thank the Department of Health Technology and Informatics colleagues for statistical assistance and the School of Nursing for image processing. The authors also thank The First Affiliated Hospital of Nanchang University, the Hong Kong Polytechnic University and Queen Mary Hospital for their generous support.

Funding

This research was partly supported by General Research Fund (15103520) of the University Research Committee, Health and Medical Research Fund (07183266, 09200576) of the Health Bureau, and PolyU (UGC) RI-IWEAR Seed Project (P0044802), The Government of the Hong Kong Special Administrative Region.

Ethical approval

This study was approved by the Institutional Review Board (Approval Number: UW 20-105). The board waived the requirement for informed consent from the participants as the study involved minimal risk to subjects and met the criteria for waiver of consent.

Conflict of interest

The authors have no potential conflicts of interest to disclose.

ORCID iDs

Mayang Zhao  <https://orcid.org/0009-0009-2069-6192>

Ren Ge  <https://orcid.org/0000-0003-2049-2682>

References

- Avesta A, Hossain S, Lin M, Aboian M, Krumholz H M and Aneja S 2023 Comparing 3D, 2.5 D, and 2D approaches to brain image auto-segmentation *Bioengineering* **10** 181
- Cerny V, Pagac J, Novak M and Jansa P 2023 Semi-automatic quantification of mosaic perfusion of lung parenchyma and its correlation with haemodynamic parameters in patients with chronic thromboembolic pulmonary hypertension *Clin. Radiol.* **78** e918–e924
- Chen S, Ma K and Zheng Y 2019 Med3d: transfer learning for 3D medical image analysis (arXiv:1904.00625)
- Chlap P, Min H, Vandenberg N, Dowling J, Holloway L and Haworth A 2021 A review of medical image data augmentation techniques for deep learning applications *J. Med. Imaging Radiat. Oncol.* **65** 545–63
- D'Armini A M 2015 Diagnostic advances and opportunities in chronic thromboembolic pulmonary hypertension *Eur. Respir. Rev.* **24** 253–62
- Deano R C, Glassner-Kolmin C, Rubenfire M, Frost A, Visovatti S, McLaughlin V V and Gomberg-Maitland M 2013 Referral of patients with pulmonary hypertension diagnoses to tertiary pulmonary hypertension centers: the multicenter RePHerral study *JAMA Intern. Med.* **173** 887–93
- Delcroix M et al 2016 Long-term outcome of patients with chronic thromboembolic pulmonary hypertension: results from an international prospective registry *Circulation* **133** 859–71
- Diogo P, Morais M, Calisto F M, Santiago C, Aleluia C and Nascimento J C 2023 Weakly-supervised diagnosis and detection of breast cancer using deep multiple instance learning 2023 *IEEE 20th Int. Symp. on Biomedical Imaging (ISBI)*
- Fedullo P, Kerr K M, Kim N H and Auger W R 2011 Chronic thromboembolic pulmonary hypertension *Am. J. Respir. Crit. Care Med.* **183** 1605–13
- Galiè N et al 2016 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS); endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT) *Eur. Heart J.* **37** 67–119
- Gopalan D, Delcroix M and Held M 2017 Diagnosis of chronic thromboembolic pulmonary hypertension *Eur. Respir. Rev.* **26** 160108
- He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*
- Jimenez-del-Toro O, Cid Y D, Platon A, Hachulla A-L, Lador F, Poletti P-A and Müller H 2020 A lung graph model for the radiological assessment of chronic thromboembolic pulmonary hypertension in CT *Comput. Biol. Med.* **125** 103962
- Kim N H, Delcroix M, Jais X, Madani M M, Matsubara H, Mayer E, Ogo T, Tapson V F, Ghofrani H-A and Jenkins D P 2019 Chronic thromboembolic pulmonary hypertension *Eur. Respir. J.* **53** 1801915
- Kulkarni A, Carrion-Martinez I, Dhindsa K, Alaref A A, Rozenberg R and van der Pol C B 2021 Pancreas adenocarcinoma CT texture analysis: comparison of 3D and 2D tumor segmentation techniques *Abdom. Radiol.* **46** 1027–33
- Lang I 2015 Chronic thromboembolic pulmonary hypertension: a distinct disease entity *Eur. Respir. Rev.* **24** 246–52
- LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
- Lee J-G, Jun S, Cho Y-W, Lee H, Kim G B, Seo J B and Kim N 2017 Deep learning in medical imaging: general overview *Korean J. Radiol.* **18** 570–84
- Li F, Chen H, Liu Z, Zhang X and Wu Z 2019 Fully automated detection of retinal disorders by image-based deep learning *Graefes Arch. Clin. Exp. Ophthalmol.* **257** 495–505
- Lu Y, Lorenzoni A, Fox J J, Rademaker J, Vander Els N, Grewal R K, Strauss H W and Schöder H 2014 Noncontrast perfusion single-photon emission CT/CT scanning: a new test for the expedited, high-accuracy diagnosis of acute pulmonary embolism *Chest* **145** 1079–88
- Nam S, Chong Y, Jung C K, Kwak T-Y, Lee J Y, Park J, Rho M J and Go H 2020 Introduction to digital pathology and computer-aided pathology *J. Pathol. Transl. Med.* **54** 125–34
- Narechania S, Renapurkar R and Heresi G A 2020 Mimickers of chronic thromboembolic pulmonary hypertension on imaging tests: a review *Pulm. Circ.* **10** 2045894019882620

- Pepke-Zaba J et al 2011 Chronic thromboembolic pulmonary hypertension (CTEPH) results from an international prospective registry *Circulation* **124** 1973–81
- Price L C, Wort S J, Finney S J, Marino P S and Brett S J 2010 Pulmonary vascular and right ventricular dysfunction in adult critical care: current and emerging options for management: a systematic literature review *Crit. Care* **14** 1–22
- Raghu M, Zhang C, Kleinberg J and Bengio S 2019 Transfusion: understanding transfer learning for medical imaging *Advances in Neural Information Processing Systems* vol 32
- Ren G, Lam S-K, Zhang J, Xiao H, Cheung A L-Y, Ho W-Y, Qin J and Cai J 2021 Investigation of a novel deep learning-based computed tomography perfusion mapping framework for functional lung avoidance radiotherapy *Front. Oncol.* **11** 644703
- Ribeiro A H, Ribeiro M H, Paixão G M, Oliveira D M, Gomes P R, Canazart J A, Ferreira M P, Andersson C R, Macfarlane P W and Meira W Jr 2020 Automatic diagnosis of the 12-lead ECG using a deep neural network *Nat. Commun.* **11** 1760
- Sarker I H 2021 Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions *SN Comput. Sci.* **2** 420
- Seiffert A P, Gómez-Grande A, Pilkington P, Cara P, Bueno H, Estenez J, Gómez E J and Sánchez-González P 2020 Automatic diagnosis of chronic thromboembolic pulmonary hypertension based on volumetric data from SPECT ventilation and perfusion images *Appl. Sci.* **10** 5360
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D 2020 Grad-CAM: visual explanations from deep networks via gradient-based localization *Int. J. Comput. Vis.* **128** 336–59
- Shahin Y, Alabed S, Alkhanfar D, Tschirren J, Rothman A M, Condliffe R, Wild J M, Kiely D G and Swift A J 2022 Quantitative CT evaluation of small pulmonary vessels has functional and prognostic value in pulmonary hypertension *Radiology* **305** 431–40
- Sostman H D, Miniati M, Gottschalk A, Matta F, Stein P D and Pistolesi M 2008 Sensitivity and specificity of perfusion scintigraphy combined with chest radiography for acute pulmonary embolism in PIOPED II *J. Nucl. Med.* **49** 1741–8
- Wang S et al 2020 A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis *Eur. Respir. J.* **56** 2000775
- Wang X, Su R, Xie W, Wang W, Xu Y, Mann R, Han J and Tan T 2023 2.75 D: boosting learning by representing 3D medical imaging to 2D features for small data *Biomed. Signal Process. Control* **84** 104858
- Zhai Z, Boon G J, Staring M, van Dam L F, Kroft L J, Hernández Girón I, Ninaber M K, Bogaard H J, Meijboom L J and Vonk Noordegraaf A 2023 Automated quantification of the pulmonary vasculature in pulmonary embolism and chronic thromboembolic pulmonary hypertension *Pulm. Circ.* **13** e12223
- Zhou L, Xiao Q, Taha M F, Xu C and Zhang C 2023 Phenotypic analysis of diseased plant leaves using supervised and weakly supervised deep learning *Plant Phenomics* **5** 0022
- Zhou Z, Sodha V, Rahman Siddiquee M M, Feng R, Tajbakhsh N, Gotway M B and Liang J 2019 Models genesis: generic autodidactic models for 3D medical image analysis *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd Int. Conf., Proc., Part IV 22 (Shenzhen, China, 13–17 October 2019)*