

# Detecção de Tromboembolia Pulmonar utilizando Redes Neurais Convolucionais e Extração de Características

Gabriel Olescki<sup>1</sup>, João Mario Clementin de Andrade<sup>2</sup>, Dante Escuissato<sup>2</sup>,  
Lucas Ferrari de Oliveira<sup>1</sup>

<sup>1</sup>Departamento de Informática (DInf) Universidade Federal do Paraná (UFPR)  
Curitiba, PR – Brasil

<sup>2</sup>Departamento de Clínica Médica Universidade Federal do Paraná (UFPR)  
Curitiba, PR – Brasil

{golescki,lferrari}@inf.ufpr.br, joaoclementin@gmail.com,  
dante.escuissato@gmail.com

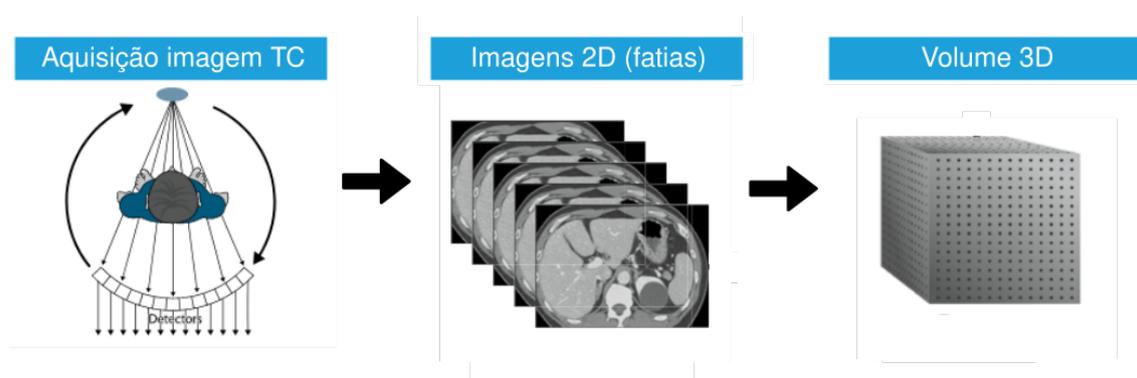
**Abstract.** *Pulmonary embolism is among the leading causes of death all over the world, once the diagnosis is made a quick response is required from the medical staff in order to save the patient. The main form of pulmonary embolism diagnosis is computed tomography, due to its fast acquisition rate and precision. With the high amount of data produced by computed tomographies, deep learning algorithms had showed good results in finding embolisms in such exams. This paper proposes a deep learning based technique using the state of the art networks aimed to find pulmonary embolisms in computed tomography exams. Using a u-net inspired model for segmentation of pulmonary embolisms candidates we achieved a Dice of 0.81 and IoU of 0.79.*

**Resumo.** *Embolia pulmonar é uma das principais causas de morte relacionadas a doenças cardiovasculares no mundo, uma vez que é feito o diagnóstico é necessária uma resposta rápida pela equipe médica para salvar o paciente. A principal forma de diagnóstico é pelo exame de tomografia computadorizada e, devido a grande quantidade de dados que o exame gera, algoritmos de deep learning têm mostrado bons resultados em encontrar embolia pulmonar de maneira autônoma. O objetivo deste trabalho é desenvolver uma rede de deep learning capaz de encontrar embolia pulmonar em exames de tomografia computadorizada. Até então, utilizando uma rede inspirada na U-net, o método segmentou trombos atingindo um Dice Score de 0.81 e um IoU de 0.79.*

## 1. Introdução

Tromboembolia pulmonar (TEP) trata-se de uma condição clínica na qual um trombo proveniente da circulação venosa profunda se desloca para a circulação arterial pulmonar. Esse trombo pode comprometer em diferentes graus o fluxo de sangue dentro do pulmão, sendo fatal caso não haja tratamento [Carson et al. 1992]. O TEP apresenta uma alta incidência na população, com uma estimativa de 200 mil casos por ano nos EUA [White 2003] e uma taxa de mortalidade de 30% [Calder et al. 2005], podendo ser reduzida para até 2% com um diagnóstico precoce [Sadigh et al. 2011].

Para chegar a um diagnóstico rápido, o método de escolha é a tomografia computadorizada (TC), por sua velocidade de aquisição e uma alta confiabilidade [Goldhaber and Bounameaux 2012, Kligerman et al. 2018]. Um exame de TC para TEP produz uma imagem tridimensional do tórax do paciente, composto por centenas de imagens (fatias) em 2-D que concatenadas formam o volume completo (figura 1). Para fazer a análise do exame, o médico deve verificar cuidadosamente todas as centenas de imagens a fim de encontrar um trombo, tarefa não trivial pela irregularidade em formatos dos TEPs e a sua aparição em diferentes áreas [Rajan et al. 2020]. Pela dificuldade do diagnóstico e fatores como fadiga e inexperiência há uma alta taxa de erro de diagnóstico de TEP, havendo uma variação na taxa de 13% no acerto de diagnóstico entre médicos que trabalham no turno noturno e diurno [Huang et al. 2020].



**Figura 1. Exemplo de aquisição de um exame de TC, suas fatias 2D e o volume final 3D. Fonte: Adaptado de [Seeram 2018]**

Por haver essa taxa de erro no diagnóstico, sistemas computacionais de auxílio ao diagnóstico (CADs) se tornam valiosos para a detecção de TEP. Pelo alto número de dados produzidos por exame, sistemas baseado em aprendizagem de dados já são capazes de encontrar a melhor taxa de acerto no diagnóstico de TEP [Tajbakhsh et al. 2019].

O objetivo deste trabalho foi desenvolver um método capaz de encontrar TEPs em imagens de TC, para isso foi desenvolvido um pipeline de redes de *deep learning* em duas etapas. A primeira etapa para a segmentação dos TEPs candidatos, utilizando uma rede baseada na U-net [Ronneberger et al. 2015]. Na segunda etapa foi utilizada uma rede capaz de classificar as imagens produzidas na primeira etapa entre TEP ou não-TEP, reduzindo o número de falsos positivos.

## 2. Trabalhos relacionados

Os sistemas CADs já existentes seguem principalmente os mesmos passos para realizar a detecção de TEPs: 1) Segmentação do pulmão e artéria pulmonar, 2) Criação de um conjunto de TEPs candidatos, 3) Extração de características dos TEPs candidatos e 4) Remoção de falsos positivos desse conjunto [Tajbakhsh et al. 2015], o último passo sendo normalmente realizado utilizando técnicas de aprendizado supervisionado.

No trabalho de [Bouma et al. 2009] os autores utilizam uma segmentação usando *thresholding* dos canais de cinza e matriz Hessiana para gerar TEPs candidatos, que são classificados com árvores de decisão, o trabalho reportou uma *recall* de 58% e uma taxa de 15 falsos positivos por *scan* (FP/s).

Já em [Tajbakhsh et al. 2015] os TEPs candidatos são encontrados utilizando o algoritmo segmentação *tobogganing* para encontrar candidatos e os classifica com uma rede CNN, o trabalho usou uma representação dos TEPs candidatos em 2.5D utilizando autovetores resumindo o componente encontrado pelo algoritmo de segmentação, foi testado em duas bases de dados, uma privada e outra pública, atingindo *recall* de 83% e 2 FP/s na base privada e 34% de *recall* e 2 FP/s na pública.

O trabalho de [Yang et al. 2019] foi o primeiro a utilizar redes U-net com convoluções 3D para encontrar TEPs candidatos, que depois são classificados com uma rede Resnet-18, também testou a rede em duas bases de dados, uma privada e outra pública, obteve um *recall* de 75,4% e 2 FP/s na base pública e *recall* de 76,3% com 2 FP/s na base privada.

Em [Rajan et al. 2020] os autores utilizaram uma U-net com convoluções 2D utilizando a vizinhança das fatias como os diferentes canais de entrada, utilizando uma imagem com 9 canais, sendo cada canal as fatias vizinhas de um pedaço do volume da TC, obteve uma área sob a curva ROC (AUCROC) de 0,94.

E em [Huang et al. 2020] utiliza uma rede própria, apelidada de PENet, baseando-se em DenseNets e U-nets para classificar se um exame contém ou não TEP, a rede também usou de *transfer learning* durante o treinamento, utilizando os pesos pré-treinados com a base *Kinects-600*, obteve uma AUROC de 0,85.

Em [Cano-Espinosa et al. 2020] foi utilizado três variações da rede U-net, uma utilizando convoluções 2D, outra que também utiliza convoluções 2D porém com imagens de entrada em 3D e a última que utiliza convoluções 3D, sendo o resultado obtido melhor na rede 3D, foi utilizado o valor da saída da rede no *voxel* mais central de cada TEP para definir se era um verdadeiro positivo, obteve um *recall* de 55% e 1 FP/s, para o treinamento e teste foi utilizada a base de dados [González et al. 2020].

### 3. Materiais e Métodos

#### 3.1. Dados

Os dados utilizados para treinar e avaliar o modelo proposto vêm de três bases de dados, sendo duas bases públicas e uma privada. Todas as bases de dados possuem marcações semânticas dos TEPs. A Tabela 1 resume as três bases de dados, mostrando a contagem de TEPs em intervalos com base nos tamanhos em mm<sup>3</sup>.

A primeira base de dados, chamada de “FUMPE” [Masoudi et al. 2018], composta por 35 exames de TC com contraste (angiogramografia), sendo que dois exames não possuem marcações de TEP. As máscaras de segmentação foram produzidas por um radiologista com 5 anos de experiência na área e revisadas por outro com 18 anos de experiência. A base possui um total de 8792 fatias, e 3438 regiões de interesse de TEP, com a altura e a largura dos *voxels* variando entre 0,52 e 0,78 mm. A sua maioria (67%) são encontradas em artérias periféricas. Além das marcações, os autores também disponibilizaram metadados, como a proporção do tamanho do ventrículo direito e ventrículo esquerdo do coração, se houve refluxo na veia cava inferior, se o septo intraventricular está reto, o diâmetro da artéria pulmonar e o valor de *Qanadli score*, que é calculado baseado na posição do trombo dentro do pulmão.

**Tabela 1. Contagem de TEPs de cada base de dados dentro de intervalos em  $\text{mm}^3$ .**

Intervalos ( $\text{mm}^3$ )	FUMPE	CAD-PE	Privado
$0, 1 \leq n < 0, 5$	0	74	35
$0, 5 \leq n < 1$	1	117	23
$1 \leq n < 2$	0	93	26
$2 \leq n < 9$	3	96	69
$9 \leq n < 38$	8	68	51
$38 \leq n < 168$	13	157	71
$168 \leq n < 740$	31	155	74
$740 \leq n < 3270$	32	80	25
$3270 \leq n < 14420$	19	54	29
$14420 \leq n \leq 63650$	11	27	5

A segunda base de dados, chamada de “CAD-PE” [González et al. 2020], é composta de 91 exames de angiotomografias e um total de 24624 fatias com tamanho dos *voxels* variando entre 0,52 e 0,92 mm. Os 40 primeiros exames foram segmentados por um conselho de radiologistas com mais de 10 anos de experiência enquanto os outros 51 exames foram segmentados por um membro deste conselho. A marcação foi feita com o auxílio de métodos semi-automatizados, nos quais as marcações realizadas pelos radiologistas eram da área de interesse (i.e. um retângulo na região do TEP), e com esta área de interesse utilizou-se um método semi-supervisionado para gerar as máscaras de segmentação semântica, foi feita então a média de cada segmentação realizada para produzir a máscara final.

A terceira base de dados é privada, composta de 40 exames, totalizando 14463 fatias, sendo 20 produzidas pelo tomógrafo GE Revolution 512 e outras 20 do Toshiba Aquilion 64. Os tamanhos dos *voxels* variam de 0,57 e 1,0 mm. Todos os exames foram avaliados por um radiologista torácico com 32 anos de experiência quanto a presença de TEP e a sua localização. Após esta revisão as máscaras foram criadas por um médico do terceiro ano da residência de radiologia. Não foi utilizado nenhum método semi-automatizado para criar as máscaras de segmentação. A base de dados foi criada em parceria com o Hospital das Clínicas (HC) da UFPR, a base de dados foi aprovada pelo comitê de ética do HC com o CAAE: 07724919.0.0000.0096.

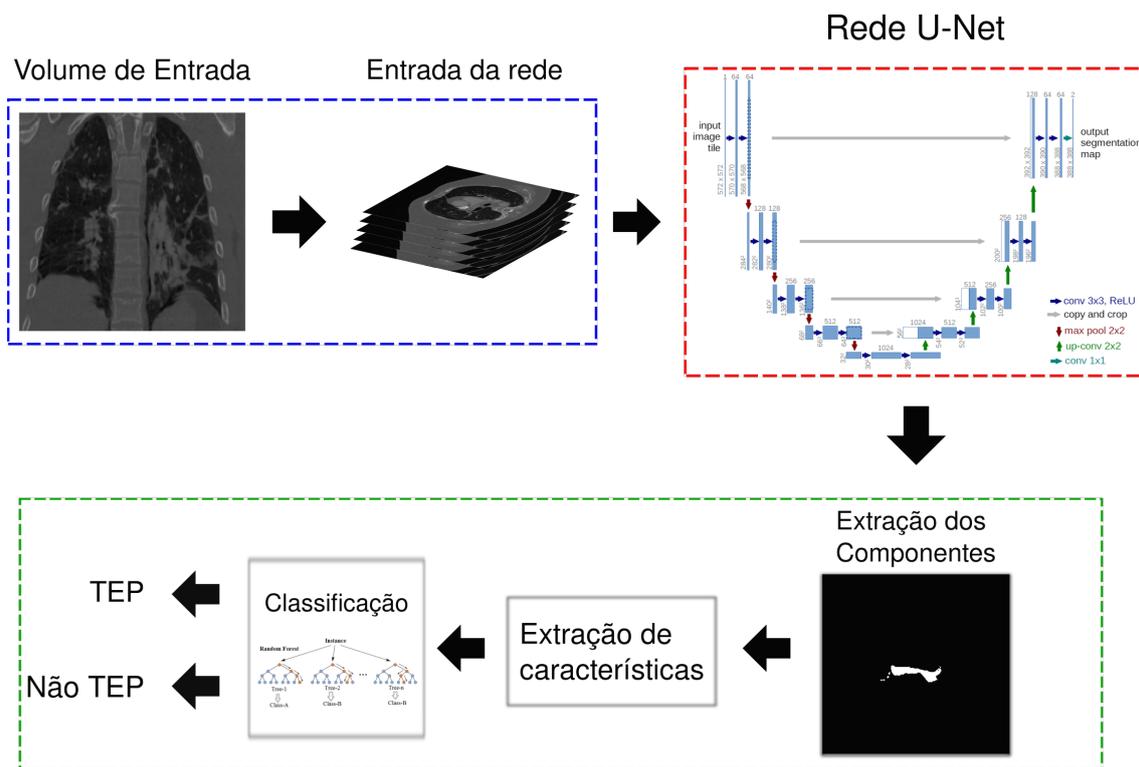
### 3.2. Método Proposto

A técnica proposta neste trabalho é composta de duas etapas principais: a localização e segmentação de TEPs candidatos e a redução de falsos positivos desses grupos candidatos (Figura 2).

Para auxiliar o treinamento do método proposto, foram aplicados os mesmos métodos de pré-processamento nas três bases de dados testadas. Primeiro foi alterado o tamanho da largura das fatias para 1 mm para todos os pacientes a fim de padronizar as entradas. Em seguida foram normalizados os valores de escala de cinza das imagens utilizando janelamento da escala de Unidades *Hounsfield* (UH), com o tamanho de janela de 350 UH e o centro da janela de 150 UH, transformando os valores entre  $[-200, 500]$ , em seguida os valores foram normalizados entre 0 e 1 visando melhorar o aprendizado

e taxa de convergência da rede. Foi realizada uma etapa de *data augmentation* na base de treino, com objetivo de aumentar a quantidade de imagens para a um melhor treino da rede, foi aplicado *flips* horizontais e verticais com probabilidade de 50% e ruído gaussiano com probabilidade de 20%. Para cada imagem também foi recortada uma janela central no tamanho  $364 \times 364$ , visando remover regiões ao redor da caixa torácica das imagens.

A rede neural de segmentação foi baseada na rede U-net de Ronnenberg (Figura 2 no destaque vermelho). A rede é dividida em duas partes, *encoder* e *decoder*. O *encoder* é construída a partir de blocos, cada bloco é composto de duas convoluções seguidas por normalização por *batch* e a função de ativação *ReLU*, todo o *encoder* é composto de cinco blocos com o número de canais na operação de convolução de cada bloco de 64, 64, 128, 256 e 512 respectivamente. Já o *decoder* possui blocos com duas operações de deconvolução, com a sequência de canais inversa a parte *encoder* (i.e. 512, 256, 128, 64 e 64), cada bloco da parte *decoder* recebe como entrada a saída do último bloco concatenado com o bloco *encoder* correspondente. Na saída do último bloco *decoder*, uma convolução 2D é aplicada com um único canal de saída, que passa por uma função sigmoide, produzindo uma saída com valores entre 0 e 1.



**Figura 2.** Representação do passo-a-passo do método proposto, em azul as etapas de pré-processamento e extração das fatias para a entrada da rede, em vermelho a representação da rede U-net adaptado de [Ronneberger et al. 2015], em verde as etapas de extração dos TEPs candidatos (componentes), extração das características e classificação.

Diferente da rede U-net original, a rede proposta recebe como entrada uma imagem de tamanho  $364 \times 364 \times 5$  e gera uma saída  $364 \times 364$ , ou seja, são utilizadas convoluções 2D para entradas representando volumes em 3D. Os cinco canais de entrada são as cinco fatias vizinhas no volume de CT, porém a saída da rede identifica apenas a

máscara do *slice* central (canal 3) das imagens de entrada.

Para encontrar os TEPs candidatos, os valores de saída da rede foram arredondados para 0 (valores  $< 0.5$ ) e 1 (valores  $\geq 0.5$ ), depois foi aplicado um algoritmo de componentes conexos (*connected-component labeling*, ou CCL), cada componente encontrado pelo CCL foi considerado como TEP candidato. Para a redução de falsos positivos, foi realizada uma operação de fechamento com um *kernel* de  $5 \times 3 \times 3$ , visando remover partículas pequenas e ruídos, candidatos encontrados em apenas uma fatia do exame também foram removidos. Foram considerados verdadeiros positivos os componentes que obtiveram uma interseção com algum componente da máscara verdadeira, os componentes sem nenhuma interseção foram considerados falsos positivos.

Após as etapas de pós-processamento, foi feita a extração de características de cada componente. Valores como o valor médio dos *voxels* centrais (escala UH), os valores centrais do componente da matriz de saída da rede U-net, a suas coordenadas (x, y e z), a maior distância do componente (nos três eixos) e seu volume ( $\text{mm}^3$ ), foram extraídos e totalizaram nove características. Para fazer a classificação entre candidatos TEPs e não-TEPs foram testados os classificadores KNN, SVM, e *Random Forest*, os hiper-parâmetros de cada classificador foram encontrados por meio de busca gulosa com validação cruzada de 5 *folds*.

### 3.3. Resultados

Para a implementação da rede foi utilizada a biblioteca PyTorch [Paszke et al. 2019]. Todas as bases de dados foram separadas em 70% para treino, 10% para validação e 20% para teste, as três bases foram treinadas e testadas separadamente.

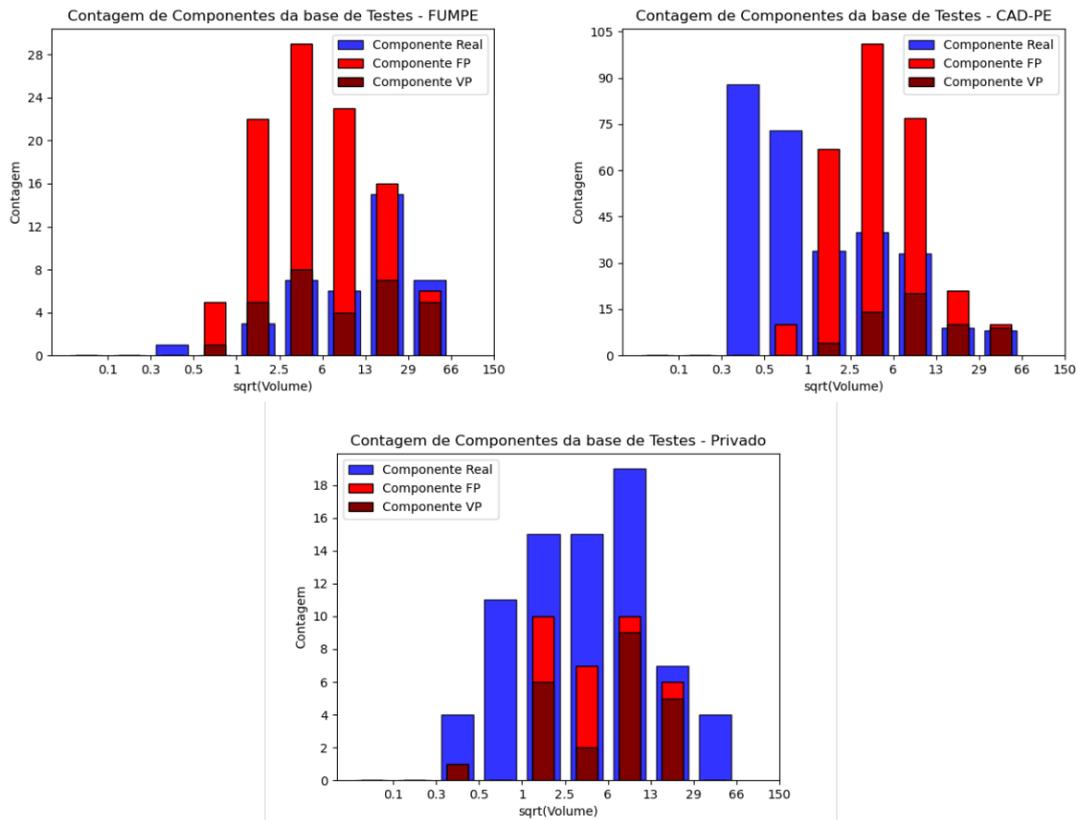
Para o treinamento da U-net foi utilizado o *optimizer* Adam, com *learning rate* de  $1e - 4$ , e a *loss* escolhida foi a Dice, a IoU também observada considerando valores acima de 0,5 como TEP para seu cálculo. O valor do *learning rate* foi reduzido em mais  $1e - 2$  caso o resultado na base de validação atingisse um platô.

Nos resultados da segmentação, nas base FUMPE a rede atingiu um IoU médio de 0,70, um Dice score de 0,72 e Recall de 0,82, na base CAD-PE o IoU médio foi de 0,79 e Dice score de 0,81 e Recall de 0,89 e na base privada, o IoU foi de 0,75, Dice de 0,78 e Recall de 0,78 (Tabela 2). Quanto a segunda parte do método, o classificador que apresentou melhor resultado foi o *Random Forest*, obtendo uma acurácia de 83% na base CAD-PE, 74% na base FUMPE e 79% na base Privada.

**Tabela 2. Resultados obtidos pela primeira rede de segmentação U-net**

Dataset	Dice	IoU	Recall
FUMPE	0,72	0,68	0,82
CAD-PE	0,79	0,81	0,89
Privado	0,76	0,77	0,79

Mesmo com as máscaras produzindo um valor de Recall acima de 0,75, foi feita a análise por tamanho dos componentes encontrados. Nessa avaliação foi identificado que a rede apresenta dificuldades para encontrar, principalmente, componentes pequenos, produzindo também altas taxas de falso positivo em componentes de tamanhos médios (Figura 3). Observa-se também que em alguns casos o número de verdadeiros positivos de



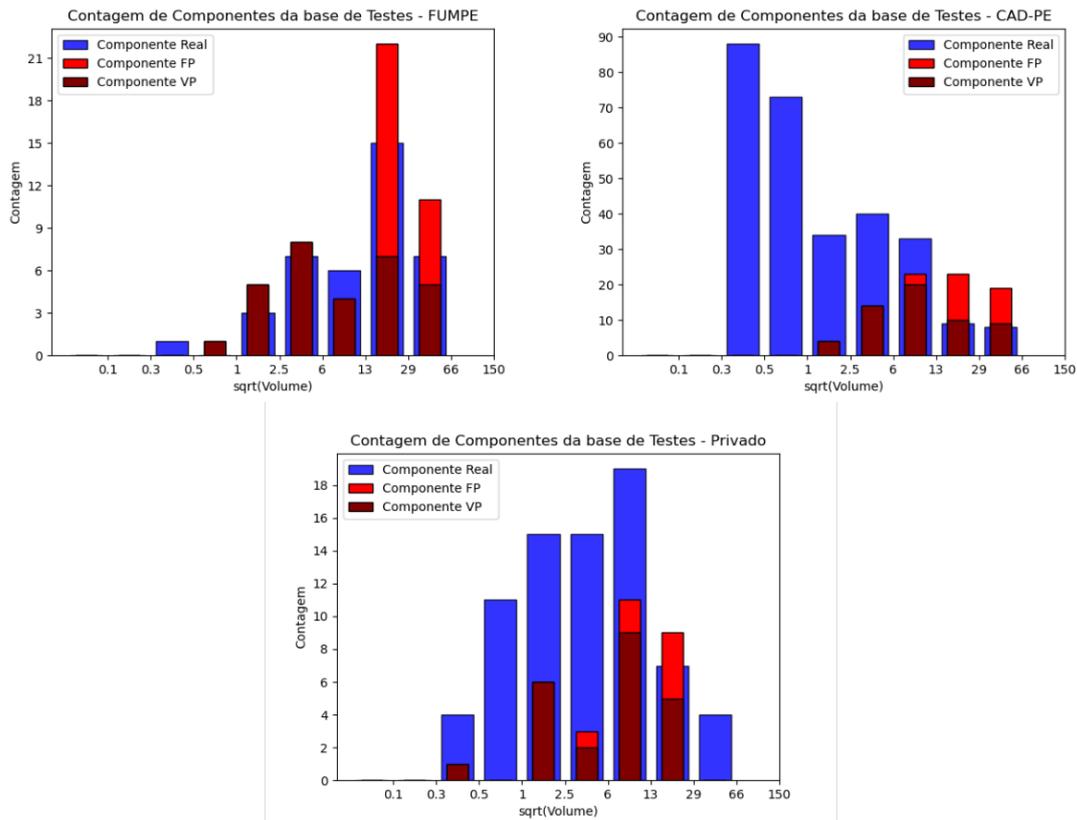
**Figura 3. Gráficos mostrando a contagem (eixo y) da raiz quadrada dos volumes (eixo x) dos componentes (TEPs) encontrados na base de teste das três bases de dados, em azul os TEPs verdadeiros (i.e. marcados pelo especialista), em vermelho escuro os TEPs verdadeiros positivos (VP) segmentados pela rede U-net e em amarelo os TEPs falsos positivos (FP) produzidos pela rede.**

um intervalo de tamanho está maior do que a quantidade de componentes reais, isso se da pelo fato do método considerar os componentes como VP caso haja alguma interseção do candidato com algum componente real, então há casos onde os TEPs foram encontrados porém não foram marcados por completo, gerando uma predição de menor volume.

A base de imagens se apresentou como um outro fator que interfere diretamente na quantidade de falsos positivos encontrados e também em quais faixas de volume esses falsos positivos se encontram. Acredita-se que esse comportamento deve-se a quantidade de elementos existentes em cada faixa de tamanho. Pela figura 4, observa-se a eficácia do modelo em reduzir os componentes falsos positivos das três bases, principalmente os componentes de menor volume, porém erra ao classificar os componentes maiores, classificando componentes verdadeiros como falsos positivos entre os maiores volumes.

#### 4. Conclusão

O método proposto, apesar de ser baseado nos outros métodos do estado da arte, ainda possui algumas limitações. A quantidade de falsos negativos, ou seja, trombos em que a rede não foi capaz de encontrar, está elevada principalmente dos TEPs de menor tamanho, mostrando que a rede não é capaz de distinguir componentes pequenos de outros ruídos ou

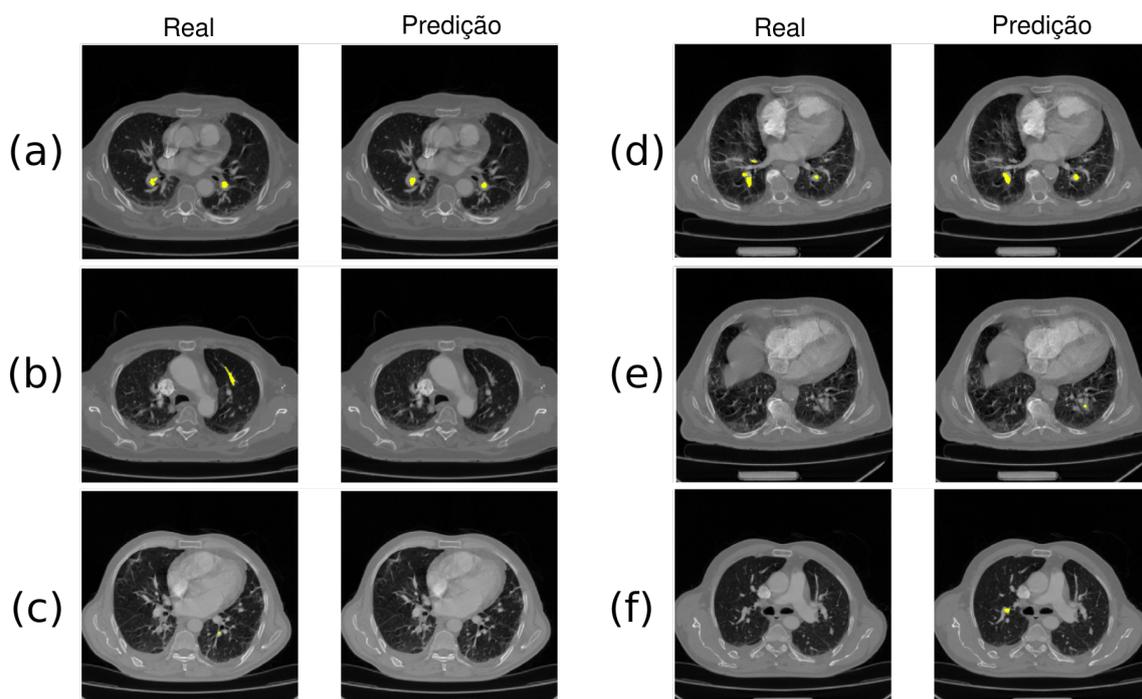


**Figura 4. Gráficos mostrando a contagem (eixo y) da raiz quadrada dos volumes (eixo x) dos TEPs encontrados na base de teste das três bases de dados, porém aplicando o classificador de redução de falsos positivos, em azul os TEPs verdadeiros (i.e. marcados pelo especialista), em vermelho escuro os TEPs verdadeiros positivos (VP) segmentados pela rede U-net e em amarelo os TEPs falsos positivos (FP) produzidos pela rede.**

artefatos. Outra limitação observada é a taxa de falsos positivos, que apesar de mitigada pela etapa de redução, ainda é presente.

O tamanho e a quantidade de trombos da base de dados influencia diretamente a capacidade do modelo. Acredita-se ser possível melhorar os resultados da primeira etapa do método com a rede U-net utilizando mais testes de diferentes hiper-parâmetros. A rede provou ser capaz de distinguir TEPs dentro dos exames mesmo produzindo alguns falsos positivos, o modelo mostrou também ter dificuldades em detectar trombos menores e mais periféricos (Figura 5).

Visando melhorar a acurácia do modelo em encontrar componentes pequenos, pode-se explorar outros hiper-parâmetros da rede relacionados a convolução e redução dos canais, por exemplo blocos de *squeeze-and-excitation* como no trabalho de [Huang et al. 2020], realizar o treinamento do modelo com combinações diferentes das bases de dados pode também aumentar a taxa de componentes encontrados. Modelos com operações de convolução em 3D também mostram um aumento na taxa de segmentação da rede como em [Cano-Espinosa et al. 2020] e [Yang et al. 2019]. Outras redes de segmentação podem também servir como alternativa, como a Mask R-CNN. No



**Figura 5. Comparação entre a máscara gerada (colunas Real) e as máscaras geradas pela rede (colunas Predição), em (a) um exemplo onde a rede encontrou todos os TEPs da fatia, em (b) um exemplo onde a rede não foi capaz de encontrar a máscara, em (c) e (d) a rede não encontrou um TEP pequeno, já em (e) e (f) mostram exemplos onde a rede marcou posições onde não há TEPs nas máscaras originais (falsos positivos)**

trabalho de [Long et al. 2021] o autor utiliza uma variação da Mask R-CNN, chamada de P Mask R-CNN, que é otimizada para encontrar objetos pequenos.

Quanto ao número de falsos positivos, abordagens como a de [Tajbakhsh et al. 2015] podem aumentar a taxa de acerto dessa etapa. Nele o autor produziu uma projeção única do TEP, chamada de 2.5D, utilizando auto-valores e auto-vetores gerou-se uma imagem com dois canais, um longitudinal e outro transversal, de cada trombo, explicado também em [Tajbakhsh et al. 2019]. Com esta representação é possível classificar os trombos com outras redes de classificação de imagens, facilitando o treinamento com a possibilidade de utilizar pesos pré-treinados, e.g. redes como a InceptionNet [Szegedy et al. 2017], ResNext [Xie et al. 2016], EfficientNet [Tan and Le 2019] e AmoebaNet [Real et al. 2019], todas possuindo uma taxa de acerto acima de 90% na base de dados da Imagenet [Deng et al. 2009].

## Referências

- Bouma, H., Sonnemans, J. J., Vilanova, A., and Gerritsen, F. A. (2009). Automatic detection of pulmonary embolism in cta images. *IEEE transactions on medical imaging*, 28(8):1223–1230.
- Calder, K. K., Herbert, M., and Henderson, S. O. (2005). The mortality of untreated pulmonary embolism in emergency department patients. *Annals of emergency medicine*, 45(3):302–310.

- Cano-Espinosa, C., Cazorla, M., and González, G. (2020). Computer aided detection of pulmonary embolism using multi-slice multi-axial segmentation. *Applied Sciences*, 10(8):2945.
- Carson, J. L., Kelley, M. A., Duff, A., Weg, J. G., Fulkerson, W. J., Palevsky, H. I., Schwartz, J. S., Thompson, B. T., Popovich Jr, J., Hobbins, T. E., et al. (1992). The clinical course of pulmonary embolism. *New England Journal of Medicine*, 326(19):1240–1245.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- Goldhaber, S. Z. and Bounameaux, H. (2012). Pulmonary embolism and deep vein thrombosis. *The Lancet*, 379(9828):1835–1846.
- González, G., Jimenez-Carretero, D., Rodríguez-López, S., Cano-Espinosa, C., Cazorla, M., Agarwal, T., Agarwal, V., Tajbakhsh, N., Gotway, M. B., Liang, J., et al. (2020). Computer aided detection for pulmonary embolism challenge (cad-pe). *arXiv preprint arXiv:2003.13440*.
- Huang, S.-C., Kothari, T., Banerjee, I., Chute, C., Ball, R. L., Borus, N., Huang, A., Patel, B. N., Rajpurkar, P., Irvin, J., et al. (2020). Penet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric ct imaging. *npj Digital Medicine*, 3(1):1–9.
- Kligerman, S. J., Mitchell, J. W., Sechrist, J. W., Meeks, A. K., Galvin, J. R., and White, C. S. (2018). Radiologist performance in the detection of pulmonary embolism. *Journal of thoracic imaging*, 33(6):350–357.
- Long, K., Tang, L., Pu, X., Ren, Y., Zheng, M., Gao, L., Song, C., Han, S., Zhou, M., and Deng, F. (2021). Probability-based mask r-cnn for pulmonary embolism detection. *Neurocomputing*, 422:345–353.
- Masoudi, M., Pourreza, H.-R., Saadatmand-Tarzjan, M., Eftekhari, N., Zargar, F. S., and Rad, M. P. (2018). A new dataset of computed-tomography angiography images for computer-aided detection of pulmonary embolism. *Scientific data*, 5:180180.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.
- Rajan, D., Beymer, D., Abedin, S., and Dehghan, E. (2020). Pi-pe: A pipeline for pulmonary embolism detection using sparsely annotated 3d ct images. In *Machine Learning for Health Workshop*, pages 220–232.
- Real, E., Aggarwal, A., Huang, Y., and Le, Q. V. (2019). Regularized evolution for image classifier architecture search. In *Proceedings of the aaai conference on artificial intelligence*, volume 33, pages 4780–4789.

- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Sadigh, G., Kelly, A. M., and Cronin, P. (2011). Challenges, controversies, and hot topics in pulmonary embolism imaging. *American Journal of Roentgenology*, 196(3):497–515.
- Seeram, E. (2018). Computed tomography: A technical review. *Radiologic technology*, 89(3):279CT–302CT.
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.
- Tajbakhsh, N., Gotway, M. B., and Liang, J. (2015). Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 62–69. Springer.
- Tajbakhsh, N., Shin, J. Y., Gotway, M. B., and Liang, J. (2019). Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation. *Medical image analysis*, 58:101541.
- Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.
- White, R. H. (2003). The epidemiology of venous thromboembolism. *Circulation*, 107(23\_suppl\_1):I–4.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2016). Aggregated residual transformations for deep neural networks. *arXiv preprint arXiv:1611.05431*.
- Yang, X., Lin, Y., Su, J., Wang, X., Li, X., Lin, J., and Cheng, K.-T. (2019). A two-stage convolutional neural network for pulmonary embolism detection from ctpa images. *IEEE Access*, 7:84849–84857.