2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI) | 979-8-3503-3458-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IRI58017.2023.00061

Weakly Supervised Confidence Aware Probabilistic CAM multi-Thorax Anomaly Localization Network

Tanushree Meena*¶, Anwesh Kabiraj†¶, Pailla Balakrishna Reddy‡‖and Sudipta Roy§¶
Artificial Intelligence and Data Science, Jio Institute, Navi Mumbai, India¶,
Artificial Intelligence Centre of Excellence Reliance AI-CoE , Hyderabad, India‖
Email: *tanushree.meena@jioinstitute.edu.in, †anwesh.kabiraj17@gmail.com,
‡balakrishna.pailla@ril.com, §sudipta1.roy@jioinstitute.edu.in

***Abstract*—Most anatomical information and anomalies are provided by chest X-ray (CXR) images and are sometimes adequate for the early diagnosis. However, by observing the radiographs it is a challenging task to recognize multiple occurring thorax diseases. As a result, there is a trend toward developing deep learning systems to assist radiologists. Motivated by this, we propose a generalized weakly supervised Confidence-Aware Probabilistic Class Activation Map (CAPCAM) classification model that localizes anomalies for 13 different thoracic disease. The CAPCAM used CX-Ultranet as the backbone with the combination of Confidence Aware Network (CAN) and Anomaly Detection Network (ADN) which helps the model to utilise all components of feature extracted therefore eliminating the necessity to train them individually and time taken. We experimentally shown that our proposed CAPCAM method sets a new state-of-the-art benchmark by achieving accuracy in terms of Intersection of Bounding Box (IoBB) in the range of 85% - 94%, and F1 scores in the range of 88% - 90% for all thirteen diseases on publicly available large-scale CXR NIH dataset. The proposed CAPCAM pooling also achieves better results than other state of the art (SOTA) pooling methods.**

***Index Terms*—CXR, weak supervision, medical image, deep learning**

I. Introduction

According to the Global Burden of Disease Study 2017 [1], thorax diseases are a major cause death in the world. Chest X-ray (CXRs) imaging is a non-invasive technique and has the potential of diagnosing multi-thorax diseases [2]. Multiple thorax diseases can be detected from a single CXR. But manual identification of these diseases from a CXR is very challenging. This challenge marks the implementation of deep learning (DL) for the detection of thoracic diseases. Recent studies [3] [4] have shown that the implementation of DL model was effective to the radiologists for disease detection. Traditional thoracic DL systems use manual feature extraction and ML techniques to analyze shape texture, contrast, intensity, and shape. Since each thoracic disease has a unique pattern, developing a standard feature set for multiple thoracic diseases is a challenging task. Automatic detection of multiple diseases from a CXR scan is crucial in healthcare. DL techniques have the potential to detect diseases automatically and improve accuracy. However, DL techniques face robustness and reliability challenges in clinical care.Traditional thoracic DL methods typically rely on ML feature extraction at the root and are often specialized for detecting a few specific diseases. In contrast, our model does not rely on training features for individual diseases. Instead, we propose an efficient and generalized weakly supervised probabilistic class activation map pooling model. Our model outperforms other state-of-the-art pooling methods in terms of IoBB and F1 scores and achieves excellent results on multiple datasets. The integration of components like CAN and ADN enables the model to utilize all extracted features without the need for individual training. Our approach addresses limitations of traditional thoracic DL systems by employing weakly supervised learning, introducing probabilistic class activation map pooling, and combining attention and discrimination networks. These advancements reduce annotation requirements, enhance pooling accuracy, and improve disease localization capabilities.

II. Related Work

In the recent years, several ML and DL techniques have been proposed for the classification of thoracic diseases. Ho et al. [5] presented multiple feature integration by using a pre-trained deep CNN model and shallow handcrafted techniques. In [6] an attention-guided curriculum learning method is used for disease classification and localization of the lesion area. For multi-label thoracic disease classification, a framework based on residual attention learning is proposed in [7]. DualCheXNet [8] enables dual level feature fusion process namely decision level and feature level which results in the complementary feature learning embedded in the network. CheXNet [9] is an approach which uses a modified version of Dense Net 121 for the detection of pneumonia and other 13 diseases from the ChestXray14 dataset. A novel lesion location attention guided network LLAGnet [10] consists of two corresponding attention modules that focus on the discriminative features from lesion locations for multi-label thoracic disease classification in CXRs. In [11], the authors used a triple learning technique which integrates an element, channel and scale wise attention modules which can instantaneously learn disease-discriminative channels, locations, and scales for effective diagnosis. The work [12] focusing on the fused semantic features from encoded X-ray features along with radiology reports which are fed to the encoder transformer to use chest radiographs and meta-data related to them. To diagnose the chest x ray images, based on the past information, medical concept graph [12] was introduced. MXT [13] is a new

979-8-3503-3458-6/23/$31.00 ©2023 IEEE
DOI 10.1109/IRI58017.2023.00061

variant of pyramid vision Transformer for multi-label chest X-ray image classification. This approach can capture both short and long-range visual information through self-attention. CheXclusion [14] is an approach which examines the degree to which state-of-the-art DL classifiers show true positive rate disparity among various privacy traits. A weakly supervised DL [15] approach for the localization and classification of thorax diseases was developed which is a combination of excitation and sequence block followed by the multi map layer and min max pooling. To address the domain and label discrepancy [16] throughout the multiple datasets, a task-specific adversarial training and uncertainty-aware temporal ensembling were introduced. However, no systematic study of the possibilities of incorporating anatomical prior to improve classification accuracy was conducted in any of the methods mentioned above. In this study, we propose a weakly supervised CAPCAM pooling for the efficient localization of diseases thoracic diseases.

III. Methodology

A. CX-Ultranet as a highly efficient feature extractor

The CAPCAM model works with a CNN. For our work we used CX-Ultranet [17], pretrained on NIH Chest X-ray dataset and CheXpert dataset. The image is initially passed through layer-by-layer mobile inverted convolutional networks. The multiclass cross entropy loss function layer is responsible for segregating the optimized feature maps and train the CNN. At this point, we have extracted the feature maps and pass it to our CAPCAM pooling network. Some other benefits of using part of the CX-Ultranet [17] as feature extractor is its high accuracy and efficiency. Since CX-Ultranet [17] [17] predicts multiple diseases with high accuracy, therefore we use the same result to aid our CA- PCAM pooling network. The network is capable of learning different anomalies by itself however its ability to classify the same is much lesser than that of CX-Ultranet. Therefore, we have used it as a post processing method for more accurate anomaly localization.

B. Probabilistic Class Activation Map (PCAM) Pooling Method

The proposed confidence aware probabilistic CAM pooling for anomaly localization consists of a feature extractor coupled with PCAM pooling, anomaly prediction network and confidence prediction network is shown in fig. 1. All the channels among constant embedding share constant weight. The fundamental assumption of attention pooling follows the multiple instance learning (MIL) framework, that treats embedding as instances. The CXRs as a bag of instances is positive, e.g., one of the 13 different thoracic diseases, if a single instance is positive. The PCAM pooling is based upon the same MIL framework but generates normalized attention weights for every embedding of the feature map. The novel PCAM pooling method is described in detail in fig. 2. Given a fully convolutional neural network trained for multiclass classification the feature map is denoted by M with shape (C, H, W) which we get from the last fully connected layer of the CNN (CX-Ultranet). Here, C denotes the channel dimension while H and W denotes the height and width of the feature map. We also take the disease with highest probability from CX-Ultranet [17] and target anomaly localization for it. The Class Activation Map of a particular thoracic disease is given by:

$$f_{i,j} = w^t M_{i,j} + b \quad | \quad i,j \in H,W \tag{1}$$

where, $M_{i,j}$ is feature embedding of the feature map M with shape (C,H,W) at position (i,j). w,b are the weights and bias of the CX-Ultranet [17] from its last fully connected layer. $f_{i,j}$ is the logit function before the sigmoid operation under multiclass classification setting which monotonically calculates the disease likelihood over $M_{i,j}$. Later it is used to guide the PCAM pooling method to generate heatmap and then refine the target anomaly location. To measure contribution of each embedding the assigns normalized attention weights are assigned. Since, $f_{i,j}$ is unbounded and exists in $(-\infty,+\infty)$ therefore we bound it with a sigmoid function $p_{i,j}$ defined in equation (2) and (3).

$$p_{i,j} = sigmoid(s_{i,j}) \tag{2}$$

$$x = \sum_{i,j}^{H,W} w_{i,j} M_{i,j} w_{i,j} = \frac{sigmoid(w^t M_{i,j} + b)}{\sum_{i,j}^{H,W} w_{i,j} sigmoid(w^t M_{i,j} + b)} \tag{3}$$

x is the pooled feature embedding layer and w(i,j) are the extracted attention weights. These are utilized to generate heatmap for the input images. Subsequently, we demonstrated that the generated heatmaps precisely covers the anomalous region The threshold is set to 95% of the threshold map for better visual clarity. Results from other compared SOTA pooling methods are also given same probability thresholding so that the visual comparisons are on equivalent. By setting a threshold, we can control the trade-off between false positives and false negatives. A higher threshold reduces the number of positive detections but increases the likelihood of true positives.

C. Anomaly Detection Network

The ADN consists of a multi layered perceptron with 100-neuron hidden layer and one output layer as shown in fig. 2. It generates an anomaly score corresponding to the input image x and is formulated as

$$\alpha = \phi(x, M, t) \tag{4}$$

M is the feature extracted by the backbone CNN and t is the weakly supervised trainable parameter. According to [18] Gaussian distributions matches well with such scores in plenty of datasets. We define an univariate gaussian distribution as r_1, r_2, $r_3 \ldots$ $r_l \sim N(\mu,\sigma^2)$. The corresponding reference score is given by

$$\mu_R = \frac{1}{l} \sum_{i=1}^{l} r_i \tag{5}$$

$$\sigma_R^2 = \frac{1}{l} \sum_{i=1}^{l} (r_i - \mu_R)^2 \tag{6}$$

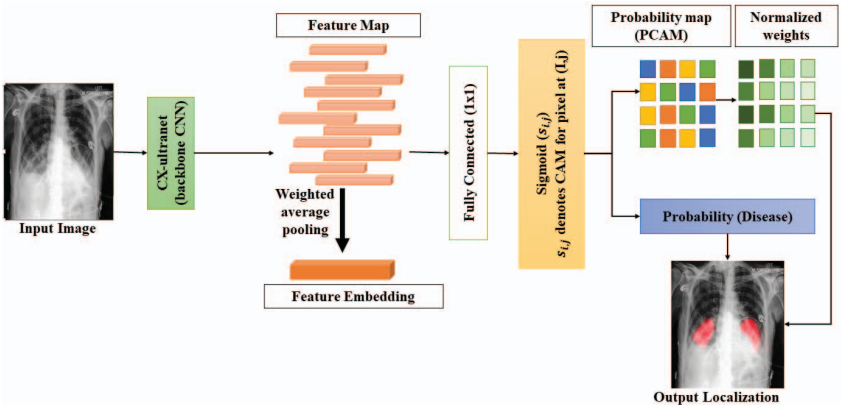

Fig. 1. The overall Generation of Probabilistic Class Activation MAP from the Input Image.

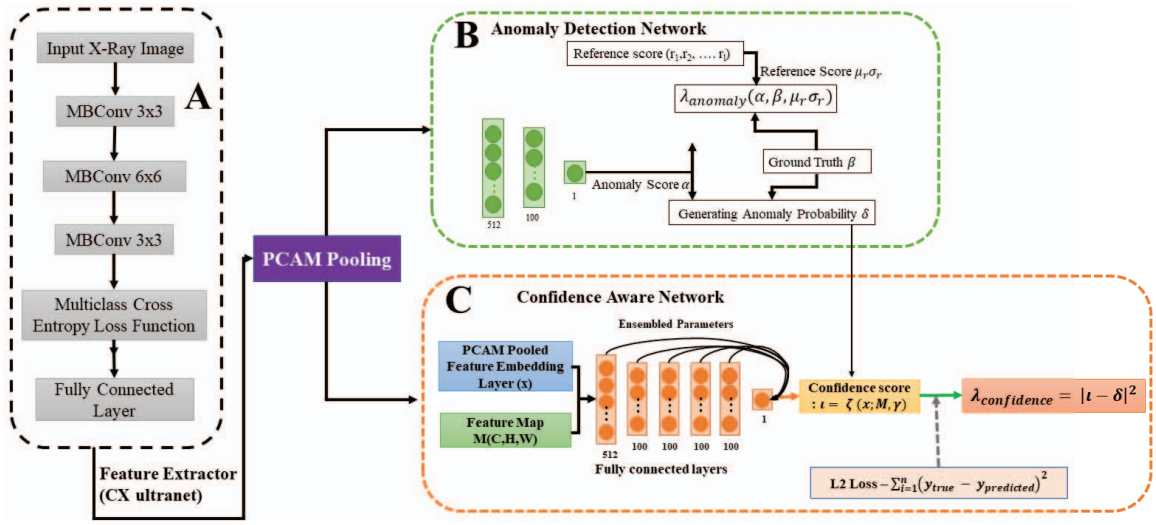

Fig. 2. The overall flow of the architecture starting with A) CX-UltranetUltraNet as feature extractor, PCAM pooling the feature maps and embeddings and then passing them through the B) ADN and C) CAN visualize the precise localizations.

Following a very similar work done in [19] we set the values of σ=1, μ=0 and l=5000. To optimize the anomaly detection module, we use the following contrastive loss [20] [21]

$$\lambda_{anomaly}(\alpha, \beta, \mu_r \sigma_r) = (1-\beta)|\frac{\alpha - \mu_R}{\sigma_R}| + \beta_{max}(0, margin - (\frac{\alpha - \mu_R}{\sigma_R}) \quad (7)$$

where σ_R is standard deviation of anomaly scores for the used normalized data generated by the univariate gaussian distribution. β is ground truth label where 0 indicates it as a neutral case while 1 indicates it as a positive case belonging to any of the 13 different classes, *margin* represents the Z-score confidence. In this work we have set the value to 5 based on empirical studies. The ADN if separately trained learns to classify between positive and neutral classes therefore making it a standalone binary classification model. In our work it is weakly supervised and we use it as a post processing method, after we already get the disease probability from CX-Ultranet. Also, the CX-Ultranet [17] is multiclass classifying D-CNN and solves any kind of class imbalance in a novel way. This makes the ADN inherently resilient to class imbalance if present in the dataset.

D. Confidence Aware Network

Our work deals with a multiclass problem at the very bottom. We consider a total of 13 thoracic diseases and aim to localize the anomaly for each one of them. In many cases a single X-ray is subject to more than one disease. The CX-Ultranet [17] can detect multiple diseases simultaneously and returns them as percentage probabilities along with confidence scores. The former plays vital importance, while the latter is used for comparison. However, if we just consider anomaly scores for localization then the result is not precise enough. Most of the current works rely on either probability thresholding or attention weights to guide the same. We have

already proposed a novel way of utilizing attention weights in the PCAM map which follows the MIL framework. Our confidence-aware network is also built upon the shared feature extractor of CX-Ultranet. It consists of the confidence aware module [22] with four 100-neuron hidden layers shown in fig. 2 for a strong confidence prediction. The computation of the CAN is formulated as

$$\iota = \zeta(x; \theta, \gamma) \tag{8}$$

Where ι represents the confidence score generated corresponding to the localized anomaly, γ represents the ensembled parameters of the confidence aware module. ζ lies between [0,1] therefore the standard l2 loss is used to optimize the confidence prediction module which is formulated as a regression task.

$$\lambda_{confidence} = |\iota - \delta|^2 \tag{9}$$

where δ is the anomaly probability which is defined as

$$\delta = \begin{pmatrix} prob, if \beta = 0 \\ 1 - prob, if \beta = 1 \end{pmatrix} \tag{10}$$

IV. Experiment / Implementation Details

A. Datasets

NIH: The NIH chest X-ray dataset [23] consists of 112,120 X-rays from 30,805 unique patients with 14 diseases. We strictly follow the official split of NIH, 70% for training, 10% for validation, and 20% for testing, for conducting experiments and fair comparison with previous works.

B. Implementation Details

Preprocessing Images with feature extractor– The CX-Ultranet [17] scales down 1024×1024 images present in NIH dataset to 512×512 images which is an optimal size for detection of anomalous regions as well as for finer feature detection. No image enhancements are done as any artificial artifact might cause the model to detect unnecessary anomalies and hence faulty results. The class imbalance problem is taken care by the novel CX-Ultranet [17] extractor. The Feature maps taken from the last fully connected convolution layer preserves the same image resolution. The image also comprises of three channels instead of just white and black.

Training scheme for guiding different disease localization- The trainable parameter t in ADN is trained explicitly to guide localization for different diseases. We used the weights and biases from CX-Ultranet [17] to train the same as it reduces need for a different training model. The weights and biases of only the disease with highest probability are considered.

V. Result and Discussion

A. Results

1) Evaluation Metrics: We test out the CAPCAM on various accuracy metrics including specificity, recall, precision and F1 scores shown in Table 1. CAPCAM shows a very high degree of precision and F1 scores owing to the fact it utilizes both attention weights as well as confidence score to guide the anomaly localization. We also perform an ablation study on three variations of the same network. F1 scores for all the thirteen thoracic diseases are more than ≅ 92% which is the harmonic mean of recall and precision. The CAPCAM network suffers from little to no class imbalance as CX-Ultranet [17] feature extractor deals with it in a very efficient way. The other reason behind achieving such high scores is due to the fact we also incorporate MIL framework while generating the PCAM. As this is a localisation task which uses the probability score map, confidence score maps and others to narrow down an anomalous region therefore it becomes necessary to define some threshold for pin pointing an area. Otherwise areas which are nowhere in the anomaly region will also have some very low +ve probability score and will come up in the region localised.

TABLE I
Specificity, Recall, Precision and F1 scores achieved by confidence aware probabilistic class activation map for the thirteen thoracic diseases.

Diseases	Specificity	Recall	Precision	F1-score
Cardiomegaly	96.24	95.47	92.24	93.82
Effusion	95.92	92.19	96.42	94.21
Emphysema	92.71	92.69	92.59	92.63
Infiltration	94.95	97.16	92.76	94.88
Mass	92.96	92.45	95.17	93.79
Nodules	95.94	95.16	95.16	95.12
Atelectasis	95.46	92.62	92.29	92.45
Pneumothorax	94.05	95.08	95.86	95.48
Pleural Thickening	92.37	91.19	96.43	93.73
Pneumonia	93.09	91.38	93.78	92.56
Fibrosis	95.18	91.76	96.9	94.22
Edema	92.76	95.11	96.13	95.61
Hernia	95.14	95.64	95.02	95.32

2) Intersection over Bounding Box (IoBB) scores: In this work we have focused mainly on localization so it is very important that the model should localize the anomalous regions as accurately as possible. Since the dataset we worked on only had bounding box as the ground truth so we will be calculating accuracy of CAPCAM with respect to the IoBB. The CAPCAM model localizes region which we later show in observations. CAPCAM pooling with confidence threshold set to 0.9 shows the best result by far than any other methods. The benefit of confidence prediction network is also clearly visible as more refined localization of the anomalous region as seen if fig. 3.

3) Visual Observation: We use the CAPCAM pooling network along with CX-Ultranet as a feature extractor to localize anomalies for 13 different thoracic disease. We calculate the confidence score for the anomaly localized separately and for our work we treat localized regions with a confidence score more than 0.9 as final anomalous regions. The threshold value is calculated empirically to have maximum IoBB scores. The network accurately targets different lesions and regions based on different diseases as visible. The input image in fig. 3A is affected to both Cardiomegaly and Fibrosis to some extent

TABLE II
IOBB AND F1SCORE FOR DIFFERENT POOLING METHODS FOR ALL 13 DISEASES

Methods	Global Average Pooling		LSE Pooling		LSE-LBA Pooling		Attention Pooling		PCAM Pooling (Baseline)		CAPCAM Pooling (Proposed method)	
Metrics	**IoBB**	**F1**	**IoBB**	**F1**	**IoBB**	**F1**	**IoBB**	**F1**	**IoBB**	**F1**	**IoBB**	**F1**
Cardiomegaly	48.68	60.3	70.63	61	71.1	68.4	65.04	69.5	73.1	77.2	**94.91**	**89.9**
Effusion	31.92	57.5	57.47	62.8	75.71	71.6	72.28	72.2	79.12	71.8	**93.72**	**90**
Emphysema	52.97	58.6	57.62	72	76.47	65.6	62.96	63	82.03	77.5	**93.55**	**90.2**
Infiltration	36.34	60.2	56.24	65.1	78.37	62.9	72.33	63.3	78.47	81.8	**97.17**	**89.4**
Mass	44.12	64.2	59.48	74.3	72.96	62.7	68.53	69.6	74.43	75.7	**93.25**	**90.7**
Nodules	59.72	58.1	53.54	65.9	73.66	68.4	71.13	68.4	83.59	75.1	**94.05**	**89.9**
Atelectasis	59.61	58	67.13	63	72.97	64.8	63.42	68.7	83.38	74.2	**94.33**	**89.9**
Pneumothorax	58.09	61.6	69.82	73.4	72.28	68	69.82	64.1	87.69	77.6	**97.83**	**89.6**
Pleural Thickening	46.42	61	66.29	69.1	72.76	65	63.84	66.2	76.88	78.9	**95.76**	**90**
Pneumonia	54.76	61.9	66.48	75.6	73.33	63	67.77	71.2	87	78.4	**97.64**	**90.4**
Fibrosis	43.33	63.9	69.18	63.8	71.9	64.6	70.91	71.3	78.37	76.5	**96.1**	**89.4**
Edema	48.08	51.7	66.45	65.8	75.04	67.6	72.55	74.8	83.55	81.6	**96.26**	**90.5**
Hernia	30.51	56.6	51.48	62.8	75.4	70.1	71.73	69.1	85.86	76.8	**93.56**	**90.1**

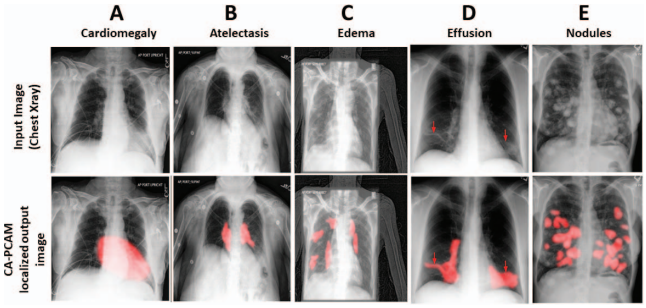

Fig. 3. Anomaly localization using CAPCAM for 13 different diseases A) Cardiomegaly, B) Atelectasis, C) Edema, D) Effusion and E) Nodules.

as symptoms of fibrosis are visible. However, the percentage probability of cardiomegaly returned by CX-Ultranet is higher than that of fibrosis, so the CAPCAM network focuses on localization for cardiomegaly which is increased heart size rather than fibrosis which is associated with fibrous white textures in the lungs. Although the CX-Ultranet is capable of multiclass classification, however in this work we just use a part of the network as extractor and the main aim is precise localization of the anomalies.

B. Discussion

We compared CAPCAM against Global Average Pooling [24], LSE pooling [25], LSE-LBA pooling [26], Attention Pooling [27] and baseline PCAM interms of IoBB and F1 score is shown in Table 2. The CAPCAM network outperforms all SOTA models by more than 20%-30% which is a major improvement. The high accuracy in detecting the correct region is majorly dependent on attention weight-based class activation map generation and then interpreting the same in a probabilistic model. This in turn is guided by an independent confidence prediction network for the activation map generated. Confidence thresholding improves CAPCAM from baseline PCAM by 10%-15% as observed in Table 2. Fig. 4 shows the visual comparison of multiple diseases generated by the three best models including attention pooling, LSE pooling and the proposed CAPCAM pooling. Attention pooling usually ends up taking more area under consideration than the actual anomalous regions which is not the case with LSE pooling. However, LSE pooling lags in situations where multiple unattached regions are needed to be localized as clearly seen from fig. 4 wherein LSE pooling fails to localize the anomalous regions in the upper portion of the lungs in the case of nodules. A limitation of this work could be that we have only utilizes one dataset, limiting the generalizability of the proposed model.

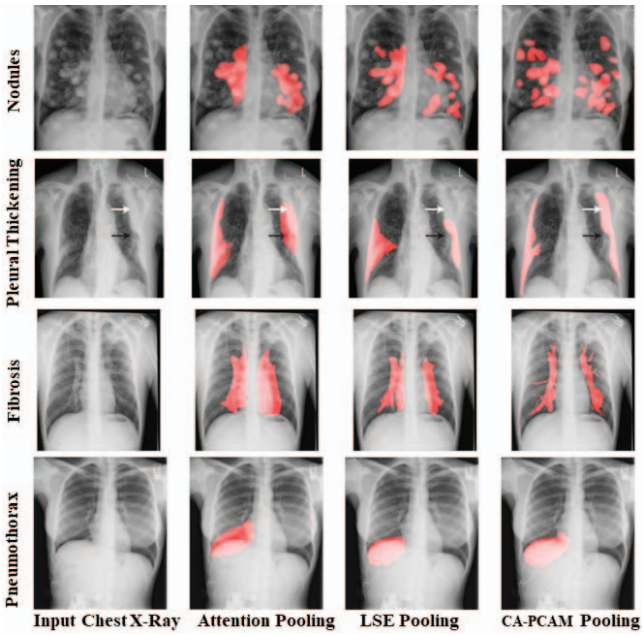

Fig. 4. Visual outputs of Attention, LSE, and CAPCAM pooling.

VI. Conclusion

In this work, we propose a weakly supervised CAPCAM pooling method for thoracic disease localization achieves superior performance on the NIH chest X-ray dataset, surpassing previous methods by 20-30% in accuracy and F1 scores. It outperforms other pooling methods, demonstrating its efficacy for identifying and localizing thoracic diseases. Thus, we can say that the proposed probabilistic class activation map pooling achieves better results than other SOTA pooling methods and is useful for identification and localization of thoracic diseases. We have only worked on NIH chest X-ray. To establish its effectiveness in diverse clinical settings, future research should involve multiple datasets for comprehensive analysis and validation. This would provide a broader understanding of the model's performance and its applicability across different populations, imaging protocols, and sources. Such an extension would enhance the model's robustness and reliability in real-world scenarios.

Acknowledgement

This research received no external funding. This research work was supported by the RFIER-Jio Institute "CVMI-Computer Vision in Medical Imaging" project under the "AI for ALL" research center. We would like to thank Dr Kalyan Tadepalli, Orthopedic surgeon, H. N. Reliance Foundation Hospital and Research Centre for his advice and suggestion.

Source code availability

The source code is available at: https://github.com/labsroy007/CA-PCAM.

References

[1] G. A. R. et al., "Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017," *The Lancet*, vol. 392, no. 10159, pp. 1736–1788, 2018.

[2] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[3] T. Meena and S. Roy, "Bone fracture detection using deep supervised learning from radiological images: A paradigm shift," *Diagnostics*, vol. 12, no. 10, 2022.

[4] S. Roy, T. Meena, and S.-J. Lim, "Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine," *Diagnostics*, vol. 12, no. 10, p. 2549, 2022.

[5] T. K. K. Ho and J. Gwak, "Multiple feature integration for classification of thoracic disease in chest radiography," *Applied Sciences*, vol. 9, no. 19, 2019.

[6] Y. Tang, X. Wang, A. P. Harrison, L. Lu, J. Xiao, and R. M. Summers, "Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs," in *Machine Learning in Medical Imaging*, Y. Shi, H.-I. Suk, and M. Liu, Eds. Cham: Springer International Publishing, 2018, pp. 249–258.

[7] Q. Guan and Y. Huang, "Multi-label chest x-ray image classification via category-wise residual attention learning," *Pattern Recognition Letters*, vol. 130, pp. 259–266, 2020, image/Video Understanding and Analysis (IUVA).

[8] "Dualchexnet: dual asymmetric feature learning for thoracic disease classification in chest x-rays," *Biomedical Signal Processing and Control*, vol. 53, p. 101554, 2019.

[9] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[10] B. Chen, J. Li, G. Lu, and D. Zhang, "Lesion location attention guided network for multi-label thoracic disease classification in chest x-rays," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 2016–2027, 2020.

[11] H. Wang, S. Wang, Z. Qin, Y. Zhang, R. Li, and Y. Xia, "Triple attention learning for classification of 14 thoracic diseases using chest radiography," *Medical Image Analysis*, vol. 67, p. 101846, 2021.

[12] D. Hou, Z. Zhao, and S. Hu, "Multi-label learning with visual-semantic embedded knowledge graph for diagnosis of radiology imaging," *IEEE Access*, vol. 9, pp. 15 720–15 730, 2021.

[13] X. Jiang, Y. Zhu, G. Cai, B. Zheng, and D. Yang, "Mxt: A new variant of pyramid vision transformer for multi-label chest x-ray image classification," *Cognitive Computation*, vol. 14, no. 4, pp. 1362–1377, 2022.

[14] L. Seyyed-Kalantari, G. Liu, M. McDermott, I. Y. Chen, and M. Ghassemi, "Chexclusion: Fairness gaps in deep chest x-ray classifiers," in *BIOCOMPUTING 2021: proceedings of the Pacific symposium*. World Scientific, 2020, pp. 232–243.

[15] C. Yan, J. Yao, R. Li, Z. Xu, and J. Huang, "Weakly supervised deep learning for thoracic disease classification and localization on chest x-rays," in *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, 2018, pp. 103–110.

[16] L. Luo, L. Yu, H. Chen, Q. Liu, X. Wang, J. Xu, and P.-A. Heng, "Deep mining external imperfect data for chest x-ray disease screening," *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3583–3594, 2020.

[17] A. Kabiraj, T. Meena, P. B. Reddy, and S. Roy, "Detection and classification of lung disease using deep learning architecture from x-ray images," in *Advances in Visual Computing: 17th International Symposium, ISVC 2022, San Diego, CA, USA, October 3–5, 2022, Proceedings, Part I*. Springer, 2022, pp. 444–455.

[18] H.-P. Kriegel, P. Kroger, E. Schubert, and A. Zimek, "Interpreting and unifying outlier scores," in *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 2011, pp. 13–24.

[19] G. Pang, C. Shen, and A. van den Hengel, "Deep anomaly detection with deviation networks," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 353–362.

[20] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1735–1742.

[21] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[22] C. Corbière, N. THOME, A. Bar-Hen, M. Cord, and P. Pérez, "Addressing failure prediction by learning model confidence," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

[23] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[24] Y. Wang, J. Li, and F. Metze, "A comparison of five multiple instance learning pooling functions for sound event detection with weak labeling," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 31–35.

[25] P. O. Pinheiro and R. Collobert, "From image-level to pixel-level labeling with convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1713–1721.

[26] L. Yao, J. Prosky, E. Poblenz, B. Covington, and K. Lyman, "Weakly supervised medical diagnosis and localization from multiple resolutions," *arXiv preprint arXiv:1803.07703*, 2018.

[27] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *International conference on machine learning*. PMLR, 2018, pp. 2127–2136.