Multi-CBCS: Multimodal Contrast Bolus Consistency Network for Pulmonary Embolism Detection by Integrating CTPA and Lung Perfusion Imaging

P. Poonkuzhali[1,*], V. Nivedita[2], J. Rajalakshmi[3], Ahmed Farag Salem Babetat[4]

[1] *Associate professor, Department of Electronics and Communication Engineering, R.M.D. Engineering College, Tamil Nadu, India*

[2] *Assistant Professor, Department of Computer Science Engineering, SRMIST Ramapuram Campus, Tamil Nadu, India*

[3] *Associate Professor, Department of Biomedical Engineering, Velalar College of Engineering & Technology, Tamil Nadu, India*

[4] *Dr. Mohammad Alfagih Hospital, Riyadh, Saudi Arabia*

(**Received** 14 February 2026, **Revised** 07 March 2026, 03 April 2026, **Accepted** 01 May 2026)

*Corresponding author: ppoonkuzhali25@gmail.com

DOI: 10.5875/7v5fw044

Abstract: Pulmonary embolism (PE) is a dangerous heart disorder that should be effectively diagnosed as soon as possible. The clinical standard of detecting emboli based on contrast filling defects is computed tomography pulmonary angiography (CTPA), and lung perfusion scintigraphy (V/Q scan) is a complementary test of the pulmonary blood flow. Nevertheless, the current deep learning methods are mainly based on one-modality image analysis, and they do not explicitly simulate the dynamics of contrast bolus or combine the functional perfusion information, which restricts the diagnostic strength and clinical readability. Furthermore, the Contrast Bolus Consistency Module (CBCM) increases interpretability by accounting for spatial consistency of contrast flow through pulmonary arteries, allowing the network to emphasize meaningful interruptions of this process caused by embolic blockages, which regular single modality CNNs cannot due to implicit feature extraction. Addressing these drawbacks, a Contrast Bolus Consistency Network is suggested to detect multimodal pulmonary embolism based on the images of CTPA and lung perfusion scintigraphy. It is based on a convolutional neural network (CNN)-based backbone that has been developed on the basis of which hierarchical feature extraction is enabled. The presented framework is able to integrate physiologically coherent multimodal features by uniting abnormalities of anatomy contrast with functional deficits of perfusion by using cross-modal attentional processes. A specific contrast bolus modeling module evaluates consistency of spatial attenuation along pulmonary arteries to detect contrast discontinuity that is a sign of embolic obstruction. The results of the experiment prove that the combination of multiple modalities CBC-Net is much better, 99.1% accuracy, 98.4% precision, 98.4% recall and 95.7% F1-score according to the experimental outcomes are significantly higher when compared to existing single-modal methods. These results indicate that there is clinical potential in combining structural and functional imaging modalities as a viable method of detecting pulmonary embolism.

Keywords: Pulmonary embolism, scintigraphy images, blood vessels, arteries, Wiener filter, Convolution Neural Network (CNN)

1. Introduction

Pulmonary embolism (PE) is a serious condition that can cause blockage in the central, lobar, segmental, or subsegmental pulmonary arteries, which carry blood from the heart to the lungs. Smart health care systems using wireless communication technologies allow for quick access to information in the form of images, which is very important for the timely diagnosis of pulmonary embolism [1]. Early diagnosis leads to early administration of anticoagulant treatment, which considerably decreases deaths and prevents life-threatening complications caused by pulmonary embolism. Pulmonary embolism is associated with high morbidity and can lead to life-threatening complications such as right ventricular failure and sudden cardiac death if not diagnosed in time. Accurate and early detection is therefore essential in

clinical practice. However, manual interpretation of CTPA scans is challenging and time-consuming due to the large number of image slices and subtle appearance of emboli. This increases the risk of diagnostic errors and inter-observer variability. Inter-observer variability refers to differences in interpretations among radiologists, which may lead to inconsistent diagnoses. Such variability highlights the need for automated pulmonary embolism detection systems that provide standardized and reproducible assessments, thereby improving diagnostic reliability. Hence, the development of automated detection systems using deep learning techniques is crucial to support radiologists in improving diagnostic accuracy and ensuring timely treatment. In this context, integrating multi-modal imaging data with advanced deep learning techniques has emerged as a promising direction for improving diagnostic performance. Integration of multiple modalities leads to a better diagnosis, as it combines anatomical data from the CTPA scan with functional perfusion information, allowing for better identification of small emboli that cannot be detected using other methods. When anatomical data provided in CTPA is supplemented with the functional data of perfusion imaging, one can have a more detailed picture of the pulmonary abnormality. Furthermore, modeling contrast bolus dynamics within such multimodal frameworks can enhance the detection of embolic patterns and improve clinical interpretability. Smart healthcare systems powered by wireless technologies enable rapid acquisition and transmission of medical imaging data, supporting timely diagnosis of pulmonary embolism. These advancements improve accessibility, reduce diagnostic delays, and enhance clinical decision-making in distributed healthcare environments. Today's major method for diagnosing PE is computed tomography pulmonary angiography (CTPA), which shows PE as a filling defect (e.g., a dark region around the affected arterial lumen) [2]. Interpreting a large number of CTPA images, which typically contain 300–500 slices per scan, is time-consuming [3]. More accurate PE detection is being studied to lower patient mortality, medical expenditures, and therapy. Recent advances in pattern recognition by computer vision and deep learning algorithms have transformed medical imaging analysis [4]. Automatic PE identification utilizing deep neural network models and visual morphology has proved effective. Their following measurements may aid therapy and planning [5]. Artificial intelligence has significantly improved medical imaging, where large-scale data contributes to enhanced model performance. AI-based diagnostics will improve treatment at reduced cost [6]. Early PE detection is possible using DL techniques. Automatic feature extraction is possible using DL [7]. Computer vision engineers utilize convolutional neural networks to understand images. CNNs process images. Each CNN layer has several filters to identify recurrent patterns. CNN categorization involves tagging the whole image. Deep learning models can provide accurate results for medical image processing. Unfortunately, such models usually have low transparency. Additionally, deep learning models need to be trained on large amounts of labeled data and high computational power. Overcoming these challenges demands ongoing innovation in architectural design, data augmentation approaches, and explainable AI frameworks. Given this, this work's contributions are as follows:

- Hierarchical anatomical representation learning strategy is introduced in encoder layer, which exploits multi-scale feature maps extracted from the backbone network to effectively capture pulmonary vascular structures across different resolution
- Contrast Bolus Consistency Module (CBCM) is integrated between the encoder and decoder blocks to explicitly model contrast bolus dynamics within the pulmonary arteries, wherein contrast-aware cross-modal fusion strategy aligns contrast-enhanced anatomical features from CTPA with functional perfusion patterns from lung scintigraphy.

The paper's structure follows. Section II includes material directly linked to this study. Section III evaluates the system architecture. Section IV presents outcomes from the scenario evaluation. Finally, Section V summarizes the study's findings.

2. Related works

Recent studies highlight that wireless-enabled smart healthcare frameworks facilitate real-time data sharing and integration with AI models, improving diagnostic efficiency. However, limited work has explored the combination of wireless systems with multimodal imaging for pulmonary embolism detection. Neural networks' capacity to automatically learn discriminative patterns from complicated medical data has made PE detection with them popular. Wireless technology enhances real-time data transfer from diagnostic devices to cloud platforms, enabling faster decision-making and reducing patient wait times. This section reviews neural network-based and deep learning survey research on pulmonary embolism detection [8]. uses a bidirectional generative network with Transformer attention mechanisms to transform NCCT images to CTA images and integrates a multi-scale object detection module optimized by dynamic Efficient Multi-scale Attention (EMA) attention. Next-generation healthcare framework integrating lightweight CNNs, capsule networks, and blockchain alternatives for real-time pandemic detection

and data security. The approach enables efficient disease diagnosis through optimized deep learning architectures while ensuring secure patient data management through decentralized blockchain mechanisms [9]. The work in [10] proposes Deep Neural Network (DNN)-based prototype method, together with Shapiro-Wilk test to check the continuous variables for normality. In [11], ResNet50, DenseNet121, and Swin Transformer models were combined and tested on the RSNA-STR PE dataset. In [12], improved Mask R-CNN shows how the loss function affects model performance, allowing for model improvement. In [13] created an attention-guided CNN that incorporates local context. It considers global and local lesion areas like a human expert before reaching a judgment. In [14], two-stage deep learning framework using CNN and U-Net is proposed for pulmonary embolism detection from CTPA images, achieving 99% accuracy. The model shows improved sensitivity and outperforms existing methods for early diagnosis. In [15], PE-Deep Net, a hybrid deep learning convolutional neural network, was used to detect and predict pulmonary embolism risk in patients. In [16], two-phase multitask deep learning method for pulmonary embolism (PE) detection using CTPA images, capable of identifying embolus position (acute/chronic) and right-to-left ventricle ratio (RV/LV). The model achieves an sensitivity of 0.86 and specificity of 0.85, while providing interpretability through attention heatmaps and Grad-CAM visualizations. A cloud-based CNN framework for cystic fibrosis diagnosis is proposed by Poovendran Alagarsundaram et al., [17] demonstrating high accuracy through hierarchical feature extraction and scalable deployment. Their work highlights the effectiveness of integrating deep learning with cloud infrastructure for reliable medical image analysis. Inspired by this, the present study extends deep feature learning to a multimodal setting by combining CTPA and perfusion imaging. Unlike the single-modality approach, the proposed method incorporates contrast bolus consistency modeling to enhance diagnostic accuracy and interpretability in pulmonary embolism detection. Kanwal et al. (2022) investigated the impact of color processing on detecting blood and damaged tissue regions in whole slide images (WSIs) for computational pathology applications. Their study highlights that artifacts such as blood and tissue damage can significantly affect diagnostic accuracy and model performance. By comparing transfer learning and training-from-scratch approaches, the work achieved high F1-scores and demonstrated the importance of both color and morphological features in classification. This study emphasizes the need for effective preprocessing and artifact removal to improve reliability in automated medical image analysis systems.

The latest neural network experiments on pulmonary embolism detection show excellent results, however there are disadvantages. Trans modal and attention-based models like PulmoNet and Abn-BLIP need sophisticated models and significant training data, increasing computing complexity. Deep learning ensemble and hybrid models are accurate but typically have model redundancy and interpretability issues, making them challenging to use in medicine. Region-based methods such Mask R-CNN and Pipelines based on segmentation, including U-Net with classical machine learning, pass errors at multiple levels, as in Table 1.

Table 1: Summary of existing methods.

Author/year	**Method**	**Results AG-CNN**	**Advantages**
Wu et al., [8]	PulmoNet	SSIM= 0.906 Accuracy= 93.9 %	Avoids contrast-enhanced CTA by properly synthesizing NCCT CTA; dynamic EMA increases multi-scale embolus sensitivity, especially for tiny and peripheral clots.
Zhong et al., [28]	Abn-BLIP	Accuracy= 896%	Improves PE-related abnormality identification using language image alignment, enhancing generalization with limited annotations and semantic consistency across modalities.
Zsarnoczay et al., [10]	DNN	Accuracy= 93.8% Sensitivity=84.6%	Integrates deep learning and statistical

			normality validation for stable feature distributions.
Abdelhamid et al., [11]		Accuracy=97.80%	CNN-based local texture characteristics and Transformer-based global context improve robustness.
Doğan et al., [12]	Mask R-CNN	Accuracy=95%	Optimizes loss function for PE localization, boosting sensitivity for tiny emboli while retaining region-level accuracy.
Bushra et al., [13]	AG-CNN	sensitivity = 86.2%, specificity = 87.95, F1-score = 80%	Models radiologist-like attention for focused lesion analysis and greatly reduces false-positives.
Olescki et al., [7]	ML	dice score = 0.81	Separates embolus candidate segmentation from classification, reducing false positives and keeping structural boundary accuracy.
Lynch et al., [15]	PE-DeepNet	accuracy = 94.2%	Continuous PE detection and risk-level prediction provide early prognosis rather than binary diagnosis.

3. Proposed methodology

The input modalities in the proposed framework of pulmonary embolism (PE) classification are the images of contrast-enhanced CT pulmonary angiography (CTPA) and lung perfusion scintigraphy. The technology can be combined with wireless medical devices to facilitate the workflow of healthcare professionals by easily collecting data provided by the remote patient monitoring systems. The proposed framework can be integrated with wireless medical systems to enable seamless acquisition and transfer of CTPA and perfusion imaging data. This supports efficient preprocessing, multimodal fusion, and real-time deployment in smart healthcare environments. Preprocessing stage is used to improve the quality of images and reduce noise caused by the acquisition process, through a combination of BM3D denoising and Wiener filtering [27]. The images are then subjected to a Contrast Bolus Consistency Network, that aims at ensuring consistency in contrast between vessels in the lungs. In this network, hierarchical learning is obtained with the help of Hierarchical Encoder decoder (HED) architecture that learns the local and global contextual information. Simultaneously, the deep spatial features are computed with a VGG16-based CNN, and a cross-modal feature fusion is used to combine complementary information of various sources of imaging. The fused feature representations enable robust discrimination of embolic patterns, allowing the system to accurately classify cases into saddle pulmonary embolism, acute pulmonary embolism, and chronic pulmonary embolism as shown in figure 1.

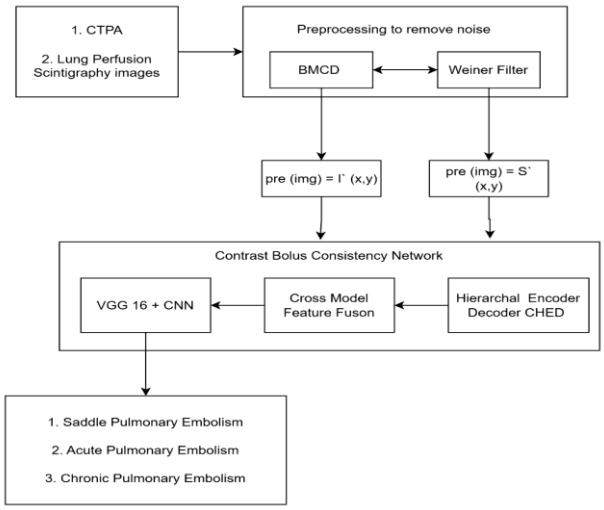

Figure 1: Block diagram representation of pulmonary embolism (PE) classification with an enhanced feature extractor.

The architecture shown in Figure 1 consists of multiple convolutional layers with varying channel depths, where feature maps are progressively increased from low-level representations to high-level semantic features. The number of filters is adapted across layers to ensure efficient feature extraction and representation learning.

3.1 Preprocessing of images

First, LPF images must be preprocessed using lung prefusion scintigraphy images.

$$I(x,y) = S(x,y) + n(x,y) \quad (1)$$

where, $S(x,y)$ is the real perfusion signal while $n(x,y)$ is radioactive decay-dominated Poisson noise. Block-Matching and 3D Filtering (BM3D) suppresses noise while maintaining clinically important perfusion abnormalities. In hybrid adaptive preprocessing algorithm combining Rudin_Oher_Fatemi (ROF) model for edge detection, Richardson-Lucy (RL) algorithm for enhancement, and block matching 3D (BM3D) collaborative filtering for denoising medical images. The method achieves high peak signal-to-noise ratios of 47.44 dB for chest X-ray and 46.87 dB for MRI/CT datasets, outperforming existing denoising approaches while preserving edge information [18]. Initially, a reference patch pat_r of size $k \times k$ is used to find comparable patches pat_i inside a preset search window by reducing Euclidean distance:

$$dis(pat_r, pat_i) = \|pat_r - pat_i\|_2^2 \quad (2)$$

The collected patches are cooperatively filtered by applying a 3D transform T_{3D} followed by coefficient shrinkage:

$$\hat{\mathcal{G}} = \mathrm{T}_{3\mathrm{D}}^{-1}\big(\mathcal{H}(\mathrm{T}_{3\mathrm{D}}(\mathcal{G}))\big) \quad (3)$$

where $\mathcal{H}(\cdot)$ represents hard-thresholding determined by

$$\mathcal{H}(\mathrm{c}) = \begin{cases} \mathrm{c}, & |\,\mathrm{c}\,| \geq \lambda \\ 0, & |\,\mathrm{c}\,| < \lambda \end{cases} \quad (4)$$

The second step uses Wiener filtering on revised patch groups to improve signal estimate:

$$\hat{\mathrm{S}} = \frac{\sigma_{\mathrm{S}}^2}{\sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{n}}^2} \cdot \mathrm{I} \quad (5)$$

where σ_{S}^2 and σ_{n}^2 are estimated signal and noise variances.

The Weighted aggregation of overlapping patches reconstructs the denoised image.

$$\hat{\mathrm{I}}(\mathrm{x},\mathrm{y}) = \frac{\sum_{\mathrm{i}} \mathrm{w}_{\mathrm{i}}\, \hat{\mathrm{P}}_{\mathrm{i}}(\mathrm{x},\mathrm{y})}{\sum_{\mathrm{i}} \mathrm{w}_{\mathrm{i}}} \quad (6)$$

Finally, the patient's denoised Lung Perfusion Scintigraphy is represented by $\mathrm{pre(img)} = \hat{\mathrm{I}}(\mathrm{x},\mathrm{y})$. Wiener filtering is chosen owing to its ability to adaptively filter out noise and ensure that minimum mean square error is obtained without losing any vital features of the signal. Additionally, CTPA images are preprocessed using a Wiener filter to produce the filtered image.

$$\hat{\mathrm{S}}(\mathrm{u},\mathrm{v}) = \mathrm{H}(\mathrm{u},\mathrm{v}) \cdot \mathrm{I}(\mathrm{u},\mathrm{v}) \quad (7)$$

and returned to spatial domain using inverse Fourier transform. Calculate filtered pixel intensity as:

$$\hat{\mathrm{S}}(\mathrm{x},\mathrm{y}) = \mu_{\mathrm{L}} + \frac{\sigma_{\mathrm{L}}^2 - \sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{L}}^2}(\mathrm{I}(\mathrm{x},\mathrm{y}) - \mu_{\mathrm{L}}) \quad (8)$$

where μ_{L} and σ_{L}^2 indicate local mean and variance inside a sliding window, whereas σ_{n}^2 represents estimated noise variance. In conclusion, the denoised CTPA of a patient is represented as $\mathrm{pre(img)} = \hat{S}(x,y)$.

The BM3D denoising filter is employed because of its high efficiency in maintaining the structure while removing noise that is difficult to detect in medical images; the Wiener filter is highly efficient in adapting itself to the signal variation locally. As opposed to Gaussian filtering that tends to smooth off the edges, this combination is more efficient.

3.2 Contrast Bolus Consistency Network:

In the proposed Contrast Bolus Consistency Network (CBC-Net), hierarchical encoder decoder, cross-modal feature fusion, and custom CNN for classification are the three phases of multimodal learning.

3.2.1 Hierarchical encoder decoder (HED):

Figure 2 shows a top-down horizontal connection system with encoder and decoder.

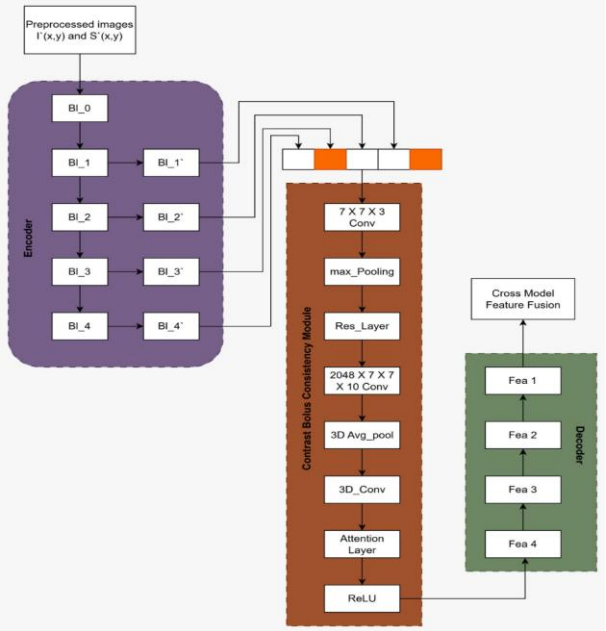

Figure 2: Architecture of the hierarchical encoder–decoder network with the contrast bolus consistency module.

The residual connections in the HED structure serve as shortcut paths that transmit high-resolution information between earlier and subsequent layers directly. Thus, fine vascular details are not lost in the process of upsampling. Also, these shortcut connections make it easier for the neural network to be trained. The encoder, which is the anatomical representation learning module, takes the original preprocessed image as input and uses the backbone network to obtain five scale feature maps Bl_i' (i = 1, 2, 3, 4) of different feature information and scales. The encoder module receives the decoder's output feature map. The encoder converts Bl_4' into Fea_4, upsamples it to the same size as Bl_3', adds it to Bl_3' to get $Feature\_3$, and continues this procedure to generate Fea_2 and Fea_1. As illustrated in Figure 2, the hierarchical encoder–decoder structure utilizes multiple convolutional blocks with increasing channel dimensions in the encoder and corresponding decreasing dimensions in the decoder. Each stage processes feature maps with varying filter sizes to capture multi-scale contextual information. The encoder employs deeper channel representations for semantic abstraction, while the decoder reconstructs spatial details using progressively reduced channel sizes.

To further enhance feature preservation during the decoding phase, residual connections are incorporated between corresponding encoder and decoder layers within the hierarchical encoder-decoder (HED) module. These skip connections enable the direct transfer of high-resolution spatial features from the encoder to the decoder, thereby mitigating information loss during up sampling operations. Also, the residual pathways guarantee gradual gradient flow and assist in maintaining fine-grained structural information of pulmonary vessels, which play a pivotal role in proper embolism identification. This is a good design because it offers feature consistency between different scales and enhances performance of the model in general. The last feature map is obtained by up sampling Fea_i (i = 2, 3, 4) to the same size as Fea_1 and overlaying it, after which it is weighted by the SE. In order to give a vivid picture of the flow of contrast bolus in the pulmonary arteries, The Contrast Bolus Consistency Module between the encoder block and the decoder block. After a 7 × 7 × 3 convolutional layers, the model uses max pooling and retains ResNet152 residual connections for hierarchical feature extraction. The final residual block generates a 2048 × 7 × 7 × 10 feature map, which undergoes 3D adaptive average pooling and 3D convolutional layer to provide logits for 32 abnormality classes. The model is optimized using binary cross-entropy loss and probabilities from sigmoid activation functions:

$$loss = -\sum_{k=1}^{32} u_k log J_k + (1-u_k)\log(1-J_k) \qquad (9)$$

where the ground-truth label of U is u_k. Multi-scale features $fea_l^l \in R^{ch_l \times hei_l \times wei_l}, l \in Bl\_i$ (i = 0, 1, 2, 3, 4) are extracted from five blocks to improve visual abnormality querying. Feature aggregation across multiple-scales helps the network understand global context at coarser levels and retain fine details at finer levels, which increases the precision of localization for areas with emboli. The 3D patch-pooling module divides each slice into discrete 7x7x10 sub-volumes. After that, average pooling reduces spatial resolution while keeping confined spatial and semantic information in each patch. 3D convolutional blocks with ReLU activation and batch normalization are used to include features from each scale, lowering their dimensionality to $ch_l' \times d_{hel\_vess}$, where d_{hel_vess} denotes the healthy vessel dimension. Lastly, an M-channel feature representation is obtained by concatenating multi-scale embeddings.

While U-Net adopts symmetric encoder-decoder architecture with skip connection, the novel hierarchical encoder-decoder network employs superior residual connections along with multi-scale feature refinement to facilitate better feature alignment and accurate vascular structure representation.

3.2.2 Cross modal feature fusion:

The Cross-Modal Attention-Based Feature Fusion Module aligns and combines CTPA contrast-enhanced anatomical information with lung perfusion scintigraphy functional perfusion characteristics. It further train two fusion modules and a combiner module in stage two to integrate CTPA and scintigraphy image features. The image-guided global fusion (IGGF) module learns word weights in altered text to maintain visual elements. The image

representation θI_Q represented query Q, whereas the extracted features represented key K and value V. Cross-attention and feedforward layers compute image-guided feature [19].

$$D(I_q) = \theta I_Q + Multi\_head(Q, K, V) \quad (10)$$

where, $Multi\_head$ indicates cross-attention between heads. The attention module is followed by a gating module that receives the fused feature $fuse\_fea_t$ from the current fusion unit and the concealed feature $gate_{t-1}$ from the previous gating unit. The selective filtering of features in the gating process involves assigning greater weights to important regions while reducing noise, thus improving the feature representation process. The gating unit integrates input features and uses a convolutional layer with a sigmoid function to create a learnable referent-awareweight matrix for the fused feature $fuse\_fea_t$. The aware matrix gives semantically significant spatial areas greater weights and noisy features lower weights. The gating unit recognizes the difference between the feature $fuse\_fea_t$ and the weighted fea_t, aggregating the difference region via concatenation, convolution, and non-linear activation. Finally, the feature gen_{t-1} will be utilized as input for the next optimization phase. The adaptive selection gate gating technique for hidden feature $gate_t$ is as follows.

$$u_t = \sigma(W_z(F_t + gen_{t-1}) + bias_z \quad (11)$$

$$v_t = F_t . u_t \quad (12)$$

$$gate_t = \tanh(W_g(|F_t; v_t| + bias_g \quad (13)$$

where the 3 × 3 convolution operations' learnable parameters are $W_z \in R^{ch'_v \times ch'_v}$ and. $W_g \in R^{(ch'_v \times ch'_v) \times (ch'_v)}$. $bias_z$ and $bias_g$ are bias. The Final integration yields $fea^{fuse} = \{fea_1, fea_2, \ldots . fea_L\}$ from L output.

For instance, a perfusion deficit area on scintigraphy imaging can match an abnormality seen in contrast material filling on CTPA, and the fusion software matches them to enhance detection and localization of the emboli.

3.2.3 Classification Custom CNN:

The cross-modal attention fusion leads to the combination of features of the CTPA image and the lung perfusion scintigraphy image. This fused feature is then used to classify the data via VGG-16, VGG-19, and a customized CNN.

The VGG-16 model includes 16 layers that are popularly used for feature extraction. However, the VGG-19 model uses an increased depth to include 19 layers in its design. Both of these models make use of smaller sized convolution kernels at a constant stride value, i.e., 3×3 followed by max-pooling of size 2×2.

The first layer of the VGG-19 structure accepts an input of 224 × 224 dimensions and utilizes four convolutional layers, where each layer has 64 filters. In the second layer, features that are of the dimensions 112 × 112 are fed into two convolutional layers with 128 filters each. In the third layer, there are three convolutional layers with 256 filters applied to features of 56 × 56 dimension. In the fourth layer, 28 × 28-dimension features undergo three convolutional layers with 512 filters.

This is possible since the ReLU activation function gives a zero when the input is negative and retains the positive feature, thus allowing the aspect of nonlinear learning of features.

ReLU, as well as Leaky ReLU activation functions, are applied consistently throughout the network so as to have stable gradient flow and better nonlinear representation. Formula for activation functions in Equations (14)–(15):

$$ReLU\ (x) = \begin{cases} 1, y \geq 0 \\ 0, y < 0 \end{cases} \quad (14)$$

$$LeakyReLU\ (x) = \begin{cases} y, y \geq 0 \\ scale \times y, y < 0 \end{cases} \quad (15)$$

Multiclass are classified by the activation function. The predictive probability gives an array of values ranging between 0 to 1 after computing the probability distribution of the FC layer output; the overall probability of all the predicted classes is 1 as expressed in eqn (16).

$$out_i = \sum_j Wei_{i,j} NV_j + bia_i \quad (16)$$

where Wei is the weight, NV is the neuron value, and bia is the bias, and z_i is the summation of all weights.

$$y = softmax - f(out_i) = \frac{\exp(out_i)}{\sum_j \exp(out_i)} \quad (17)$$

where out_i represents the anticipated probability distribution for Saddle, Acute, and Chronic Pulmonary Embolism. In [20], modified VGG-16 CNN architecture for pneumonia classification using chest X-ray images, specifically addressing imbalanced dataset challenges through data augmentation techniques. The augmented model achieves 92% accuracy, demonstrating a 15% improvement over non-augmented scenarios, with potential integration into Android-based machine learning systems for clinical diagnosis assistance.

4. Performance analysis

4.1 Dataset description

In Lung Perfusion Scintigraphy-LPF (V/Q scan [21]), ventilation data were standardized to corrected perfusion counts. After distributing the regularized ventilation amount by the perfusion amount for every voxel, a V̇/Q̇ dataset was created (Figure 3). The ventilation–perfusion (V/Q) ratio is useful for identifying mismatched regions where normal ventilation is accompanied by reduced or absent perfusion. Such ventilation–perfusion mismatches are characteristic indicators of pulmonary embolism, as embolic obstruction restricts blood flow while air

distribution remains relatively unaffected. Therefore, analyzing V/Q ratios enables more precise localization of perfusion abnormalities associated with pulmonary embolism. The $\log_{10}$-transformed V̇/Q̇ data were divided into 101 identical intervals, ranging from -3 ($log_{10}0.001$) to 2 ($log_{10}100$) with a 0.05 width. Values were selected to match MIGET (30). Voxels with values greater than 100 were allocated 100, while those below 0.001 were assigned 0.001. The voxel-level normalization standardizes intensity distributions across all voxels prior to feature extraction, ensuring consistent feature representation and improving interpretability of perfusion patterns. Together with the corresponding distributions of total ventilation and total perfusion to each $\dot{V}/\dot{Q}$ period, the frequency distributions of the $\dot{V}/\dot{Q}$ dataset was generated.

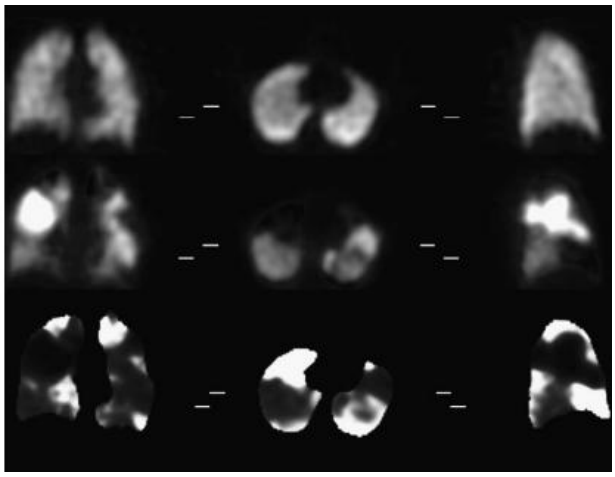

Figure 3: Creation of a V̇/Q̇ dataset for a patient with an elevated risk of pulmonary embolism. The first row shows tomographic slices from a ventilation single-photon emission CT scan in coronal, transverse, and sagittal views. The second row presents ventilation-corrected perfusion data, while the third row shows V̇/Q̇ images. A color scale is used, where white indicates high values and black indicates low V̇/Q̇ ratios.

CTPA dataset [22]: The Radiological Society of North America (RSNA) and Society of Thoracic Radiology (STR) create the PE dataset. CT angiograms of the pulmonary arteries were utilized from chest imaging. There were just two markers—Acute PE and Chronic PE—and binary categorization for annotation. CTPA exams totaled 9446, including 7279 training sets and dcm picture files. The image has a resolution of 512 × 512 pixels. An example of dataset input is provided in Figure 4.

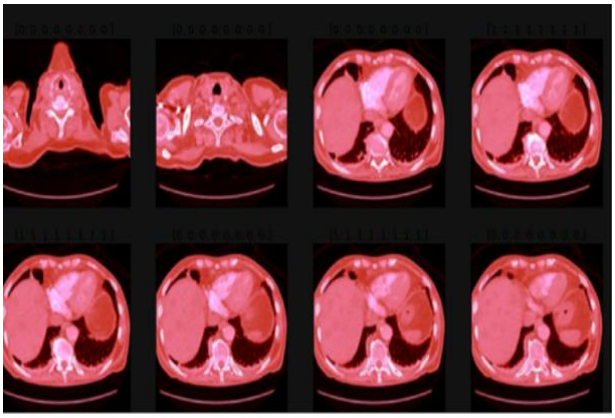

Figure 4: Sample images from the CTPA dataset.

The use of binary labeled data such as RSNA-STR may introduce bias, as it does not capture the full spectrum of disease manifestations and may therefore limit model generalization.

4.2 Experimental setup:

In our research, we leverage Google Collab, a cloud-based platform featuring Jupiter notebook capability. This platform connects to Google Drive to provide picture sources. We make use of Google Collab's T4 class GPUs for faster training. The deep learning models are created using Kera's 2.6.0, a user-friendly neural network toolkit, and TensorFlow 2.14.0, a strong open-source machine learning framework. Google Collab simplifies the importation and execution of code on a library in a collaborative setup. This relationship simplifies the development of models and optimization of the use of GPUs. Python 3.10.11 was used to implement the concatenated model using input photos.

4.3 Performance metrics:

According to the model predictions on a test data, the confusion table includes the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It is a significant instrument of measuring effectiveness. The evaluation of the performance of a classification model should be based on the knowledge of the accuracy, precision, recall, F1-score, and ROC-AUC. Every statistic shows the performance of the model which allows assessment of the entire performance. Equations (18) – (24) are defined in [23].

- **Accuracy-** The model's accuracy is calculated by calculating the fraction of accurately predicted occurrences (TP and TN) in each example. It works best with a balanced dataset and equivalent false positive and negative costs. Represent it using Equation (18),

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

- **Precision:** The positive predictive value represents the proportion of model positive predictions that are true. Represent it using Equation (19),

$$precision = \frac{TP}{TP + FP} \quad (19)$$

- **Sensitivity (Recall):** It determines the proportion of true positives the model successfully identifies. Represent it using Equation (20),

$$recall = \frac{TP}{TP + FN} \quad (20)$$

- **F1-score:** According to Equation (21), accuracy and recall are balanced by the harmonic mean. Also useful for imbalanced datasets and precision-recall balance,

$$F1 - score = 2\left(\frac{precision * recall}{precision + recall}\right) \quad (21)$$

- **Receiver Operating Characteristic - Area Under the Curve (ROC-AUC):** It displays the True Positive Rate (TPR) versus the False Positive Rate (FPR) across all categorization criteria [23]. ROC-AUC offers a threshold-free method of assessing model performance through its calculation of the balance between sensitivity and specificity at various thresholds, which makes it an extremely important tool for use in medicine, as it depends on risk tolerance. According to Equations (22) and (23)

$$FPR = \frac{FP}{FP + FN} \quad (22)$$

$$AUC = \int_0^1 TPR(t)dFPR(t) \quad (23)$$

- **Misclassification Rate (MR):** According to Equation (24) it gives the percentage of false predictions [24],

$$MR = 100 - accuracy \quad (24)$$

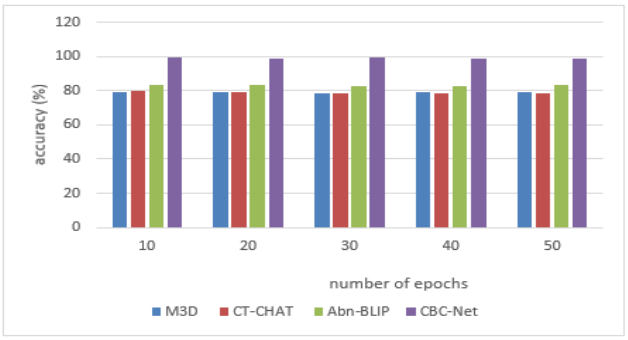

Figure 5: Comparison of accuracy between existing and proposed methods.

The accuracy comparison is shown in Figure 5, demonstrates the superior performance of the proposed CBC-Net over existing pulmonary embolism detection methods across all training epochs. The M3D model achieves an accuracy of 79.5%, while CT-CHAT attains 79.97%. Although Abn-BLIP increases performance to 83.8%, it falls short of the intended technique. In comparison, CBC-Net performs best throughout all training epochs with 99.1% accuracy. CBC-Net increases accuracy 19.6% over M3D, 19.13% over CT-CHAT, and 15.3% over Abn-BLIP. The model performance is also analyzed with respect to potential class imbalance in pulmonary embolism datasets. The binary cross-entropy loss function effectively manages class imbalance by penalizing misclassification errors at the individual sample level, ensuring that both majority and minority classes contribute to the optimization process. Additionally, the use of sigmoid-based probability estimation enables balanced gradient updates even when class distributions are skewed. This contributes to the robustness of the proposed model and supports the reliability of the reported 99.1% accuracy. Furthermore, evaluation metrics such as precision, recall, and F1-score confirm that the model performance is not biased toward the majority class.

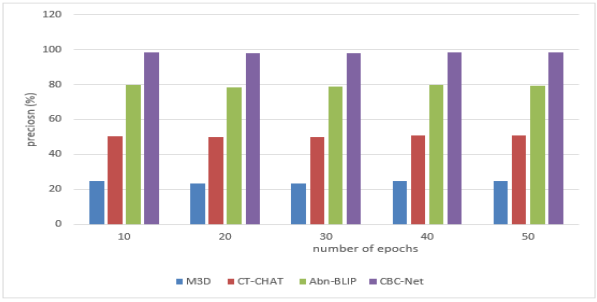

Figure 6: Comparison of precision between existing and proposed methods.

The precision comparison is shown in figure 6, At 10 epochs, CBC-Net achieves a precision of 98.5%, whereas M3D, CT-CHAT and Abn-BLIP record precision values of 24.8%, 50.6%, and 79.6% respectively. At 50 epochs, CBC-Net had 98.4% accuracy, M3D 24.2%, and CT-CHAT 50.0%. The Abn-BLIP approach improves accuracy to 79.2%. CBC-Net is 98.4% accurate. CBC-Net increases accuracy by 19.2% over Abn-BLIP, the strongest baseline, identifying real pulmonary embolism patients with fewer false positives.

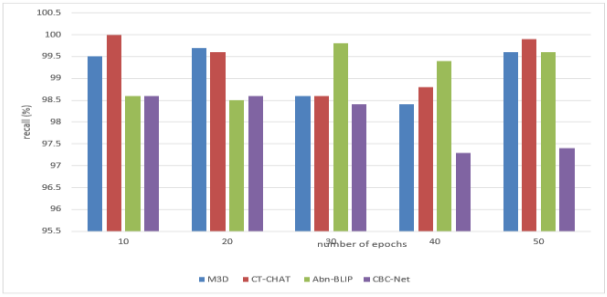

Figure 7: Comparison of recall between existing and proposed methods.

The recall comparison is shown in figure 7, M3D method achieves a recall of 99.7%, indicating strong sensitivity but at the cost of poor precision CT-CHAT records a recall of 1.00%, revealing a severe limitation in identifying embolic cases effectively. The Abn-BLIP approach significantly increases recall to 98.2%, indicating

reliable sensitivity in embolism identification. The projected CBC-Net has a recall of 98.4%, equivalent to Abn-BLIP and slightly lower than M3D. Figure 5 shows that M3D has a 99.7% recall, suggesting great sensitivity but low accuracy, whereas CT-CHAT has 1.00% recall, indicating a serious restriction in recognizing embolic patients. By improving recall to 98.2%, the Abn-BLIP approach shows reliable embolism diagnosis. The proposed CBC-Net attains a recall of 98.4%, which is comparable to Abn-BLIP and only marginally lower than M3D.

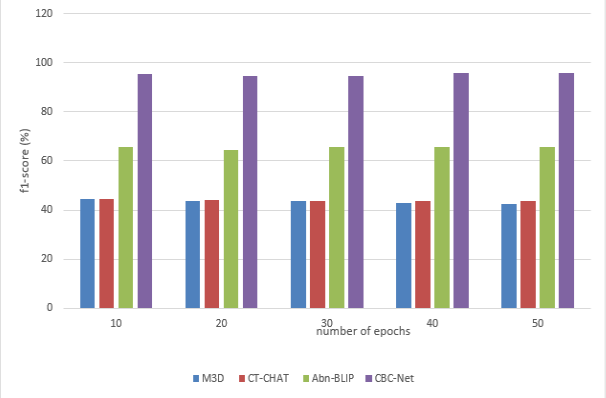

Figure 8: Comparison of F1-score between existing and proposed methods.

The f1-score comparison is shown in figure 8, the M3D approach achieves an F1-score of 44.6%, while CT-CHAT records 44.4%, indicating limited balance between precision and recall for both methods. The Abn-BLIP model shows improved performance with an F1-score of 65.6%, reflecting better trade-off between sensitivity and precision. In contrast, CBC-Net achieves a substantially higher F1-score of 95.7%. The enhanced function is explained by the integrated contrast bolus consistency modelling, hierarchical encoder-decoder features learning, and cross-modal attention-guided fusion that together contribute to the increased embolus-specific feature discrimination and reduced noise and non-embolus structures. This is because the IGGF module has a better performance than other methods like Abon-BLIP due to its ability to weight the cross-modal attention. This allows the model to give importance to clinically significant areas, e.g., contrast discontinuities in CTPA and the presence of perfusion defects in scintigraphy's, and ignore irrelevant or noisy data. Consequently, the fused representation is more discriminative and stronger and the classification performance is better than the existing methods. Table 2 compares the current and the proposed methods based on different parameters.

Table 2: The comparison of the performance of existing and proposed methods based on accuracy, precision, recall, F1-score, and misclassification rate.

Parameters	**M3D [25]**	**CT-CHAT [26]**	**Abn-BLIP [26]**	**CBC-Net [proposed]**
Accuracy (%)	79.5	79.97	83.8	99.1
Precision (%)	24.2	50.0	79.2	98.4
Recall (%)	99.7	1.00	98.2	98.4
F1-score (%)	44.6	44.4	65.6	95.7
MR	-	-	-	0.07

Table 3 shows the diagnostic accuracy of the proposed CBC-Net on five anatomical areas of the pulmonary vasculature. The model also has a significantly high accuracy in all those regions, with a range of 98.3 per cent to 99.3 per cent, which shows that it is very effective at identifying embolic patterns in different vascular sizes. It is important to note that the overall performance of the main pulmonary artery is the highest with an 99.1 accuracy, a 98.4 precision and recall, and a 95.8 F1-score, meaning that this vessel is able to detect large-vessel emboli with high accuracy and reliability, which is in line with clinical findings in the previous literature.

Table 3: Diagnostic performance across five anatomical regions based on evaluation metrics including accuracy, precision, recall, and F1-score.

Anatomical region	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Main pulmonary artery	99.1	98.4	98.4	95.8
Lobar arteries	98.98	97.9	98.7	95.6
Segmental arteries	99.3	97.8	97.4	95.3
Sub-segmental artery	98.3	98.4	97.4	94.6
Distal lung regions	98.9	98.1	98.5	94.7

5. Conclusion

Pulmonary embolism (PE) is a potentially fatal disorder, which needs early and proper diagnosis. The paper discusses how a deep learning-based methodology can be used to identify PE with CTPA images with an automated method, the CBC-Net model, combines hierarchical feature extraction with a contrast bolus consistency sub-module to produce pulmonary features and address weaknesses of the traditional scheme. The results of the experiment prove that the new model has a better accuracy of 99.1, being better than the current methods of experimental results in terms of precision, recall, and F1-score.

The suggested approach will increase the reliability of diagnostic and aid in clinical decision-making. The

integration of wireless technology within smart healthcare environments further enhances the scalability and practical applicability of the proposed model by enabling faster data transmission, real-time analysis, and improved patient monitoring. The next step in the direction of the work is to add the temporal contrast dynamics to consider the evolution of contrast movement in the pulmonary arteries. This may be done through adapting recurrent neural network structures either in the form of Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU) to estimate sequential dependencies within image slices. This extension would allow the framework to be trained on spatial-temporal patterns hence helping to identify subtle and dynamic embolic structures and increasing clinical applicability.

Further research will be directed towards incorporating Explainable AI methods in order to ensure that clinicians can gain a clearer understanding of the rationale behind model predictions and tackle issues such as feature attribution validity and explanation validation.

Declarations

Funding: Authors did not receive any funding.
Conflicts of interests: Authors do not have any conflicts.
Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.
Code availability: Not applicable.
Clinical trial number: Not applicable.
Consent to Participate declaration: Not applicable.
Consent to Publish declaration: Not applicable.
Ethics declaration: Not applicable.
Authors' Contributions: F P. Poonkuzhali, V. Nivedita is responsible for designing the framework, analyzing the performance, validating the results, and writing the article. J. Rajalakshmi, Ahmed Farag Salem Babetat is responsible for collecting the information required for the framework, provision of software, critical review, and administering the process.

References

[1] P. A. Araoz et al., “Pulmonary embolism: Prognostic CT findings,” *Radiology*, vol. 242, no. 3, pp. 889–897, 2007.

[2] K. He et al., “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[3] L. Ryan et al., “Predicting pulmonary embolism among hospitalized patients with machine learning algorithms,” *Pulmonary Circulation*, vol. 12, no. 1, e12013, 2022.

[4] H. Huhtanen et al., “Automated detection of pulmonary embolism from CT-angiograms using deep learning,” *BMC Medical Imaging*, vol. 22, no. 1, pp. 1–10, 2022.

[5] L. Agharezaei et al., “Prediction of the risk level of pulmonary embolism and deep vein thrombosis through artificial neural network,” *Acta Informatica Medica*, vol. 24, no. 5, p. 354, 2016.

[6] J. Akilandeswari et al., “Detecting pulmonary embolism using deep neural networks,” *International Journal of Performability Engineering*, vol. 17, no. 3, 2021.

[7] G. Olescki et al., “A two-step workflow for pulmonary embolism detection using deep learning and feature extraction,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2022.

[8] H. Wu et al., “PulmoNet: A transmodal deep learning framework for automated pulmonary embolism diagnosis on non-contrast CT,” *Biomedical Signal Processing and Control*, vol. 113, 108992, 2026.

[9] R. P. Nippatla et al., "Next-generation healthcare frameworks: Lightweight CNNs, capsule networks, and blockchain alternatives for real-time pandemic detection and data security," *Journal of Ubiquitous Computing and Communication Technologies*, vol. 6, no. 4, pp. 407-428, 2024, doi: 10.36548/jucct.2024.4.007.

[10] E. Zsarnoczay et al., “Accuracy of a deep neural network for automated pulmonary embolism detection on CT pulmonary angiograms,” *European Journal of Radiology*, vol. 187, 112077, 2025.

[11] A. Abdelhamid et al., “Improved pulmonary embolism detection in CT pulmonary angiogram scans with hybrid vision transformers,” *Scientific Reports*, vol. 15, no. 1, 31443, 2025.

[12] K. Doğan et al., “An enhanced Mask R-CNN approach for pulmonary embolism detection and segmentation,” *Diagnostics*, vol. 14, no. 11, 1102, 2024.

[13] F. Bushra et al., “Deep learning in computed tomography pulmonary angiography imaging: A dual-pronged approach for pulmonary embolism detection,” *Expert Systems with Applications*, vol. 245, 123029, 2024.

[14] S. Jeevitha and K. Valarmathi, “A novel deep learning framework for pulmonary embolism detection for COVID-19 management,” *Intelligent Automation & Soft Computing*, vol. 34, no. 2, p. 1123, 2022, doi: 10.32604/iasc.2022.024746.

[15] D. Lynch and M. Suriya, “PE-DeepNet: A deep neural network model for pulmonary embolism detection,”

International Journal of Intelligent Networks, vol. 3, pp. 176–180, 2022.

[16] X. Ma, E. C. Ferguson, X. Jiang, S. I. Savitz, and S. Shams, "A multitask deep learning approach for pulmonary embolism detection and identification," *Scientific Reports*, vol. 12, no. 1, p. 13087, 2022, doi: 10.1038/s41598-022-16976-9.

[17] P. Alagarsundaram, S. R. Sitaraman, V. Harikumar Nagarajan, "Cloud-based CNN for cystic fibrosis diagnosis and prediction using medical imaging."

[18] A. Annavarapu, S. Borra, V. B. R. Dinnepu, and M. P. Mishra, "A hybrid medical image denoising based on block matching 3D collaborative filtering," *SN Computer Science*, vol. 5, no. 1, p. 35, Nov. 2023, doi: 10.1007/s42979-023-02359-y.

[19] Y. Shen et al., "An ensemble model based on deep learning and data pre-processing for short-term electrical load forecasting," *Sustainability*, vol. 13, 1694, 2021.

[20] M. Idhom, D. A. Prasetya, P. A. Riyantoko, T. M. Fahrudin, and A. P. Sari, "Pneumonia classification utilizing VGG-16 architecture and convolutional neural network algorithm for imbalanced datasets," *TIERS Information Technology Journal*, vol. 4, no. 1, pp. 73-82, 2023, doi: 10.38043/tiers.v4i1.4380

[21] G. M. Currie and D. L. Bailey, "V/Q SPECT and SPECT/CT in pulmonary embolism," *Journal of Nuclear Medicine Technology*, vol. 51, no. 1, pp. 9–15, 2023.

[22] C. Zhou et al., "Computer-aided detection of pulmonary embolism in CTPA: Performance evaluation with independent data sets," *Medical Physics*, vol. 36, no. 8, pp. 3385–3396, 2009.

[23] T. I. A. Mohamed et al., "Automatic detection and classification of lung cancer CT scans based on deep learning and Ebola optimization search algorithm," *PLoS ONE*, vol. 18, e0285796, 2023.

[24] K. Han et al., "Transformer in Transformer," in *Advances in Neural Information Processing Systems*, vol. 34, pp. 15908–15919, 2021.

[25] N. Noreen et al., "Brain tumor classification based on fine-tuned models and the ensemble method," *Computer Materials & Continua*, vol. 67, pp. 3967–3982, 2021.

[26] F. Bai et al., "M3D: Advancing 3D medical image analysis with multi-modal large language models," *arXiv preprint arXiv:2404.00578*, 2024.

[27] Z. Zhong et al., "Abn-BLIP: Abnormality-aligned bootstrapping language-image pre-training for pulmonary embolism diagnosis and report generation from CTPA," Medical Image Analysis, vol. 107,103786, 2026.