

Attention based CNN-LSTM Network for Pulmonary Embolism Prediction on Chest Computed Tomography Pulmonary Angiograms

Sudhir Suman¹, Gagandeep Singh², Nicole Sakla², Rishabh Gattu², Jeremy Green², Tej Phatak², Dimitris Samaras³, and Prateek Prasanna⁴

¹ Dept. of Electrical Engineering, Indian Institute of Technology, Bombay, India

² Dept. of Radiology, Newark Beth Israel Medical Center, NJ, USA

³ Dept. of Computer Science, Stony Brook University, NY, USA

⁴ Dept. of Biomedical Informatics, Stony Brook University, NY, USA

prateek.prasanna@stonybrook.edu

Abstract. With more than 60,000 deaths annually in the United States, Pulmonary Embolism (PE) is among the most fatal cardiovascular diseases. It is caused by an artery blockage in the lung; confirming its presence is time-consuming and is prone to over-diagnosis. The utilization of automated PE detection systems is critical for diagnostic accuracy and efficiency. In this study we propose a two-stage attention-based CNN-LSTM network for predicting PE, its associated type (chronic, acute) and corresponding location (leftsided, rightsided or central) on computed tomography (CT) examinations. We trained our model on the largest available public Computed Tomography Pulmonary Angiogram PE dataset (RSNA-STR Pulmonary Embolism CT (RSPECT) Dataset, $N=7279$ CT studies) and tested it on an in-house curated dataset of $N=106$ studies. Our framework mirrors the radiologic diagnostic process via a multi-slice approach so that the accuracy and pathologic sequela of true pulmonary emboli may be meticulously assessed, enabling physicians to better appraise the morbidity of a PE when present. Our proposed method outperformed a baseline CNN classifier and a single-stage CNN-LSTM network, achieving an AUC of 0.95 on the test set for detecting the presence of PE in the study.

Keywords: Computer-Aided Diagnosis · CNN · LSTM · Pulmonary Embolism.

1 Introduction

Clinical Motivation. Pulmonary embolism (PE) is the most common preventable cause of hospital death in the United States with diagnostic delay and diagnostic challenges among the most common etiologies for resultant mortality

¹ This work will be presented at MICCAI 2021

[6]. The time dependent nature of PE diagnosis is especially critical given that the treatment involves anti-coagulant therapy which must be administered in an attempt to halt thrombus growth. Delayed diagnosis and treatment enables undiagnosed clots to increase in size. The patient’s mortality therefore becomes compounded by not only the resultant ischemic insult to the pulmonary parenchyma but also by the potential development of decompensated right heart failure. Currently, attempts to reduce diagnostic delays have predominantly focused on the speed of PE symptom identification in the emergency setting and little to no changes made to the imaging diagnosis paradigm through which treatment is ultimately decided [6]. Through the application of machine learning techniques, the diagnostic delays and errors in PE identification may be mitigated, thereby decreasing patient mortality.

Technical Motivation. Similar to the neoplastic applications of deep learning paradigms [22, 5, 3], the utilization of automated systems for PE detection is critical when diagnostic accuracy and efficiency are considered. Early works in automated PE diagnosis have mostly relied on traditional feature engineering and pulmonary vessel segmentation to reduce the search space which is quite computationally intensive [14, 16, 27, 17, 18]. Furthermore, most of these studies have reported results on small datasets. Recent Convolutional Neural Network-Long Short-Term memory (CNN-LSTM) methods [19, 20] and an end-to-end deep learning model, PENet [8], have demonstrated promise when diagnosing PE on Computed Tomography Pulmonary Angiogram (CTPA) examinations, but are overall limited in their binary classification of whether a PE is present or not. Methods based on blood clot segmentation [19, 25] for detecting PE are quite computationally-intensive, due to their use of only segmented blood clot regions; other regions which may show pulmonary abnormalities suggestive of clots get ignored.

In this work, we present a two-stage attention-based network consisting of a CNN and a Sequence model (LSTM + Dense) for prediction of PE and its associated characteristics. Current detection algorithms are largely modeled on the premise that every single CT slice is evaluated as an independent unit versus analyzing the relationship between successive CT slices. Our model extrapolates information from this 3D relationship in an attempt to better mimic the human cognitive process when examining a cross-sectional image. Due to the fact that PE is often observed in a small subset of 2D slices of a CT volume, the proposed attention module assists the network in locating insightful slices and assigns them higher weights in a bag-level feature aggregation mechanism. The slices with a higher attention are more likely to contain important information for PE identification thereby aiding in subsequent PE prediction. Besides, unlike other methods, our framework is trained to identify forms of PE on each CTPA examination slice and other PE attributes such as laterality, chronicity, and the RV/LV (Right Ventricular to Left Ventricular) ratio which can prove clinically useful in determining patients at risk [13, 7]. Our contributions can be summarized as follows:

1. Our attention based network provides image-level (ψ_{image}) PE prediction for each of the CTPA examination slices, study-level (ψ_{study}) prediction of PE in the CTPA volume and other associated PE characteristics.
2. Multiple Instance Learning (MIL) pooling [10] is used as an attention mechanism that provides insight into the contribution of each CTPA examination slice and the aggregate feature representation for study-level prediction.
3. Our network is trained on the largest publicly available PE dataset (RSPECT $N>7000$ studies) [4] and its performance has been evaluated on an external test dataset ($N=106$ studies).
4. Our network is designed and trained to explicitly obey clinically-defined label hierarchy and avoid making conflicting label predictions.

2 Proposed Methodology

Our network architecture comprises of two stages: 1) a CNN classifier to capture the image properties and study labels and 2) a sequence model for learning inter-slice dependencies. The CNN is used to extract features from every slice from the study and the sequence model combines these *spatially dependent* features and captures long-range dependencies to give the network information about the global change around each CT slice.

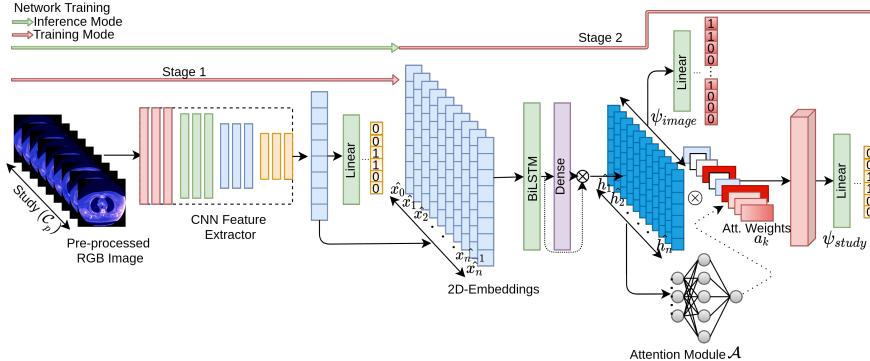


Fig. 1: Overview of our two-stage attention based PE detection model. The first stage involves training with all slices in study C_p for all image and study labels. The 2D-embedding $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_3\}$ from stage 1 after the pooling layer is used for the stage 2 LSTM-based training with an attention module for generating final predictions

2.1 Pre-processing

CT scans capture tissue radiodensity information. Voxel intensities range from -1000 to 3000 Hounsfield Units. Different window levels are used by radiologists

to evaluate lung pathologies. We first convert the single channel images to a 3-channel RGB format [2], where each channel is a distinct window [24] corresponding to the lung (window level = -600, width = 1500), PE (window level = 100, width = 700), and mediastinal ranges (window level = 40, width = 400) . The PE and mediastinal windows enable detection of blood clots. The lung window, on the other hand, may show pulmonary abnormalities suggestive of a clot. The constructed 3-channel RGB images were used to train our network.

2.2 Stage 1: CNN Classifier to capture different label properties

We used a Convolutional Neural Network (CNN) Efficient-Net [23] architecture as a backbone to extract features, followed by average pooling and a linear layer to classify each image and study. The pooling layer reduces the size of features maps after the convolutional layer so that we can make use of all the slices present in the study at once while training the Sequence model in Stage 2. This CNN classifier is trained for multi-label prediction on the RSPECT [4] dataset to capture the properties for different study labels. We make use of spatial feature maps generated in Stage 1 to train our Sequence model in Stage 2.

2.3 Stage 2: Sequence Model for learning inter-slice dependencies

The Sequence model, consisting of a bidirectional long short-term memory (Bi-LSTM) [9] and a dense layer, is used to capture long-range dependencies in CT scans. It makes use of the extracted features from study slices using the trained CNN classifier from Stage 1 and passes them through a Bi-LSTM and a dense layer to provide the network additional contextual information about the global changes around the slices. The features extracted using CNN-LSTM network capture spatio-temporal information in the CT scan volume. The ‘temporal’ aspect refers to the global relationship between successive slices. These features are subsequently used for the final prediction. The Sequence model has two heads - one for image (ψ_{image}) and another for study (ψ_{study}) level predictions. The ψ_{image} head is used for detecting the presence of PE on each slice in the study and the ψ_{study} head provides predictions regarding different characteristics of PE in the given study consisting of hundreds of slices.

Attention Mechanism. Features from the Sequence Model’s Bi-LSTM and dense layers, $(\hat{h}_1, \hat{h}_2, \dots, \hat{h}_n)$, corresponding to slices \mathcal{C}_{pq} , $q \in [1, 2, \dots, n]$ in study volume \mathcal{C}_p are fed into the attention module \mathcal{A} to obtain bag level features for \mathcal{C}_p . The aggregated bag-level features are used for global prediction by passing it through the the ψ_{study} classifier. Attention pooling is the weighted average of the slice-wise features present in the study and these attention weights a_k are learned from the neural network as $a_k = \frac{\exp\{w^T \tanh(V \hat{h}_k^T)\}}{\sum_{j=1}^n \exp\{w^T \tanh(V \hat{h}_j^T)\}}$, where w and V are learnable parameters and a_k denotes the learned attention weight distribution of \mathcal{A} . Features from each slice have a corresponding weight, and the bag-level feature set is obtained as $z = \sum_{k=1}^n a_k \hat{h}_k$. The attention module enables the network to locate informative slices present in the study. During

feature aggregation, it assigns higher weight to the more informative slices for final study-level prediction.

Loss Function. We use a custom loss function for training, similar to the performance metrics for the RSNA STR Pulmonary Embolism Detection challenge [1]. The study-level weighted log loss is defined as

$$L_{ij} = -w_j * [y_{ij} * \log(p_{ij}) + (1 - y_{ij}) * \log(1 - p_{ij})],$$

where i and j denote the study and label respectively, y_{ij} is ground truth label $\{0,1\}$ and p_{ij} is prediction probability of label j for study i that $y_{ij} = 1$. w_j signifies the weight of label j . Similarly, the image-level weighted log loss is defined as

$$L_{ik} = -[w * \frac{\sum_{k=1}^{n_i} y_{ik}}{n_i}] * [y_{ik} * \log(p_{ik}) + (1 - y_{ik}) * \log(1 - p_{ik})]$$

Here i is the study and $k = 1, 2, 3.., n_i$, where n_i is the total number of images in study i , y_{ik} is ground truth $\{0,1\}$ and p_{ik} is prediction probability (for presence of PE) on image k in study i . The total loss of study i is the sum of the image-level and study-level loss divided by the total weights, given by

$$L_i = \frac{\sum_{k=1}^{n_i} L_{ik} + \sum_{j=1}^n L_{ij}}{\sum_{j=1}^n w_j + w * \sum_{k=1}^{n_i} y_{ik}}.$$

Label Consistency. Our model ensures logical consistency of the predicted

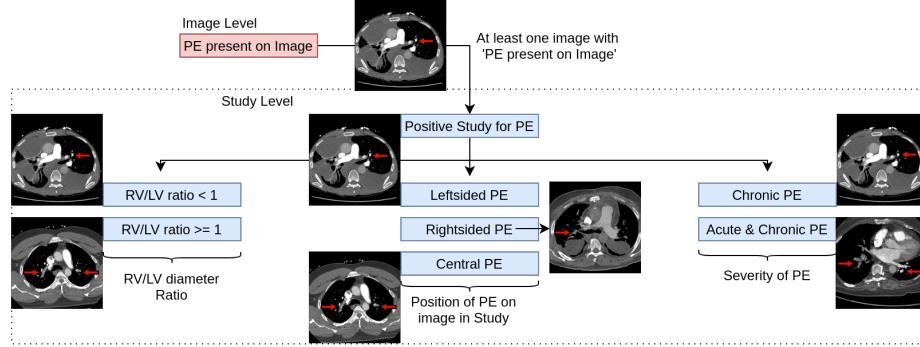


Fig. 2: Flowchart outlining relationships between labels

labels using constraint-based modified activation function. According to the rules in [1], a label is considered as predicted if it is assigned a probability (p) of more than 0.5. Label consistency rules of PE are as follows:

- If any image in a study is predicted positive for the presence of PE

- At least one of the three labels, i.e, *Right Sided, Left Sided or Central* PE must be assigned a probability greater than 0.5.
- Only one of $RV/LV < 1$ or $RV/LV \geq 1$ will have $p > 0.5$ and one of it must be present if at least one image is found positive for PE in study.
- Both *Chronic & Acute* and *Chronic* cannot be predicted positive at the same time , i.e, only one can have $p > 0.5$.
- If no image in the study volume is found positive for presence of PE, then
 - Either *indeterminate* or *negative for PE* labels must have a $p > 0.5$; both cannot have a $p > 0.5$.
 - All positive study related labels, i.e, *right sided PE, left sided PE, RV/LV < 1, RV/LV \geq 1, central PE* must have a $p < 0.5$.

Table 1: Data distribution in training, validation and test sets.

| labels | Train | Val | Test |
|-------------------------------|---------|--------|-------|
| # studies | 5824 | 1455 | 106 |
| # studies with positive PE | 1878 | 490 | 56 |
| # study with $RV/LV \geq 1$ | 754 | 186 | 43 |
| # study with $RV/LV < 1$ | 1878 | 490 | 63 |
| # study with leftsided PE | 1242 | 302 | 51 |
| # study with rightsided PE | 1486 | 389 | 53 |
| # study with central PE | 318 | 83 | 43 |
| # study with Chronic PE | 220 | 72 | 2 |
| # study with Acute&Chronic PE | 116 | 29 | 0 |
| # slice | 1431318 | 359276 | 87764 |
| # slice with PE | 77453 | 19087 | – |

3 Experimental Design

3.1 Dataset Description

Training Dataset. We use the largest publicly available annotated PE dataset (the RSNA Pulmonary Embolism CT, RSPECT dataset), comprising more than 12,000 CT studies [4]. This dataset, composed of CTPA scans and associated ground truth annotations has been contributed to by five international research centers and labeled by a group of more than 80 expert thoracic radiologists. There are multiple labels for each study and one label for each slice present in the study. The flowchart outlining the relationships between different labels is shown in Figure 2. 1,790,594 slices corresponding to 7279 studies have a complete set of ground truth labels available. 5824 of these studies were randomly selected as the training (D_1) and 1455 studies as the validation (D_2) cohorts.

Independent test set. We retrospectively collected data from 106 patients (D_3) who received CTPA studies under the PE protocol (GE Revolution CT, GE

Lightspeed VCT, GE Lightspeed 16, Phillips Brilliance 64) at *BlindedInstituteX*. The images were anonymized in accordance with Health Insurance Portability and Accountability Act and our Institutional Review Board guidelines. Both male and female adult patients (over 18 years) were selected with random sampling utilized in order to obtain similar numbers of positive versus negative PE cases (56 positive and 50 negative). CTPA studies with extensive artifacts (i.e. beam hardening, motion artifact, over/under exposure) were excluded. The ground truth study-level labels were provided by three expert readers working in consensus. Two readers (>15 yr and >10 yr exp) are board certified radiologists and the third third reader (>3 yr exp) is a radiology resident.

3.2 Framework Implementation

We trained our PE network in two stages. First, we trained the CNN classifier using all pre-processed 3-channel slices in D_1 , with both image and study-level labels as targets. Each batch consists of slices from the same study and the final study-level prediction was based on the prediction with maximum confidence. This multitask learning stage not only captures the presence of PE in slices but also learns other PE properties based on different study-level targets. We extracted the features of each slice from study \mathcal{C}_p using the trained CNN network and map it to a 2-D embedding $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_3\}$ using global average pooling. To exploit the sequential inter-slice spatial relationship and learn global intensity variations, we passed the embedding in order of their captured time points to the LSTM layer. The extracted features from the LSTM module passes through the attention module to yield bag-level features for the study volume \mathcal{C}_p , which is subsequently fed to the final study-level classifier(ψ_{study}); features from the LSTM are directly provided to the image-level classifier(ψ_{image}) for detecting the presence of PE on individual slices. Both stages were trained with the Adam optimizer [12] and the custom loss function discussed in section 2.3. The model with the lowest loss value on D_2 was chosen for evaluation on D_3 . Both stages were implemented in PyTorch 1.4 and trained using one NVIDIA Tesla T4 Google Cloud Instance GPU.

4 Results

We used area under the receiver operating characteristic curve (AUC) as our performance metric. Our proposed method was compared with a CNN classifier (Baseline 1) and a CNN-LSTM network without the attention module (Baseline 2). The AUCs of Baseline 1, Baseline 2, and our method were 0.5, 0.94, and 0.95, respectively. The corresponding accuracies were 0.74, 0.65, and 0.88, respectively. The performance curves of our best model for different images and study labels on D_2 and D_3 are shown in Figure 3. Our proposed network achieved an AUC of 0.82 for detecting PE on D_2 and 0.95 on the D_3 . CT volumes in D_3 had different slice thickness and higher number of slices per study as compared to those in D_1 and D_2 , which potentially provides more contextual information to the model.

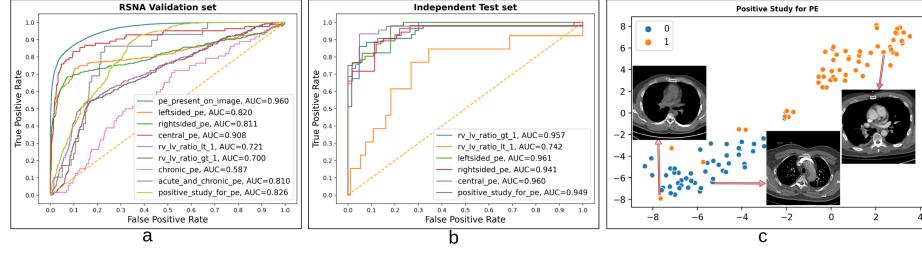


Fig. 3: (a,b) ROC curves for different image and study-level labels on the validation and independent test set. (c) 2D scatter plot representation using t-SNE of bag level features from Attention Module corresponding to studies in D_3

To provide an intuitive understanding into the features and the context learnt by the network, class activation maps (CAMs) [26] were computed by averaging the feature maps from the final convolutional layer of CNN network. As may be observed on a few representative slices in Figure 4, there is a strong activation over the regions surrounding the clots, as verified by our collaborating radiologists (see interpretation in Figure 4 caption). This corroborates the clinical characteristics of PE diagnosis. A t-Distributed Stochastic Neighbor Embedding (t-SNE) [15] was applied to the attention module’s bag level features. The 2D scatter plot visualization is shown in Figure 3.

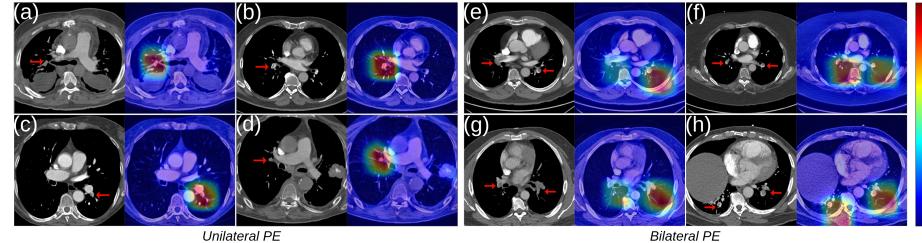


Fig. 4: **Visualization of class activation maps.** CTPA slices (a-d) demonstrate unilateral PE marked by red arrows. High intensity regions in the CAMs represent the precise location of PE. (a, b, d) show right-sided lobar/segmental PE. (c) shows lower segmental PE. Slices (e-h) demonstrate bilateral PE. (f, h) show lobar/segmental PE and CAMs depict the exact location of PE. In (e, g), the CAMs do not accurately define the maximal intensity on the right-sided central/lobar PE; however, they depict left-sided PE accurately.

As may be observed in Figure 3, our two-stage CNN-LSTM network with an attention mechanism is able to learn and predict not only presence of PE, but also its associated attributes with a high AUROC. Note that the number of

ROC curves are fewer in D_3 as compared to D_2 in Figure 3 since image-level annotations and enough number of studies for *chronic* and *acute & chronic* PE were not available for D_3 .

5 Conclusion

In this work, we developed an attention based two-stage network for CTPA classification on the largest publicly available PE dataset [4] and tested its performance on a held out validation and an independent test set from a different clinical site. We believe the contribution is significant not only due to the large-scale multi-institutional validation, but also because of how the framework is designed to explicitly obey a clinically-defined label hierarchy. Our framework is among the first to identify forms of PE on each CTPA slice as well as other attributes such as laterality, chronicity, and the RV/LV ratio which can prove clinically useful in determining patient risk. Our model is also designed to mimic the human cognitive process when examining cross-sectional scans. Furthermore, the generated CAMs have been validated by radiologists. Our motivation behind a two-stage 2D CNN-LSTM training approach stems from the GPU memory limitations in the PE challenge [1]. The results of the challenge are also suggestive of the fact that a 2D-CNN with a Sequence model works better than a 3D-CNN for this task. We achieved a loss of < 0.180 on the Kaggle test set in the final leaderboard [1]. The state-of-the-art 3D-CNN, PENet [8], achieved a mean AUC of around 0.85 for detecting PE. Our study-level PE prediction performance is similar to PENet; however, unlike PENet, our network provides slice-level predictions as well as predictions for associated PE attributes. There are a few limitations to our approach. The CAMs are able to detect blood clots as informative regions with high certainty in unilateral PE. However, in bilateral PE, we can observe that the some discriminative regions are not precisely from the PE regions. This will be investigated in our future work. Additionally, we will evaluate the efficacy of an end-to-end network and radiomic features [21, 11] for the prediction tasks. Refining the radiologic PE diagnostic paradigm through the use of a two-stage network may enable radiologists to reduce inherent human error and also decrease patient morbidity and mortality.

Acknowledgment: Dimitris Samaras was partially supported by the Partner University Fund, the SUNY2020 Infrastructure Transportation Security Center, and a gift from Adobe.

References

1. RSNA-STR pulmonary embolism CT (RSPECT) dataset, copyright rsna, <https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pe-detection-challenge-2020>

2. Anam, C., Budi, W., Haryanto, F., Fujibuchi, T., Dougherty, G.: A novel multiple-windows blending of ct images in red-green-blue (rgb) color space: Phantoms study. *Scientific Visualization* **11**(5) (2019)
3. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama* **318**(22), 2199–2210 (2017)
4. Colak, E., Kitamura, F.C., Hobbs, S.B., Wu, C.C., Lungren, M.P., Prevedello, L.M., Kalpathy-Cramer, J., Ball, R.L., Shih, G., Stein, A., et al.: The rsna pulmonary embolism ct dataset. *Radiology: Artificial Intelligence* **3**(2), e200254 (2021)
5. Couدرay, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A.: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine* **24**(10), 1559–1567 (2018)
6. Friedman, T., Winokur, R.S., Quencer, K.B., Madoff, D.C.: Patient assessment: clinical presentation, imaging diagnosis, risk stratification, and the role of pulmonary embolism response team. In: *Seminars in interventional radiology*. vol. 35, pp. 116–121. Thieme Medical Publishers (2018)
7. Ghaye, B., Ghysen, A., Bruyere, P.J., D’Orio, V., Dondelinger, R.F.: Can ct pulmonary angiography allow assessment of severity and prognosis in patients presenting with pulmonary embolism? what the radiologist needs to know. *Radiographics* **26**(1), 23–39 (2006)
8. 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.: Penet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric ct imaging. *NPJ digital medicine* **3**(1), 1–9 (2020)
9. Huang, Z., Xu, W., Yu, K.: Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991* (2015)
10. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: *International conference on machine learning*. pp. 2127–2136. PMLR (2018)
11. Khorrami, M., Prasanna, P., Gupta, A., Patil, P., Velu, P.D., Thawani, R., Corredor, G., Alilou, M., Bera, K., Fu, P., et al.: Changes in ct radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non-small cell lung cancer. *Cancer immunology research* **8**(1), 108–119 (2020)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
13. Lankeit, M.: Always think of the right ventricle, even in “low-risk” pulmonary embolism (2017)
14. Liang, J., Bi, J.: Computer aided detection of pulmonary embolism with tobogganning and multiple instance classification in ct pulmonary angiography. In: *Biennial International Conference on Information Processing in Medical Imaging*. pp. 630–641. Springer (2007)
15. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. *Journal of machine learning research* **9**(11) (2008)
16. Masutani, Y., MacMahon, H., Doi, K.: Computerized detection of pulmonary embolism in spiral ct angiography based on volumetric image analysis. *IEEE Transactions on Medical Imaging* **21**(12), 1517–1523 (2002)
17. Özkan, H., Osman, O., Şahin, S., Boz, A.F.: A novel method for pulmonary embolism detection in cta images. *Computer methods and programs in biomedicine* **113**(3), 757–766 (2014)

18. Park, S.C., Chapman, B.E., Zheng, B.: A multistage approach to improve performance of computer-aided detection of pulmonary embolisms depicted on ct images: Preliminary investigation. *IEEE transactions on biomedical engineering* **58**(6), 1519–1527 (2010)
19. Rajan, D., Beymer, D., Abedin, S., Dehghan, E.: Pi-pe: A pipeline for pulmonary embolism detection using sparsely annotated 3d ct images. In: *Machine Learning for Health Workshop*. pp. 220–232. PMLR (2020)
20. Shi, L., Rajan, D., Abedin, S., Yellapragada, M.S., Beymer, D., Dehghan, E.: Automatic diagnosis of pulmonary embolism using an attention-guided framework: a large-scale study. In: *Medical Imaging with Deep Learning*. pp. 743–754. PMLR (2020)
21. Singh, G., Manjila, S., Sakla, N., True, A., Wardeh, A.H., Beig, N., Vaysberg, A., Matthews, J., Prasanna, P., Spektor, V.: Radiomics and radiogenomics in gliomas: a contemporary update. *British Journal of Cancer* pp. 1–17 (2021)
22. Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.W., Snead, D.R., Cree, I.A., Rajpoot, N.M.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging* **35**(5), 1196–1206 (2016)
23. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*. pp. 6105–6114. PMLR (2019)
24. Wittram, C., Maher, M.M., Yoo, A.J., Kalra, M.K., Shepard, J.A.O., McLoud, T.C.: Ct angiography of pulmonary embolism: diagnostic criteria and causes of misdiagnosis. *Radiographics* **24**(5), 1219–1238 (2004)
25. Yang, X., Lin, Y., Su, J., Wang, X., Li, X., Lin, J., Cheng, K.T.: A two-stage convolutional neural network for pulmonary embolism detection from ctpa images. *IEEE Access* **7**, 84849–84857 (2019)
26. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2921–2929 (2016)
27. Zhou, C., Chan, H.P., Patel, S., Cascade, P.N., Sahiner, B., Hadjiiski, L.M., Kazerooni, E.A.: Preliminary investigation of computer-aided detection of pulmonary embolism in threedimensional computed tomography pulmonary angiography images. *Academic radiology* **12**(6), 782 (2005)