Optimizing Pulmonary Embolism detection through diverse UNET architectural variations

Renu Vadhera[1], Meghna Sharma[2]
[1]*Ph.D Research Scholar (CSE), The NorthCap University, Gurugram, 122017, India*
Email: renu19csd003@ncuindia.edu[1]
[2]*Associate Professor (CSE), The NorthCap University, Gurugram, 122017, India*
Email:meghnasharma@ncuindia.edu[2]

Abstract

A Pulmonary Embolism (PE) is a critical condition that poses a life-threatening risk when blood vessels in the lungs become obstructed. To detect and precisely locate PE, medical professionals typically employ a specialized X-ray technique known as Computed Tomography Pulmonary Angiography (CTPA). Leveraging Deep Learning methodologies, particularly U-shaped encoder-decoder architectures, has emerged as a promising avenue for automating PE segmentation from CTPA images. This research endeavours to assess and compare the performance of several U-shaped networks, including UNET, UNET++, Residual UNET, ARUX, and Attention UNET, in accurately segmenting PE regions. Utilizing a publicly available PE challenge dataset, comprehensive training, validation, and testing procedures are conducted, with a meticulous evaluation based on metrics such as the Dice Coefficient, Jaccard Similarity Index, Sensitivity, alongside considerations of training time and model parameters. The findings from this study offer valuable insights into the efficacy and suitability of various deep learning architectures for PE segmentation, paving the way for enhanced diagnostic capabilities in clinical settings.

Keywords: Pulmonary Embolism (PE) ; Computed Aided Design; CTPA; UNET; Deep Learning; PE segmentation;

1. Introduction

Pulmonary Embolism (PE) stands as the predominant cardiovascular ailment in the United States, associated with significant morbidity and disability [1]. PE is the blood clots in the pulmonary arteries, which migrate through the bloodstream from other regions of the body to the lung. The obstruction in the pulmonary arteries can result in insufficient oxygen supply to the body. This may lead to unexpected death and severe cases of extremely low blood pressure may be there [2]. Doctors must confirm the PE illness using a range of diagnostic techniques to quickly diagnose it, although its symptoms resemble with other diseases as well [5]. Chest pain, shortness of breath, hemoptysis, hypotension, loss of consciousness and tachycardia are some of the non-specific symptoms of pulmonary embolism. It is very challenging task to

diagnose PE by looking at symptoms only. Prompt medical attention is crucial for diagnosing and managing any type of pulmonary embolism, as it can be a serious condition if not treated promptly [5]. Some of the diagnostic procedures for PE include blood test, an electrocardiogram, computed tomography (CT) scan [6], chest X-ray and CTPA [2]. The most popular and effective approach for its detection is Computed Tomography Pulmonary Angiography (CTPA), due to greatly better visualization [8]. Single CTPA consists of number of slices (images), so manually study of single CTPA is time consuming task. The increasing patient population and the shortage of physicians could influence both medically underserved individuals and modern healthcare institutions [3]. Numerous automated systems have been created for automatically detection of PE, showing outperform the expert panels. According to a study, the sensitivity range for a typical radiologist is between 77% and 94%. These automated systems can help radiologists in automatic PE detection with greater sensitivity and speed [8].

2. Related Work

The diagnosis of PE in CTPA is not an easy task because PE's position is not defined. It can be in central position, segmental or sub-segmental arteries. Radiologists should carefully examine the extensive system of pulmonary arteries using several CT slices in order to look for a filling defect. Pulmonary Embolism comes in many shapes and intensities. Radiologists must pay close attention to the visual signs of PE because of artery blockage. Third, when other pulmonary disorders are present or the quality of the CT picture is poor, the ability to diagnose PE may be hindered. The use of computer-aided diagnosis for pulmonary embolism has demonstrated effectiveness in enhancing the diagnostic abilities of radiologists when assessing PE [27]. The medical industry has not adopted computer aided diagnosis software because shows high number of false positives. Many researchers are attempting to lower the number of false positives with new methodologies and strategies [7]. CTPA images are most effective method used by CAD systems to diagnose PE. CAD systems commonly handle the tasks of classifying, detecting, and segmenting pulmonary embolism. In the context of medical imaging, detection refers to finding location of PE in the image (CT scan) and classification involves categorizing images as either having or lacking pulmonary embolism, while segmentation entails isolating and extracting the specific area corresponding to pulmonary embolism from an image.

A procedure known as segmentation that divides an image into parts with comparable characteristics such as grey level, different color, brightness, texture and contrast medical image segmentation is very helpful in the diagnosis of various diseases [29]. UNET is a well-known deep learning (DL) model that may integrate shallow information with leveraging profound information to generate impactful segmentation outcomes is crucial in the realm of semantic segmentation applied to biomedical applications. For PE lesion

segmentation, recent studies have only used supervised learning techniques. UNET models with various dimensions (2D, 2.5 D, 3D) are used to segment PE from CTPA and analyzed how these models different [17]. Table 1 summaries all the approaches used in detection, segmentation, and classification of PE.

Table 1. Summary of related papers

Paper	Year of publication	Dataset	Size of dataset	Methodology	Evaluation techniques	Category	Investigative Lapse
[12]	2007	Private Dataset	12 CT scans	Tobogganing, ANN & kNN	sensitivity	Machine Learning (ML)	segmentation of the vascular structure hampers the overall performance
[10]	2008	Private Dataset	38 CTA	Feature calculation & classification techniques	sensitivity	Machine Learning (ML)	computationally complex, require high-quality data, and risk overfitting
[13]	2015	Public Dataset (PE challenge and Private dataset	141 CTPA	Tobogganing & CNN	sensitivity	Machine Learning and Deep Learning	CNN struggles to effectively eliminate false positives
[14]	2017	Private Dataset	33 CTPA	ANN, SVM & kNN	sensitivity	Machine Learning (ML)	Dataset is very small.
[15]	2019	Both Public (PE challenge) dataset & Private dataset	149 CTPA	Two stage CNN	sensitivity	Deep Learning	Require high computation power
[4]	2020	Private Dataset	2592 CT studies	2D UNet & 2D Convolution LSTM	AUC	Deep Learning	Only few parameters are used.
[16]	2020	Private dataset	590 CTPA	UNET	sensitivity & specificity	Deep Learning	It is sensitive to input variations
[9]	2020	Private dataset (Kinetics-600 dataset)		3D CNN	AUROC	Deep Learning	Performance with imaging artifacts is not checked

[17]	2020	Public dataset	80 CTPA	2D UNET, 2.5D UNET & 3D UNET	sensitivity, Dice Coefficients, FROC curve	Deep Learning	Challenged in efficiently detecting sub-0.5 mL emboli.
[19]	2020	Private dataset	5856 (positive) and 5196 (negative) Scans	CNN & RCNN	AUC	Deep Learning	Extra computation power required
[20]	2021	Public (RSPECT) dataset	7279 TPA	CNN & LSTM	AUC	Deep Learning	Extra computation power required
[21]	2021	Public (CHAOS) dataset & Private dataset	70 CTPA	Residual Unet	Dice coefficient, Precision, Recall	Deep Learning	It segments only pulmonary arteries
[11]	2022	Private dataset	20 CTPA	ANN with k-NN, Tobogganing & Genetic Algorithm	sensitivity	Machine Learning (ML)	True Positive rate is very low and small dataset is used.
[22]	2021	Public (CHAOS) dataset & Private dataset	120 CTPA	Residual Deep UNET	Precision, Recall & SSIM, DSE	Deep Learning	only segments the pulmonary arteries
[23]	2022	Private dataset	800 CTPA	CNN & LSTM	sensitivity and specificity	Deep Learning	models lack the capability to offer precise information regarding the localization of emboli
[24]	2022	Public (FUMPE) dataset & private (NCKUH) dataset	200 CTPA	Classification & segmentation			Small individual lesions are prone to misdiagnosis

[18]	2023	Public (FUMPE) dataset & private (NCKUH) dataset	132 CTPA	HRNet architecture	sensitivity	Deep Learning	model's accuracy did not improve on the testing data with the FUMPE dataset

3. Material and Methods

Chest CT scans, specifically in the form of CTPA, undergo analysis where distinct sections of the images are identified and labeled through the application of a widely recognized semantic segmentation technique. Semantic segmentation facilitates the partitioning of images into regions characterized by comparable intensity, homogeneity, and texture. In this work various deep learning based UNET models are used to segment PE. Input of model is CTPA image and output is segmented mask (which highlights the affected area). The performance of models is evaluated using computing dice coefficient, Jaccard similarity index and sensitivity.

3.1 Dataset

In this work publicly available PE challenge is used [31]. This dataset includes medical images and the corresponding reference standard in the form of mask images. It contains CTPA of 91 patients. Each CTPA consists of 3D volumetric data has axial plane, sagittal plane and coronal plane. The dataset was created for the IEEE International Symposium of Biomedical Imaging (ISBI) challenge cad-pe. Medical data was acquired from six different hospitals that emphases the Radiologic Imaging of the Madrid. Image size is of 512 × 512. The thickness of each reconstructed slice falls within the range of 0.75 to 1.5 mm while the image size of pixel is between 0.58 and 0.85 mm. Each CTPA is labelled by at least one radiologist having experience more than 10 years. Each radiologist diagnosed all emboli present in an image separately by finding the effective area in each axial slice. Sagittal and coronal images were examined to determine whether emboli were present. Segmentation is done based on thresholding followed by connected component analysis. Image specialist eliminated useless pixels and defined boundary for each segmentation in every image [28]. Figure 1 shows the axial plane that is a cross-sectional view of the body. It is like slicing a loaf of bread horizontally. In a CT scan, axial images are obtained as if you were looking down at a person or object from above. Figure shows the coronal plane, which is another vertical section, separates the body or object into its front and rear portions. In CT scans, coronal images offer a frontal view, akin to directly facing a person and figure 3 sagittal plane represents a vertical division that separates the body or object into its right and left sides. Sagittal CT images offer a lateral view, enabling the observation of structures from a sideways perspective.

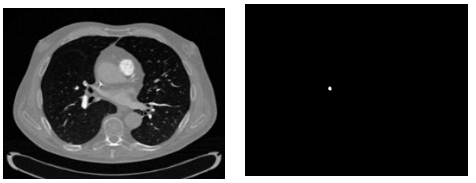
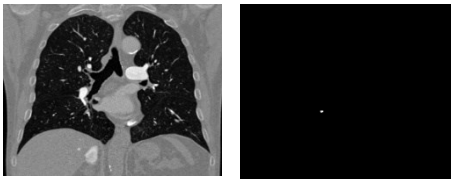
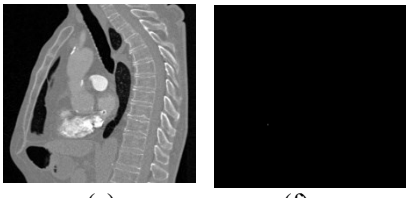


Figure 1. (a) Axial view of image
(b) Axial view of corresponding mask



(c) (d)

Figure 2. (c) coronal view of image
(d) coronal view of corresponding mask



(e) (f)

Figure 3. (e) Sagittal view of image
(f) Sagittal view of corresponding mask

3.2 Pre-processing of Data

Data is pre-processed before feeding to the model. Firstly, all CTPA axial slices are extracted, then normalize the values of pixels to a desired range, such as between 0 and 1. This step helps in standardizing the intensity values and making them suitable for further processing or analysis. Min-max normalization is employed to standardize the data, involving the calculation of the minimum and maximum pixel values within the image. The formula used to normalize each pixel value in the image is shown in equation (1).

$$\boldsymbol{Yi = max(x) - min(x)/(xi - min(x))} \quad (1)$$

where

$\boldsymbol{Yi}$: normalized intensity value of i^{th} pixel

$\boldsymbol{xi}$: i^{th} pixel intensity value

$\boldsymbol{max(x)}$: The highest intensity value in the image. $\boldsymbol{min(x)}$: The lowest intensity value in the image

Figure 4. presents a normalized CTPA for a single patient, comprising all axial slices within a stack. The number of slices in each CTPA can range from 102 to 350. Additionally, Figure 5 displays the corresponding ground truth mask images for the slices presented in Figure 4.

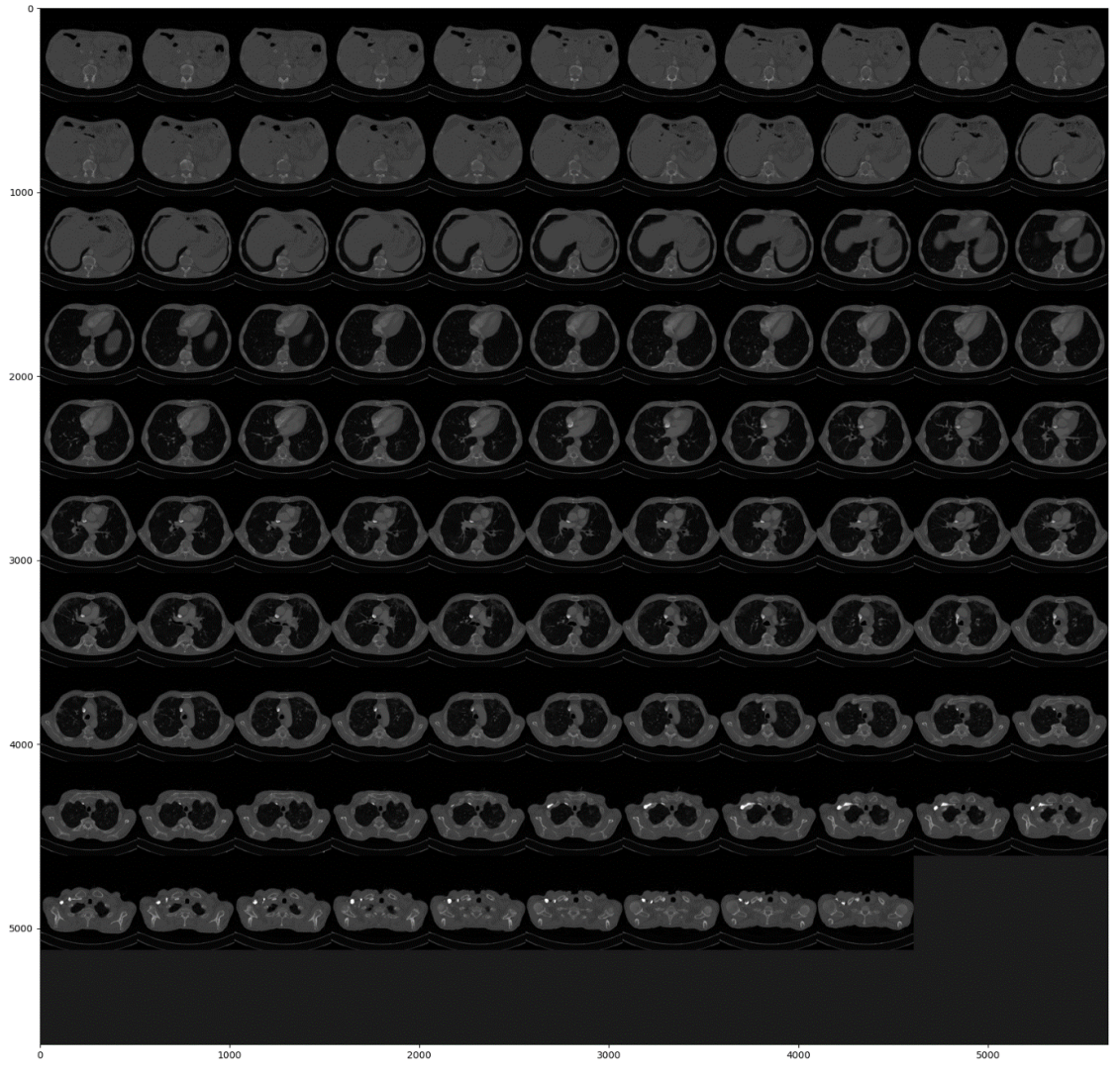


Figure 4. CTPA images of one patient

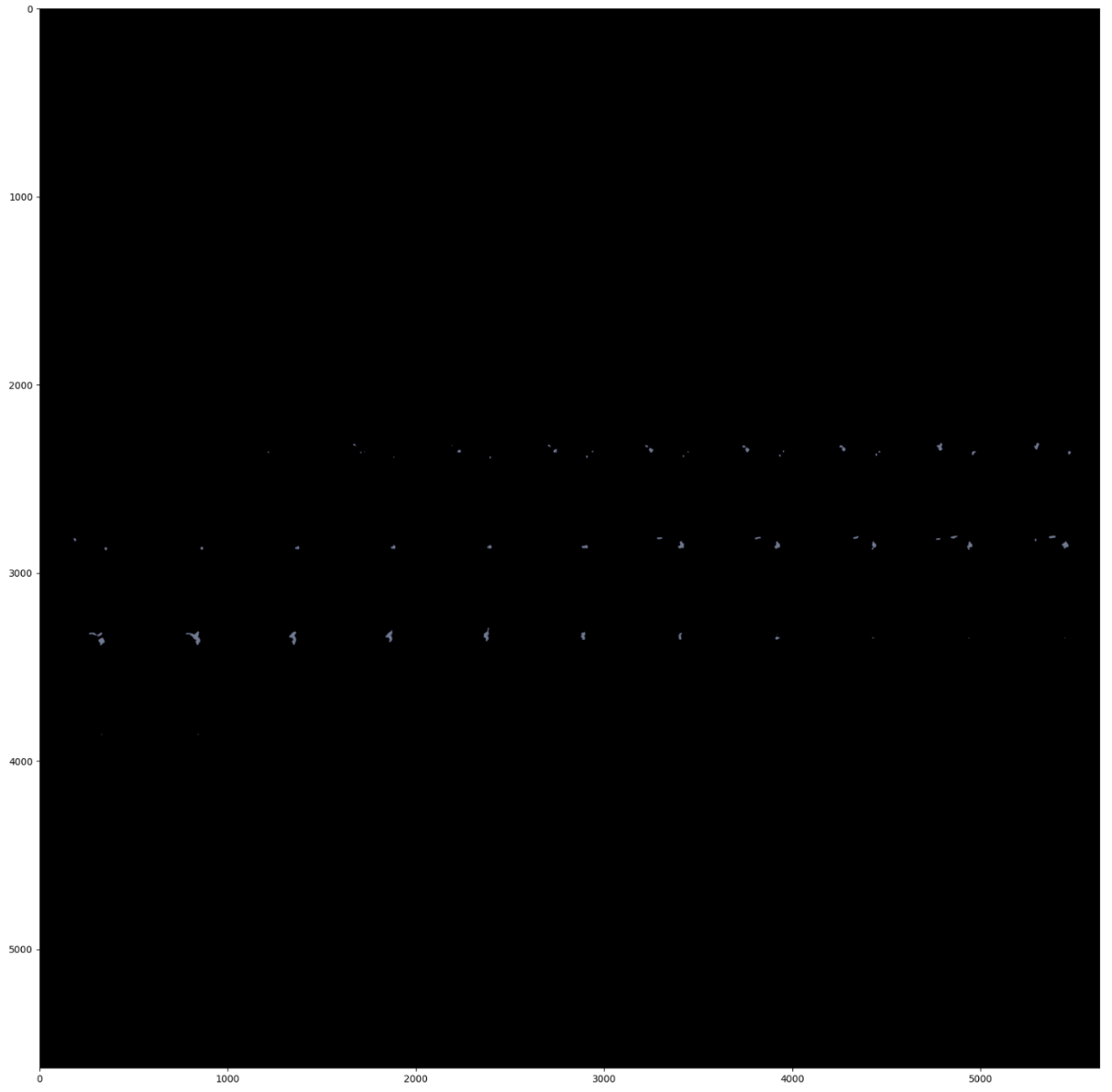


Figure 5. Masked images of corresponding images shown in Figure 4

After reading and visualizing the all slices and its corresponding mask images. it has been observed that some of upper and lower slices (images) are blank. It means upper and lower slices does not contain much information about presence of PE. The required information or features needed to detect PE are available only in middle slices. To improve accuracy and save resources some of upper and lower slices are removed from each CTPA as shown in figure 6. To speedup training and reduce computation cost all slices are resized to 256 × 256. At last data is divided into training and testing.

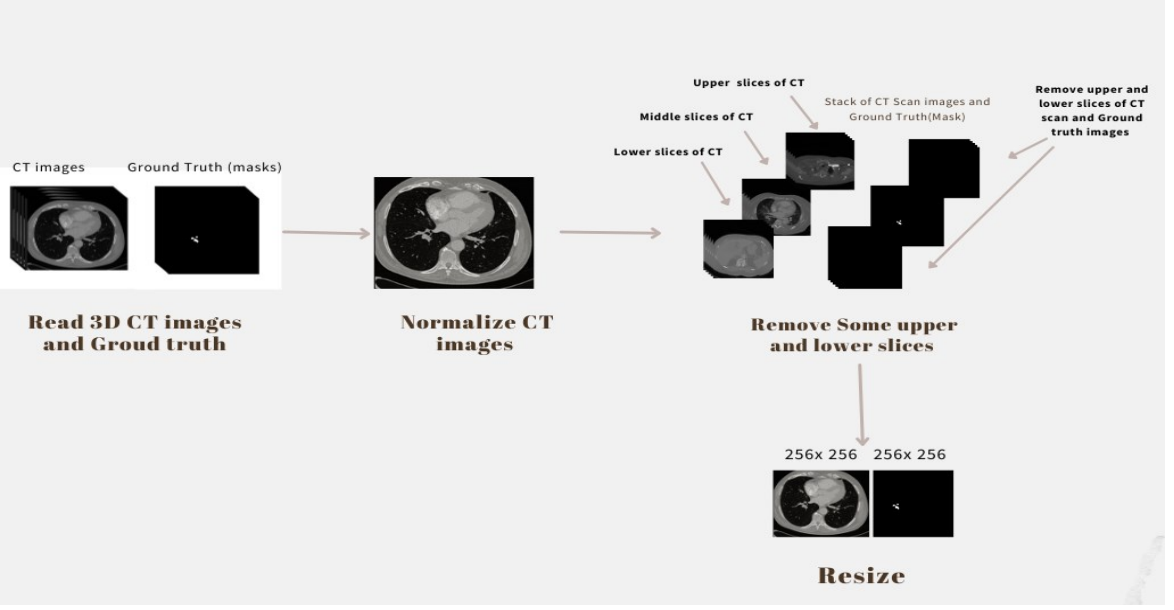


Figure 6. Pre-processing of Data

3.3 Network Architectures

This work explores and implements several segmentation models, including UNET, Residual UNet, UNet++, Attention UNet, and ARUX models. A thorough analysis is conducted on diverse similarity matrices, and the results are presented through both qualitative and quantitative comparisons.

3.3.1 UNET model

Deep learning based UNET model is widely used medical image segmentation [29] as shown in figure 7. UNET model consist of Encoder and Decoder part and there are bridge layers to combine the encoder and decoder part of UNET. Four blocks are there in each encoder and decoder part. The Encoder path refers to the standard convolutional network architecture where each block uses a convolution layer,

followed by ReLU, and down sampling is accomplished using a 2x2

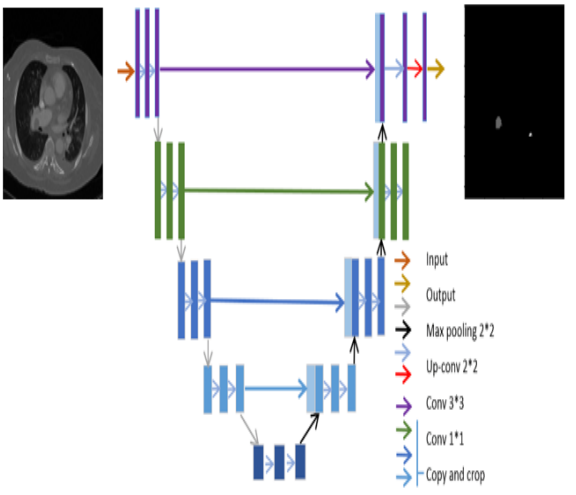


Figure 7. UNET model architecture [17]

max pooling process. Every time in downscale feature channels is doubled. To keep the input and output size consistent during the convolutional operation, the same padding is always applied to

the input tensors. To combine Encoder and decoder part, in the bridge we used two convolution layers one after the another. One layer takes input from encoder and another layer provide the input to decoder. The decoding pathway begins with an up-sampling of a feature map at each step, followed by a 2x2 convolution that results in a halving of the number of feature channels. Subsequently, there is a concatenation with the uniformly cropped feature map from the contraction path, followed by a 3x3 convolution with a ReLU activation. Last layer employs 1x1 convolution. UNET model is effective for pixel level segmentation, provides fast inference. UNET's disadvantage is that it prevents convergence during training and results in vanishing gradients. There is another problem of insufficient feature extraction in UNET [17].

3.3.2 ResUNet (Residual UNet) model

To overcome the problem of insufficient feature extraction in UNET [17] another enhanced ResUNet [21] model is used for PE segmentation. It facilitates effective information flow as its architecture is shown in figure 8.

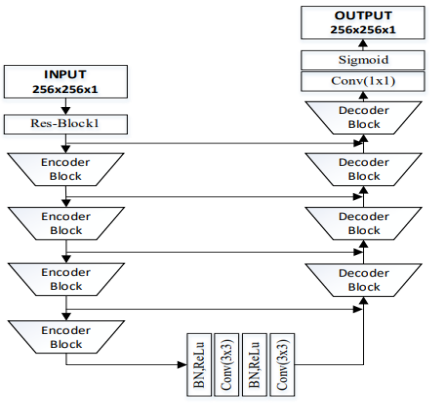


Figure 8. ResUNet model architecture [21]

The UNET structure is first chosen as the base structure. Second, this model adds residual connections to enhance the flow of gradients within the network. In improved ResUNet the simple convolution layers are replaced with residual blocks that mix the data from the bottom and top layers to reduce loss while feature extraction. Four encoder and decoder blocks are used in improved ResUNet. The key addition in ResUNet is the inclusion of residual connections. These types of connections permit the network to directly pass the input features from one layer to a deeper layer, thereby creating shortcut connections. These shortcuts enable the network to efficiently propagate gradients and learn residual mappings, which can help to improve the overall training process by improving the vanishing gradient problem. In the decoding path of ResUNet, the spatial resolution is gradually increased through up sampling operations. Like the encoding path, each block in the decoding path often contains two convolutional layers. Finally, the output of ResUNet is generated by a 1x1 convolutional layer, which maps the extracted features to the desired output. By incorporating residual connections, ResUNet aims to improve gradient flow, enhance the learning of deep representations, and potentially boost the segmentation performance.

3.3.3 UNet++ model

To reduce information, lose during the encoding and decoding stages of UNET [17] and Res-UNet [21], another model called UNet++ [25] was introduced. Ts architecture is shown in figure 9.

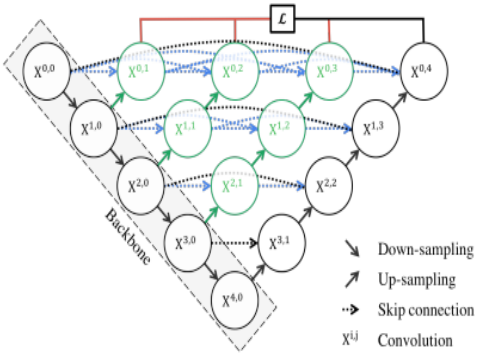


Figure 9. UNet++ model architecture [25]

It represents a refinement of the UNET architecture, incorporating multiple adjustments to improve overall performance and segmentation accuracy. The key idea behind UNET++ is to capture multi-scale contextual information by incorporating nested and densely connected skip pathways within the network. This helps the model effectively capture both local and global context while maintaining high-resolution feature maps. The structure comprises an encoding path that gradually diminishes spatial dimensions through a combination of convolutional layers and pooling operations. In decoding path, it employs nested and densely connected skip pathways. These skip connections facilitate the integration of multi-scale contextual information, allowing the model to capture local and global context while maintaining high-resolution feature maps. UNet++ has dense blocks within each stage, that promotes feature reuse and efficient learning. The transition up operation increases spatial resolution, and the final output is generated through a 1x1 convolutional layer. It shows better results than ResUNet and UNET. UNet++ needs lots of computation power [25], thus it is slow than base model.

3.3.4 Attention-UNet model

Attention-UNet [26] is another form of the UNET model that includes self-attention mechanisms to improve model's capability to emphasis on relevant image features through segmentation process. As shown in figure 10, It combines the UNET's U-shaped architecture with the attention mechanism, which allows the network to selectively attend to different spatial locations and feature channels.

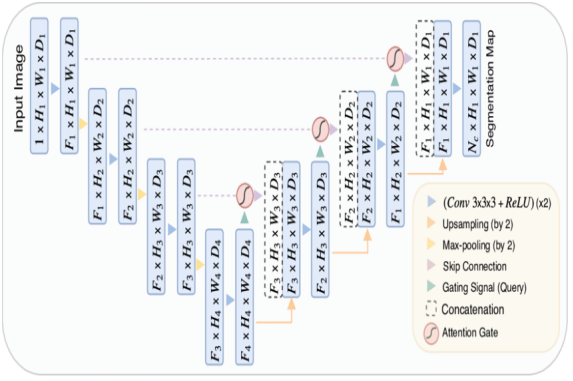


Figure 10. Attention-UNet model architecture [26]

In Attention-UNet, the encoding path follows the standard UNet structure, gradually reducing the spatial dimensions through convolutional layers and pooling operations. In every phase of the encoding path, there are several convolutional layers followed by a down sampling operation. The key addition in Attention-UNet is the integration of self-attention modules in the decoding path. Self-attention mechanisms calculate attention weights for each spatial

location within a feature map based on the relationships between different spatial locations. These attention weights are then used to weight the importance of different features. In the decoding path of Attention-UNet, the attention modules are typically applied before the up-sampling operations. The segmentation accuracy is increased by the attention mechanism, which allows the model to choose focus on relevant spatial areas and feature channels while suppressing noise or irrelevant regions. Additionally, Attention- UNet often employs skip connections, like the original UNET architecture, to transfer features from encoding part to decoding part. These skip connections help maintain the high-resolution information in both local and global framework. Self-attention mechanisms of Attention-UNet aims to improve the model's ability to attend to informative features while suppressing irrelevant ones, leading to better segmentation performance.

3.3.5 ARUX (Attention Residual UNET) model

Two CNN structures make up the ARUX-Net [24] architecture as shown in figure 11.

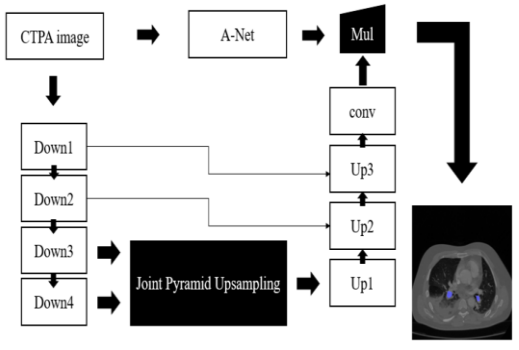


Figure 11. ARUX-Net model architecture [24]

RUX-Net constitutes the architecture of an encoder-decoder, while A-Net represents the architecture of an attention mechanism. CNN structure consists of four layers and an A-Net. It executes an encoder and decoder blocks in RUX-Net utilizing four down sampling layers, a join pyramids up sampling instrument, and three up sampling layers. Each down sampling layer used a squeeze-and-excitation layer, two convolution sets, and down sampling. Every convolution set included a convolution layer, cluster standardization layer, and actuation layer. there is a squeeze-and-excitation layer, which involves the concatenation of Down3 and Down4's feature maps and the refinement of the feature representation through joint pyramid up sampling. After that, three up sampling layers were used to up sample the feature map, and a shortcut is to integration of features from Down1 and Down2 was achieved using U-Net. It provides the benefit of both Attention [26] and residual blocks [21]. It solves the vanishing gradient problem via residual block and attention mechanisms, enabling the model to concentrate on the target areas.

3.5 Evaluation matrices

To evaluate the segmentation models, two-similarity metrics Dice Coefficient [30], Jaccard Similarity Index [30] are used as shown in table 2. These metrics provide a measure of similarity between the predicted output results and ground truth annotations. The alignment between the anticipated segmentation masks and the actual segmentation masks is evaluated quantitatively using these evaluation matrices. They provide as a means of evaluating the predictability, excellence, and similarity of the segmentation findings in an objective manner. The Dice coefficient is

frequently employed in image segmentation to evaluate the similarity between predicted masks (P) and ground truth masks (G). It quantifies the overlap or intersection between the regions identified as true positives and those predicted by the model. It is defined as twice the overlap between the predicted and ground truth regions to the sum of the sizes of the predicted and true positive regions. The Jaccard Similarity Index (JSI) which is also called as the Jaccard coefficient or Jaccard index, Intersection over Union evaluate the overlap between the predicted masks and the ground truth masks. It measures the agreement between the two sets of

Table 2. Similarity matrices used in this study

sSimilarity Coefficient	**Formula (P=Predicted, G=Ground truth)**	**Equation (p=predicted, g=ground truth)**
Dice Similarity Coefficient [30]	$DSC(P,G) = 2 * \|P \cap G\| / (\|P\| + \|G\|)$	$\frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}$
Jaccard Similarity Index [30]	$J(P,G) = \|P \cap G\| / \|P \cup G\|$	$\frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 - \sum_{i=1}^{N} p_i g_i}$

Sensitivity of all the models is also calculated. The formula to calculate the sensitivity is shown in equation 2 where p_i represents the predicted mask and g_i represents the ground truth or mask image.

$$\text{Sensitivity} = \frac{\sum_{i=1}^{N} \min(1, p_i g_i)}{\sum_{i=1}^{N} \min(1, g_i)} \quad (2)$$

3.6 Implementation Details

In this study an extensive investigation is conducted into training deep learning models for the detection of Pulmonary Embolism (PE) in Computed Tomography Pulmonary Angiography (CTPA) images. To train the models, a dataset consisting of CTPA images from 60 patients, sourced from the PE challenge dataset is utilized. Specifically, 80% of the dataset, totaling 16,789 slices, was allocated for training, while the remaining 20% (3,357 slices) was reserved for validation purposes for optimizing the models, the Adam optimizer is used as it is recognized for its superior performance compared to other optimizers. Adam represents an advancement in stochastic gradient descent algorithms, integrating adaptive learning rates and momentum. This optimizer maintains adaptive learning rates for each parameter in the model, computed by aggregating first and second moment estimations of the gradients. After hyperparameter tuning the learning rate is initialized at 0.005 for training, ensuring effective optimization. Additionally, a dropout rate of 10% was fixed to prevent overfitting. To introduce non-linearity and enable feature extraction, the Rectifier Linear Unit (ReLU) activation function is applied to all layers except the final one. The models underwent

training for a total of 200 epochs, with the implementation of early stopping mechanisms. This strategy aimed to mitigate overfitting issues and enhance model generalization. Training was halted when there was no improvement in performance on the validation set for 40 consecutive epochs. The hyperparameters are meticulously documented and utilized for model training and validation, as outlined in Table 3. Furthermore, the models are trained on hardware equipped with 15GB of GPU RAM, 40GB of system RAM, and an 80GB hard disk, ensuring computational efficiency and storage adequacy for model training and evaluation. This exhaustive study aimed to provide the comprehensive insights into the training process of deep learning models for PE detection in CTPA images, leveraging state-of-the-art techniques and optimizing model performance for enhanced diagnostic accuracy and efficiency.

Table 3. Summary of the hyperparameter used in this study

Batch Size	32
Dropout Rate	10%
Optimizer	Adam
Learning rate	0.005
Epochs	200 (Early stopping is used)
Training Data	80%
Validation Data	20%
Matrices	dice coefficient, Jaccard similarity Index, Sensitivity
Loss function	1-dice coefficient

4. Experimental Results and analysis

Comparative analysis of Deep learning models UNET[17], Residual UNet[21], Attention UNET[26], ARUX[24], UNET++[25] has been done. To compare all the models on an equal footing, all models have been trained with the same hyperparameter configuration. Google collaboration platform is used with Keras framework and TensorFlow as backend. Training and validation performance with dice coefficient, Jaccard Similarity index and sensitivity with Adam optimizer are shown in figure 12(a), 12(b) and 12(c) respectively.

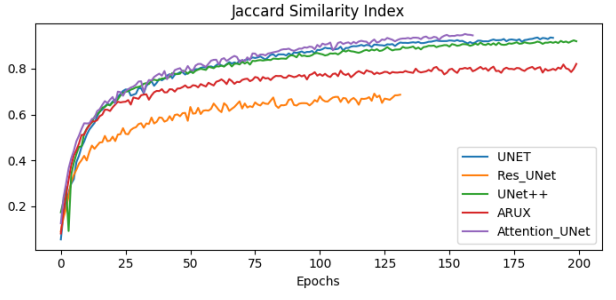


Figure. 12(a) Jaccard Similarity Index Analysis

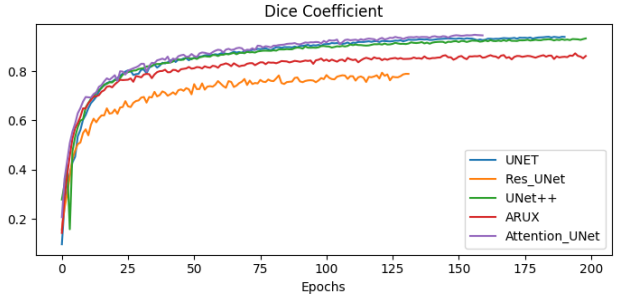


Figure 12(b). Dice Coefficient Analysis

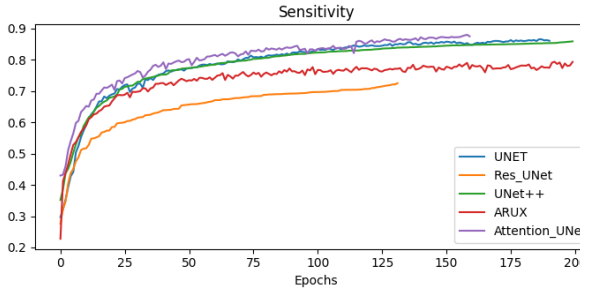


Figure 12(c). Sensitivity Analysis

4.1 Comparative study of the models

This section shows the quantitative analysis of the models. Table 4 shows performance metrics and table 5 represents the resource requirement and training parameters, which is graphically represented in figure 14 (a & b). These models have been evaluated using various evaluation criteria, including the Dice Coefficient, Jaccard Similarity Index, and Sensitivity. Figure 13 shows the comparative analysis on test data.

Table 4. Model results

sModel	Dice Coefficient	Jaccard Similarity Index	Sensitivity
UNET [17]	**0.9175**	0.9074	0.8673
ResUNET [21]	0.7952	0.7024	0.6516
UNET++ [25]	0.9017	**0.9132**	**0.8677**
ARUX [24]	0.8736	0.8175	0.8295
Attention- UNet [26]	0.9092	0.8720	0.8121

Table 5. Training parameters and resource requirements

Model	Training Time	No. of parameters trained
UNET [17]	2 hr 32 min	8641697
ResUNET [21]	3 hr 35 min	14152417
UNET++ [25]	3 hr 4 min	1493521
ARUX [24]	3 hr 6 min	476799
Attention-UNET [26]	2 hr 55 min	8211297

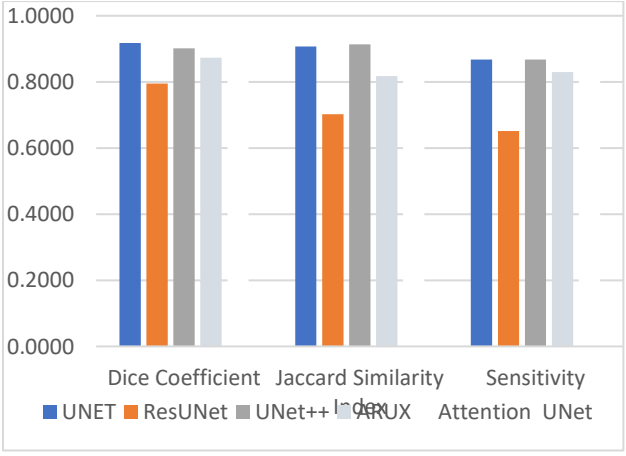


Figure 13. comparative analysis of models

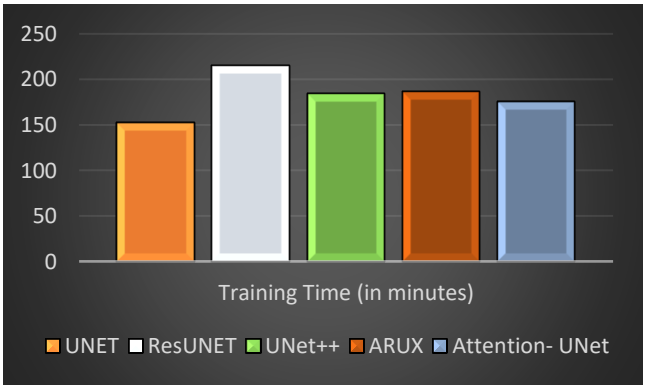


Figure 14 (a). Model's Training time

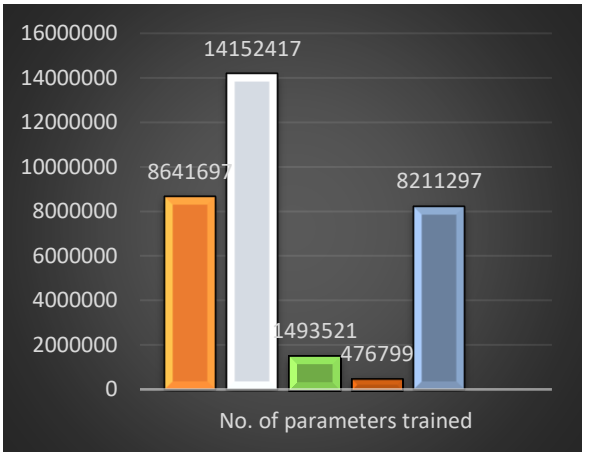


Figure 14 (b). Number of parameters trained by Models

In comprehensive analysis of semantic segmentation models after comparing their performance based on Dice Coefficient, Jaccard Similarity Index, and Sensitivity the results shows that UNET and Attention-UNet stand out with the highest Dice Coefficient values of 0.9175 and 0.9092, respectively, while UNET++ excels in Jaccard Similarity Index at 0.9132. Sensitivity is highest for UNET at 0.8673 and Attention-UNet at 0.8121. In terms of training parameters, UNET has the shortest training time at 2 hours and 32 minutes, and ARUX is the most lightweight with 476,799 parameters, while ResUNET requires the longest training time and has the most parameters at 14,152,417. These findings highlight the trade-offs between segmentation accuracy, computational efficiency, and model complexity, emphasizing the importance of choosing a model that aligns with specific application requirements and resource constraints. Qualitative segmentation results are shown in figure 15 Input is CTPA image (slice) the and its corresponding ground truth (mask image). Segmentation output of UNET, ResUNET, UNET++, ARUX, Attention UNET model is checked and compared with ground truth image.

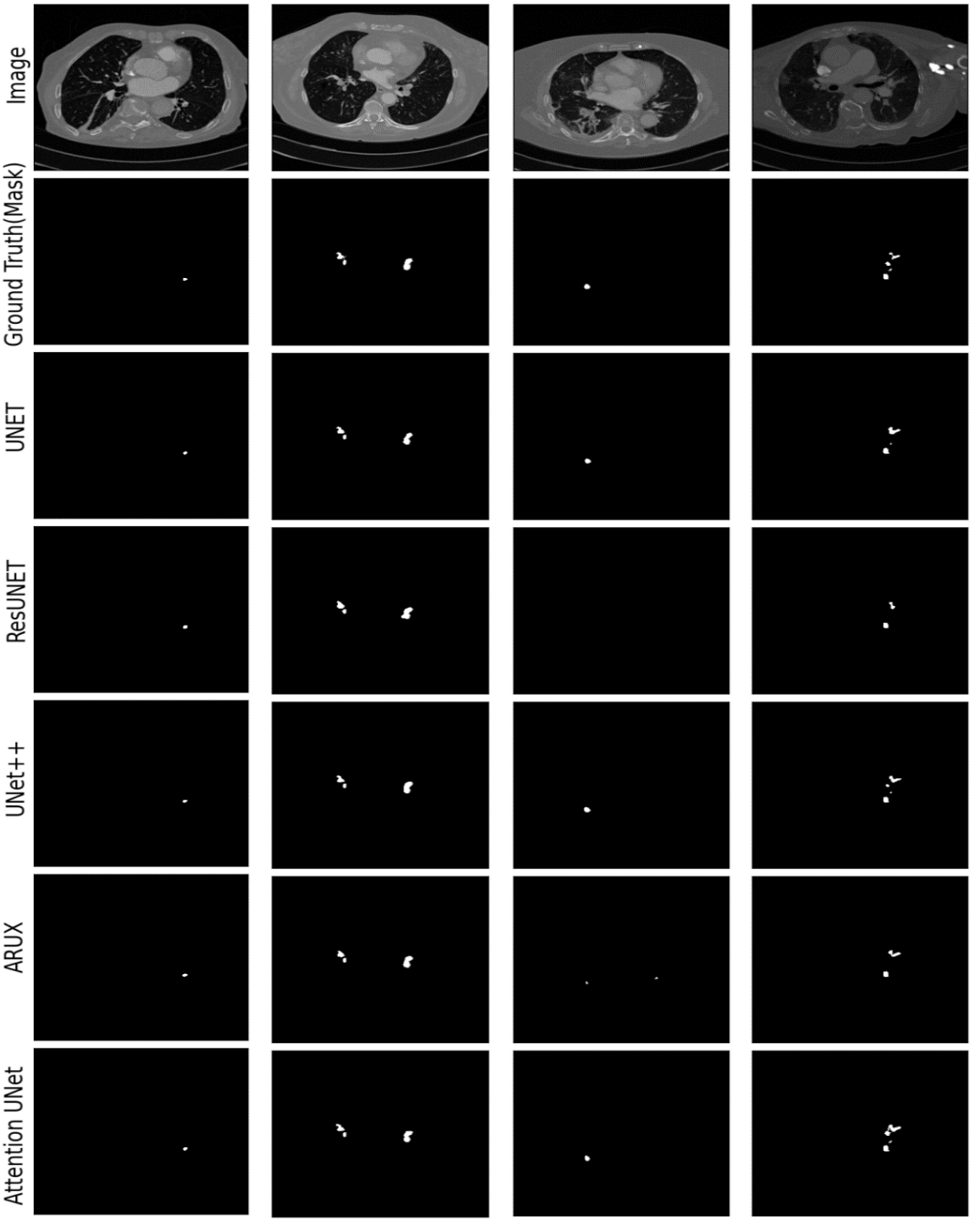


Figure 15. Segmentation results: The initial row displays the original image. Next row shows the labelled annotated image by an expert radiologist (ground truth), third row onwards shows the segmentation result (predicted mask) provide by different model.

5. Conclusion

In conclusion, the comparative analysis of semantic segmentation models for automatically segmenting Pulmonary Embolism (PE) in Computed Tomography Pulmonary Angiography (CTPA) images provides valuable insights for the development of efficient computer-assisted systems in healthcare. By facilitating prompt diagnosis and treatment of patients, these systems play a crucial role in improving clinical outcomes. Our study highlights nuanced trade-offs in performance and resource requirements among various deep learning architectures. UNET and Attention-UNET emerge as top performers, demonstrating high Dice Coefficients and sensitivities, while UNET++ excels in Jaccard Similarity Index. The selection of a model should be driven by specific application needs and computational constraints. UNET proves efficient with the shortest training time, whereas ARUX offers a lightweight option with minimal parameters. Despite longer training times and higher parameter counts, ResUNET may be preferred in scenarios prioritizing segmentation accuracy over computational efficiency. Our analysis underscores the importance of considering segmentation metrics alongside computational efficiency, training times, and model complexity to select the most suitable model for a given application. This comprehensive evaluation contributes to the advancement of computer-assisted systems, ultimately enhancing the diagnostic capabilities of healthcare professionals and benefiting patient care.

6. Future scope

Future work could involve proposing a novel model architecture that combines features from different models, enhancing segmentation performance for conditions like Pulmonary Embolism (PE). Incorporating image patches in training may improve result visualization. Evaluating the model with 2.5D and 3D data can provide a comprehensive performance assessment. The model's versatility extends to segmenting other diseases, such as pulmonary nodules, showcasing its potential for broader medical imaging applications. These suggestions aim to advance the current model, ensuring it remains adaptable and impactful in evolving medical imaging landscapes.

Declarations:
Ethical Approval
Not applicable
Funding
This research received no external funding.
Availability of data and material
The dataset used in this work is available on this link: https://ieee-dataport.org/open-access/cad-pe
Conflict of interest statement
No conflict of interest

References

1. Masotti, Luca, Marc Righini, Nicolas Vuilleumier, Fabio Antonelli, Giancarlo Landini, Roberto Cappelli, and Patrick Ray. "Prognostic stratification of acute pulmonary embolism: focus

on clinical aspects, imaging, and biomarkers." *Vascular Health and Risk Management*, pp. 567-575, 2009.

2. Rucco, Matteo, David Sousa-Rodrigues, Emanuela Merelli, Jeffrey H. Johnson, Lorenzo Falsetti, Cinzia Nitti, and Aldo Salvi. "Neural hypernetwork approach for pulmonary embolism diagnosis." *BMC research notes,* vol. 8, pp. 1-11, 2015.
3. Long, Kun, Lei Tang, Xiaorong Pu, Yazhou Ren, Mingxiu Zheng, Li Gao, Chunjiang Song, Su Han, Min Zhou, and Fengbin Deng. "Probability-based Mask R-CNN for pulmonary embolism detection." *Neurocomputing* 422, pp. 345-353, 2021
4. Rajan, Deepta, David Beymer, Shafiqul Abedin, and Ehsan Dehghan. "Pi-PE: a pipeline for pulmonary embolism detection using sparsely annotated 3D CT images." In *Machine Learning for Health Workshop* PMLR, pp. 220-232, 2020.
5. Friedman, Tamir, Ronald S. Winokur, Keith B. Quencer, and David C. Madoff. "Patient assessment: clinical presentation, imaging diagnosis, risk stratification, and the role of pulmonary embolism response team." In *Seminars in interventional radiology*, vol. 35, no. 02, pp. 116-121. Thieme Medical Publishers, 2018.
6. Zhang, Long Jiang, Guang Ming Lu, Felix G. Meinel, Andrew D. McQuiston, James G. Ravenel, and U. Joseph Schoepf. "Computed tomography of acute pulmonary embolism: state-of-the-art." *European radiology,* vol. 25, pp. 2547-2557, 2015.
7. Moore, Alastair JE, Jason Wachsmann, Murthy R. Chamarthy, Lloyd Panjikaran, Yuki Tanabe, and Prabhakar Rajiah. "Imaging of acute pulmonary embolism: an update." *Cardiovascular diagnosis and therapy,* vol. 8, no. 3, pp. 225, 2018.
8. Schoepf, U. Joseph. "Diagnosing pulmonary embolism: time to rewrite the textbooks." *The international journal of cardiovascular imaging,* vol. 21, pp. 155-163, 2005.
9. Huang, Shih-Cheng, Tanay Kothari, Imon Banerjee, Chris Chute, Robyn L. Ball, Norah Borus, Andrew Huang et al. "PENet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging." *NPJ digital medicine,* vol. 3, no. 1, pp. 61, 2020.
10. Serpen, Gürsel, D. K. Tekkedil, and M. Orra. "A knowledge-based artificial neural network classifier for pulmonary embolism diagnosis." *Computers in biology and medicine,* vol. 38, no. 2, pp. 204-220, 2008.
11. Hofsäß, Cornelia, Roman Johannes Gertz, Tanja Lossau, Jens-Peter M. Zemke, Tobias Klinder, Alexander C. Bunck, and Hannes Nickisch. "Pulmonary Embolus Detection with Dual-Energy CT Data Augmentation." In *Medical Imaging with Deep Learning*, 2022.
12. Liang, Jianming, and Jinbo Bi. "Computer aided detection of pulmonary embolism with tobogganing and mutiple instance classification in CT pulmonary angiography." In *Biennial International Conference on Information Processing in Medical Imaging*, pp. 630-641, 2007.
13. Tajbakhsh, Nima, Michael B. Gotway, and Jianming Liang. "Computer-aided pulmonary

embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks." In *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18*, pp. 62-69. Springer International Publishing, 2015.

14. Özkan, Haydar, Gökalp Tulum, Onur Osman, and Sinan Şahin. "Automatic detection of pulmonary embolism in CTA images using machine learning." *Elektronika Ir Elektrotechnika*, 2017.
15. Yang, Xin, Yi Lin, Jianchao Su, Xiang Wang, Xiang Li, Jingen Lin, and Kwang-Ting Cheng. "A two-stage convolutional neural network for pulmonary embolism detection from CTPA images." *IEEE Access,* vol. 7 pp. 84849-84857, 2019.
16. Liu, Weifang, Min Liu, Xiaojuan Guo, Peiyao Zhang, Ling Zhang, Rongguo Zhang, Han Kang et al. "Evaluation of acute pulmonary embolism and clot burden on CTPA with deep learning." *European radiology,* vol. 30, pp. 3567-3575, 2020.
17. Cano-Espinosa, Carlos, Miguel Cazorla, and Germán González. "Computer aided detection of pulmonary embolism using multi-slice multi-axial segmentation." *Applied Sciences,* vol.10, no. 8, pp. 2945, 2020.
18. Cheng, Ting-Wei, Yi Wei Chua, Ching-Chun Huang, Jerry Chang, Chin Kuo, and Yun-Chien Cheng. "Feature-enhanced adversarial semi-supervised semantic segmentation network for pulmonary embolism annotation." *Heliyon*, vol. 9, no. 5, 2023.
19. Shi, Luyao, Deepta Rajan, Shafiq Abedin, Manikanta Srikar Yellapragada, David Beymer, and Ehsan Dehghan. "Automatic diagnosis of pulmonary embolism using an attention-guided framework: A large-scale study." In *Medical Imaging with Deep Learning*, pp.743-754, 2020.
20. Suman, Sudhir, Gagandeep Singh, Nicole Sakla, Rishabh Gattu, Jeremy Green, Tej Phatak, Dimitris Samaras, and Prateek Prasanna. "Attention based CNN-LSTM network for pulmonary embolism prediction on chest computed tomography pulmonary angiograms." In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24*, pp. 356-366. 2021.
21. Liu, Zhenhong, and Hongfang Yuan. "An Res-Unet method for pulmonary artery segmentation of CT images." In *Journal of Physics: Conference Series*, vol. 1924, no. 1, pp. 012018, 2021.
22. Yuan, Hongfang, Zhenhong Liu, Yajun Shao, and Min Liu. "ResD-Unet research and application for pulmonary artery segmentation." *IEEE Access,* vol. 9, pp. 67504-67511, 2021.
23. Huhtanen, Heidi, Mikko Nyman, Tarek Mohsen, Arho Virkki, Antti Karlsson, and Jussi Hirvonen. "Automated detection of pulmonary embolism from CT-angiograms using deep learning." *BMC Medical Imaging,* vol. 22, no. 1, pp. 43, 2022.
24. Yu, Ching-Yuan, Ming-Che Chang, Yun-Chien Cheng, and Chin Kuo. "Convolutional Neural Network for Early Pulmonary Embolism Detection via Computed Tomography Pulmonary Angiography." *arXiv preprint arXiv:2204.03204,*

2022.

25. Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. "Unet++: A nested u-net architecture for medical image segmentation." In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pp. 3-11, 2018.
26. Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori et al. "Attention u-net: Learning where to look for the pancreas." *arXiv preprint arXiv:1804.03999,* 2018.
27. Engelke, Christoph, Stephan Schmidt, Annemarie Bakai, Florian Auer, and Katharina Marten. "Computer-assisted detection of pulmonary embolism: performance evaluation in consensus with experienced and inexperienced chest radiologists." *European radiology* 18 (2008): 298-307.
28. González, Germán, Daniel Jimenez-Carretero, Sara Rodríguez-López, Carlos Cano-Espinosa, Miguel Cazorla, Tanya Agarwal, Vinit Agarwal et al. "Computer aided detection for pulmonary embolism challenge (CAD-PE)." *arXiv preprint arXiv:2003.13440,* 2020.
29. Liu, Xiangbin, Liping Song, Shuai Liu, and Yudong Zhang. "A review of deep-learning-based medical image segmentation methods." *Sustainability,* vol. 13, no. 3, pp. 1224, 2021.
30. Nai, Ying-Hwey and Teo, Bernice W and Tan, Nadya L and O'Doherty, Sophie and Stephenson et al*." Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset "* Computers in Biology and Medicine vol. 134, pp. 104497, 2021
31. Gonzalez, G. CAD-PE Challenge Website. Available online: https://ieee-dataport.org/open-access/cad-pe