Original Article

Development and validation of a self-attention network-based algorithm to detect mediastinal lesions on computed tomography images

Sizhu Wu[1#^], Shengyu Liu[1#^], Ming Zhong[1^], Erik R. de Loos[2], Marc Hartert[3^], Álvaro Fuentes-Martín[4], Alessandra Lenzini[5], Dejian Wang[6^], Qing Qian[1^]

[1]Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China; [2]Division of General Thoracic Surgery, Department of Surgery, Zuyderland Medical Center, Heerlen, The Netherlands; [3]Department of Thoracic Surgery, Katholisches Klinikum Koblenz-Montabaur, Koblenz, Germany; [4]Department of Thoracic Surgery, Hospital Clínico Universitario de Valladolid, Valladolid, Spain; [5]Department of Critical Area and Surgical, Medical and Molecular Pathology, University of Pisa, Pisa, Italy; [6]Department of R&D, Hangzhou Healink Technology, Hangzhou, China

Contributions: (I) Conception and design: Q Qian; (II) Administrative support: D Wang; (III) Provision of study materials or patients: S Wu; (IV) Collection and assembly of data: S Liu; (V) Data analysis and interpretation: M Zhong; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Qing Qian, PhD. Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, 3 Yabao Road, Chaoyang District, Beijing 100020, China. Email: qian.qing@imicams.ac.cn.

Background: Diagnosis of mediastinal lesions on computed tomography (CT) images is challenging for radiologists, as numerous conditions can present as mass-like lesions at this site. This study aimed to develop a self-attention network-based algorithm to detect mediastinal lesions on CT images and to evaluate its efficacy in lesion detection.

Methods: In this study, two separate large-scale open datasets [National Institutes of Health (NIH) DeepLesion and Medical Image Computing and Computer Assisted Intervention (MICCAI) 2022 Mediastinal Lesion Analysis (MELA) Challenge] were collected to develop a self-attention network-based algorithm for mediastinal lesion detection. We enrolled 921 abnormal CT images from the NIH DeepLesion dataset into the pretraining stage and 880 abnormal CT images from the MELA Challenge dataset into the model training and validation stages in a ratio of 8:2 at the patient level. The average precision (AP) and confidence score on lesion detection were evaluated in the validation set. Sensitivity to lesion detection was compared between the faster region-based convolutional neural network (R-CNN) model and the proposed model.

Results: The proposed model achieved an 89.3% AP score in mediastinal lesion detection and could identify comparably large lesions with a high confidence score >0.8. Moreover, the proposed model achieved a performance boost of almost 2% in the competition performance metric (CPM) compared to the faster R-CNN model. In addition, the proposed model can ensure an outstanding sensitivity with a relatively low false-positive rate by setting appropriate threshold values.

Conclusions: The proposed model showed excellent performance in detecting mediastinal lesions on CT. Thus, it can drastically reduce radiologists' workload, improve their performance, and speed up the reporting time in everyday clinical practice.

Keywords: Mediastinal lesions; computed tomography image (CT image); self-attention network

^ ORCID: Sizhu Wu, 0000-0003-4540-9910; Shengyu Liu, 0000-0002-5262-1744; Ming Zhong, 0000-0002-0751-1546; Marc Hartert, 0000-0003-1217-1555; Dejian Wang, 0000-0002-6724-1859; Qing Qian, 0000-0002-9072-586X.

Submitted Apr 24, 2024. Accepted for publication May 17, 2024. Published online May 29, 2024.
doi: 10.21037/jtd-24-679
View this article at: https://dx.doi.org/10.21037/jtd-24-679

Introduction

The mediastinum, located between the two pleural cavities in the thoracic compartment, extends from the sternum to the vertebral column anteroposteriorly and from the superior thoracic aperture to the diaphragm superoinferiorly (1-3). Diagnosing mediastinal lesions presents a complex challenge for pulmonologists, radiologists, and pathologists due to the diverse range of conditions, including non-neoplastic, neoplastic, primary, and metastatic lesions that may appear as mass-like entities in this region (4-6). This complexity often results in significant diagnostic workload, potential delays, workflow interruptions, and an increased misinterpretation rate.

Recent advances in artificial intelligence (AI) have notably enhanced the interpretation of chest radiographs, especially in computed tomography (CT) imaging, which is pivotal in detecting abnormalities. AI improves the diagnostic precision of radiologists by identifying subtle tissue changes and streamlining the detection process, thereby facilitating better evaluation of disease progression, treatment efficacy, and early diagnosis of critical conditions such as lung cancer. Furthermore, AI tools enable more accurate and timely decision-making, enhancing patient care outcomes (7-12). Although various AI algorithms have successfully identified conditions like lung nodules, pneumothorax, and tuberculosis, their effectiveness remains constrained in diagnosing mediastinal lesions, where they have yet to achieve the performance standards of human experts (13-15).

Recognizing the increasing global reliance on low-dose CT (LDCT) for lung cancer screening, which typically does not use intravenous (IV) contrast, highlights an essential area of potential for AI applications. Many patients worldwide undergo LDCT screening for lung cancer, where the ability of AI to accurately detect not only lung nodules but also pathological mediastinal lesions could substantially change the landscape of screening for chest diseases. This capability would avoid the higher costs and radiation exposure associated with standard contrast-enhanced CT scans (16-18).

Our study addresses these challenges by developing a novel self-attention network-based AI algorithm to detect and localize mediastinal lesions across various compartments effectively. By closing existing gaps in lesion detection and significantly enhancing diagnostic accuracy, our approach aims to improve the precision of diagnostic tools, reduce the workload on radiologists, minimize diagnostic errors, and improve the speed and quality of patient care. This comprehensive solution offers a pivotal advancement in medical imaging, particularly in optimizing LDCT for lung cancer screening.

Key innovations of this study include:

- ❖ Advanced self-attention mechanisms: utilizing self-attention layers that analyze the spatial relationships

Highlight box

Key findings

- We developed a self-attention network to detect mediastinal lesions on computed tomography (CT) images.
- Our method has achieved high accuracy, significantly outperforming traditional methods such as convolutional neural networks (CNNs).
- The method has been validated on a large CT image dataset, confirming its applicability in real-world scenarios.

What is known and what is new?

- Deep learning, especially CNNs, has been widely employed in medical image analysis for tumour detection and organ segmentation tasks. These techniques have demonstrated the potential to enhance diagnostic accuracy and alleviate radiologists' workloads. However, they frequently encounter challenges with the complexity and variability of mediastinal lesions, which are compounded by overlapping structures and diverse pathologies.
- This study introduces a novel self-attention network specifically designed for detecting mediastinal lesions on CT images, offering significant improvements over traditional CNNs. Our model enhances detection accuracy and minimizes false positives by concentrating on pertinent features and relationships within the images. Validated on a comprehensive dataset, it demonstrates robustness and generalizability, underscoring its potential for precise and reliable medical imaging applications in real-world clinical environments.

What is the implication, and what should change?

- Accurate detection of lesions is crucial for early diagnosis of various conditions.
- Future research should focus on integrating these advanced diagnostic tools into clinical workflows to enhance early detection and treatment planning.

Table 1 The detailed information of the CT images enrolled in this study

Variable	NIH DeepLesion, pretraining set	MICCAI 2022 MELA		P value
		Training set	Validation set	
Number of CT slices (n)	921	704	176	–
Number of lesions (n)	1,672	707	177	–
Diameter of lesions (min/max, mm)	3/78	13/204	10/201	–
Thickness of lesions (min/max, mm)	0.5/5	0.7/2.5	0.7/2	–
Age (mean ± SD, years)	53.8±16.1	57.1±12.5	55.3±13.9	0.26
Gender (male, %)	59.8	49.2	47.2	0.13

CT, computed tomography; NIH, National Institutes of Health; MICCAI, Medical Image Computing and Computer Assisted Intervention; MELA, Mediastinal Lesion Analysis; SD, standard deviation.

within CT scans to enhance the detection accuracy of complex mediastinal lesions.

- ❖ Specialized mediastinal focus: tailoring AI methodologies specifically to the challenges of mediastinal lesion detection, a step beyond the general focus on lung nodules and masses.
- ❖ Enhanced diagnostic efficiency: the algorithm is designed to integrate seamlessly into clinical workflows, providing real-time analysis and results that expedite clinical decision-making and potentially reduce diagnostic errors.

By pushing the boundaries of AI in medical imaging, this study aims to set new standards in the accuracy and efficiency of mediastinal lesion detection, ultimately improving patient outcomes by enabling earlier and more precise interventions. We present this article in accordance with the TRIPOD reporting checklist (available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-679/rc).

Methods

Data acquisition and lesion annotation

Two separate large-scale open datasets were collected: the National Institutes of Health (NIH) DeepLesion dataset (18) for the pretraining stage and the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2022 Mediastinal Lesion Analysis (MELA) Challenge dataset (19) for the training and validation stages. The NIH DeepLesion dataset included 10,594 abnormal CT images from 4,427 patients accumulated in the NIH Clinical Center's Picture Archiving and Communication Systems (PACS) system (20). Only 921 abnormal CT images with mediastinal lesions were enrolled in the pretraining set. For training and validation, 880 abnormal CT images were collected from the MICCAI Challenge dataset acquired between 2009 and 2020 in an ultra-high volume tertiary hospital (Shanghai Pulmonary Hospital, Shanghai, China) (21). The dataset was randomly split into a training set and a validation set in a ratio of 8:2 at the patient-level. The detailed information on the CT images enrolled in this study is listed in *Table 1*.

Each CT image was reviewed by experienced radiologists who annotated abnormalities with bounding boxes. While our dataset captures a range of mediastinal lesions, it does not specify detailed classifications, such as the density or exact borders of the lesions. This study focuses on detecting these lesions regardless of their specific types, addressing the challenge of detecting well-defined and less distinct abnormalities. In this study, we uniformly transferred the annotations in the format of [left(x), top(y), width, height] for network training and further validation.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective study was approved by Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China review board (IRB number: IMICAMS/02/24/HREC, approval date: 30/01/2024).

Image preprocessing

The size of all CT images was the same (512×512 pixels) with different x-y spacing in the two datasets, and the z-axis pixel spacing ranged from 0.5 to 5 mm and from 0.7 to 2.5 mm, respectively. To effectively distinguish the "pathologic" mediastinal lesion from "normal" mediastinal structures,

we set the minimum and maximum window as –175 and 275 Hounsfield units (HU).

Notably, the image slice thickness can vary across patients and data sources, so trilinear and nearest sampling methods were separately applied to standardize the data to 2 mm in original images and corresponding annotations (bounding box) to avoid device bias. To balance the tradeoff between memory limitation and contextual information, the 3D images fed into the network included only the key slice and one more extended slice in forward and backward directions on the z-axis, which allowed the construction of a fixed size image of 3×512×512 pixels. Moreover, each 3D image was standardized by the min–max approach for denoising and efficient convergence in the training period.

Multiscale feature aggregation

The model utilized a ResNet50-based feature pyramid network (FPN) for feature aggregation (22), incorporating two-pathway feature convolution, upsampling, and connections. In the FPN, the downsampling of feature maps in the first three levels was achieved using a stride of 2 during the convolution operations. This downsampling process reduces the spatial resolution of the feature maps, enabling the network to capture larger receptive fields and more abstract features as it progresses. The subsequent region proposal network (RPN) was designed with three layers. To address potential feature inconsistency issues as network depth increases, dilated convolutions were applied between the last two layers. Dilated convolutions expand the receptive field without reducing spatial resolution, thus maintaining feature integrity and improving the model's ability to learn contextual information. Additionally, the residual blocks (23) employed in the network utilized two types of shortcut connections corresponding to fixed and changing sizes of feature maps. These connections effectively integrate shallow features with deeper ones, enhancing the convergence speed during training and boosting overall network performance by preserving essential spatial information throughout the deeper layers.

Channel-aware attention block (CAAB)

In the model development, the CAAB was adopted to capture the pixel dependence globally for more indicated information in the aggregated features from the backbone network. Specifically, in this approach, pixel relationships are quantitatively measured using the attention feature maps via aggregation of the pixel point with the same weight and suppression with different direction. As shown in *Figure 1*, the input feature map Z is generated to three vectors, *Q*, *K*, and *V*, which represent the height, width, and channel feature, respectively.

$$A = Softmax\left(Q' \times K'\right) \quad [1]$$

$$P_{mn} = \frac{\exp\left(h_m w_n\right)}{\sum_{m=1}^{w} \exp\left(h_m w_n\right)} \quad [2]$$

$$Y = Z + Conv1 \times 1\left(V' \times A\right) \quad [3]$$

where p_{mn}, h_m, and w_n are the value of each pixel on the spatial similarity matrix, the feature vector value of width, and the feature vector value of height, respectively; and Y represents the final output feature map.

Model training and evaluation

The structures of feature aggregation, the RPN, and detection branches in our proposed model were similar to those of the faster region-based convolutional neural network (R-CNN) model (24). The size and ratio of anchors were set as 16, 24, 32, 48, and 96 and as 1:2, 1:1, and 2:1, respectively. The classification and regression heads were used to predict the score (i.e., confidence score) and the location of the detected lesions, respectively, which were calculated by the network's last layer (i.e., fully connected layer). The softmax function was further used to normalize the confidence score as follows:

$$p\left(z_f\right) = \frac{e^{z_f}}{\sum_{c=1}^{C} e^{z_f}} \quad [4]$$

where z_f is the value of the foreground, and C is the number of classes.

For data augmentation, horizontal and vertical flip methods were adopted as general transformers for raw inputs into the network. To ensure better network performance, the pretrained model was first developed on the DeepLesion dataset, which was then fine-tuned by transfer learning on the training set using five-fold cross-validation. For training, the model was developed on two GeForce RTX 2080 Ti GPUs with the PyTorch framework. The learning rate was initially set to 0.001, the decay rate was 0.1 every 20 epochs, the minibatch size was 16 for 500 epochs at maximum, and the early-stopping function was set to 20 consecutive epochs. The report evaluation for our model training

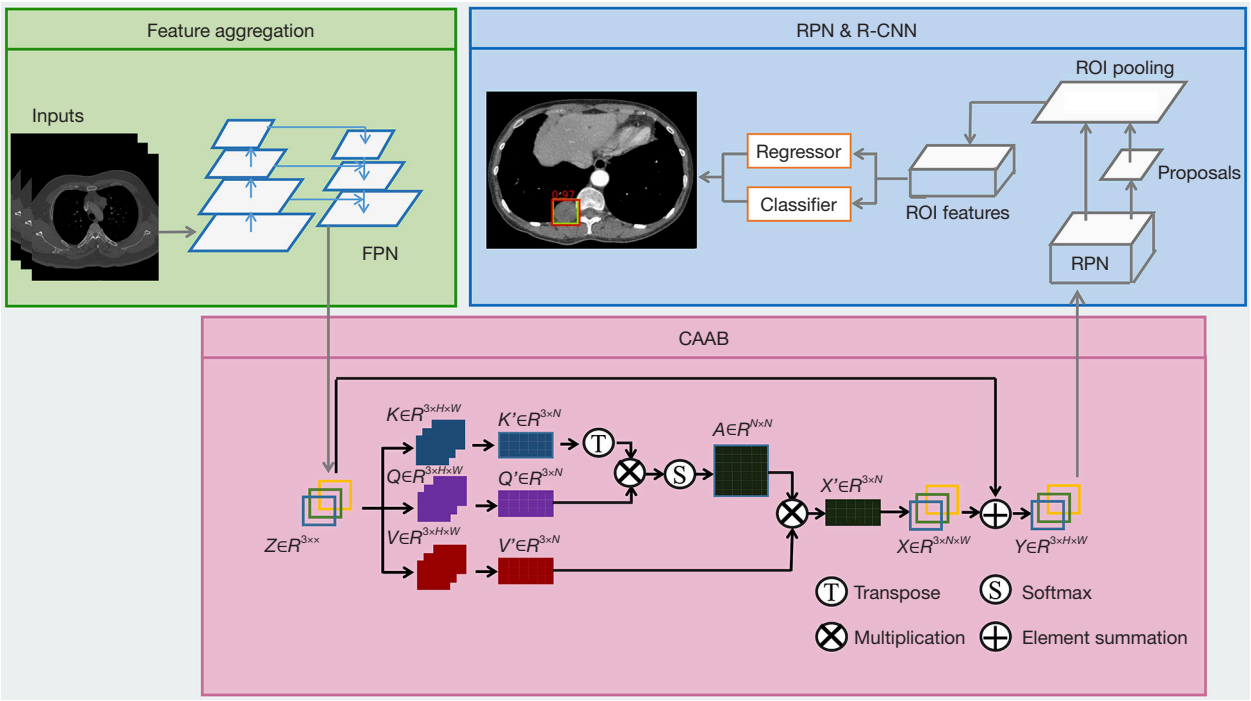


Figure 1 Overview of the proposed network. The red box signifies the prediction of true positives, aligning with the ground truths and indicating the accurately detected lesion locations by the model, while the numerical value in red represents the confidence score of lesion detection. Conversely, the yellow box denotes false positives, indicating regions erroneously labeled as lesions by the model. FPN, feature pyramid network; RPN, region proposal network; R-CNN, region-based convolutional neural network; ROI, region of interest; CAAB, channel-aware attention block.

process was considered to be the value of the Dice similarity coefficient (DSC), which was calculated as follows:

$$DSC(A,B) = 2|A \cap B| / (|A| + |B|) \quad [5]$$

Therefore, the loss function for lesion detection was the following:

$$Loss = -DSC(A,B) \quad [6]$$

Precision was included as the standard metric for the object detection in this study and was calculated as follows:

$$Precision = TP/(TP + FP) \quad [7]$$

where TP is true positive, and FP is false positive, with TP and FP representing the number of correct positive predictions and the number of incorrect positive predictions with respect to the ground truth (GT), respectively. To quantitatively evaluated multiple lesions per image, the average sensitivity (AS) at several FPs as calculated by different thresholds was defined. In this study, we only evaluated AS at six values of FPs (0.25, 0.5, 1, 2, 3, and 4), considering that images with one lesion accounted for the majority of the enrolled data. Sensitivity was calculated as follows:

$$Sensitivity = TP/(TP + FN) \quad [8]$$

where FN is false negative, representing the number of negative incorrect predictions with respect to the GT. The free-response receiver operating characteristic (FROC) curve was defined to determine the value of AS in relation to the different numbers of FPs per image. Furthermore, a competition performance metric (CPM) was used to evaluate the average level of sensitivities from the six FP rates.

Statistical analysis

The proposed model was implemented using PyTorch (version 1.7.1). All statistical analyses were conducted with R version 3.5.3 (The R Foundation for Statistical Computing, Vienna, Austria). The Student's *t*-test and Chi-squared (x^2) test were used for continuous and categorical

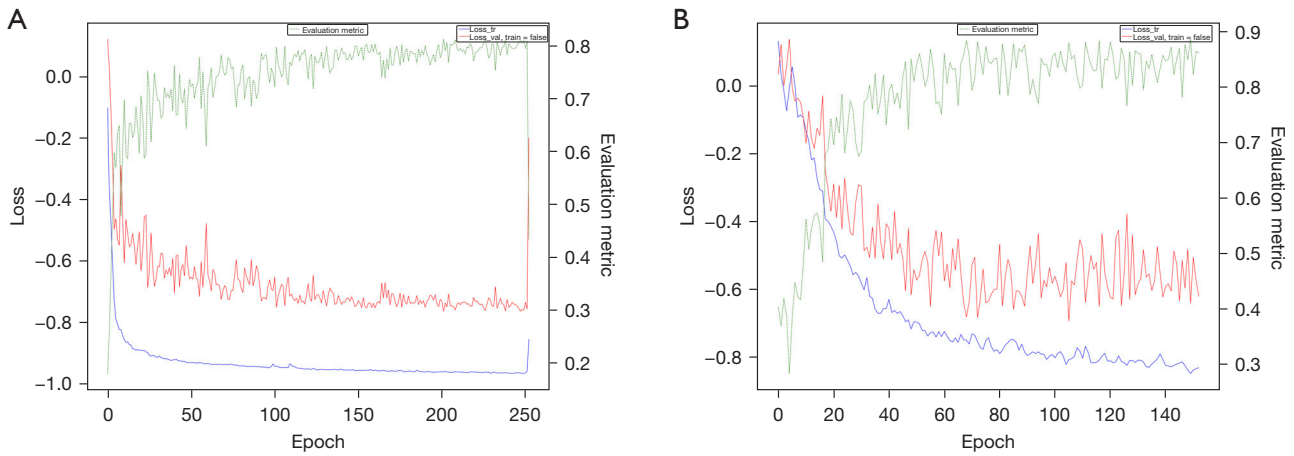


Figure 2 Model training and evaluation. (A) Pre-training stage using the DeepLesion dataset. The blue line represents the training loss, indicating how the model's learning progresses with the training data. The red line shows the validation loss, reflecting the model's performance on unseen data from the same dataset. The green line depicts the evaluation metric—specifically, the AP—achieved by the model on the validation set. The AP measures the model's accuracy in identifying positive instances (such as lesions) while minimizing false positives. A higher AP signifies superior performance in terms of both precision and recall. (B) Training stage using the MICCAI 2022 MELA challenge dataset. As in (A), the blue and red lines represent training and validation losses, respectively. The green line illustrates the evaluation metric, AP, on the validation set, demonstrating the model's generalization ability on new data from the MELA challenge. tr, training set; val, validation set; AP, average precision; MICCAI, Medical Image Computing and Computer Assisted Intervention; MELA, Mediastinal Lesion Analysis.

data respectively, and a P value <0.05 indicated a statistically significant difference.

Results

Average precision (AP) of the proposed model in mediastinal lesion detection

During the pretraining stage with the DeepLesion dataset, the model demonstrated optimal performance around 250 epochs, achieving an 82.2% AP score in mediastinal lesion detection before gradually entering an overfitting stage (*Figure 2A*). In the training stage with the MICCAI 2022 MELA Challenge dataset, the proposed model achieved rapid convergence, resulting in reduced training and validation loss within approximately 100 epochs, and attained an 89.3% AP score in mediastinal lesion detection (*Figure 2B*). The model consistently identifies lesions larger than 10 mm with high confidence; however, sensitivity decreases for smaller lesions due to the intricate anatomical structures within the mediastinum and the subtle nature of these lesions. This variability in sensitivity underscores the importance of lesion size and contrast in optimizing detection performance.

Confidence score of the proposed model in mediastinal lesion detection

In this study, we found that our proposed model could identify comparable large lesions with high confidence scores (over the value of 0.8), indicating its ability to discern lesions (*Figure 3A*) effectively. Moreover, the intersection-over-union (IoU) threshold was set to 0.5, which means the candidates with overlapping areas between themselves and corresponding GTs >0.5 were considered TPs, making the area of TPs as large as possible. In contrast, for some lesions, detection was challenging due to the influence of similar shape, location, or texture with that of GTs or due to their extremely small area (*Figure 3B*).

Sensitivity comparison between the faster R-CNN model and the proposed model

The FROC curve was plotted to compare the sensitivity between the faster R-CNN model (a classical two-stage object detection algorithm) and our proposed model in the validation set. As shown in *Figure 4A* and *Table 2*, our proposed model achieved a performance increase of almost 2% at the level of CPM compared to the faster R-CNN

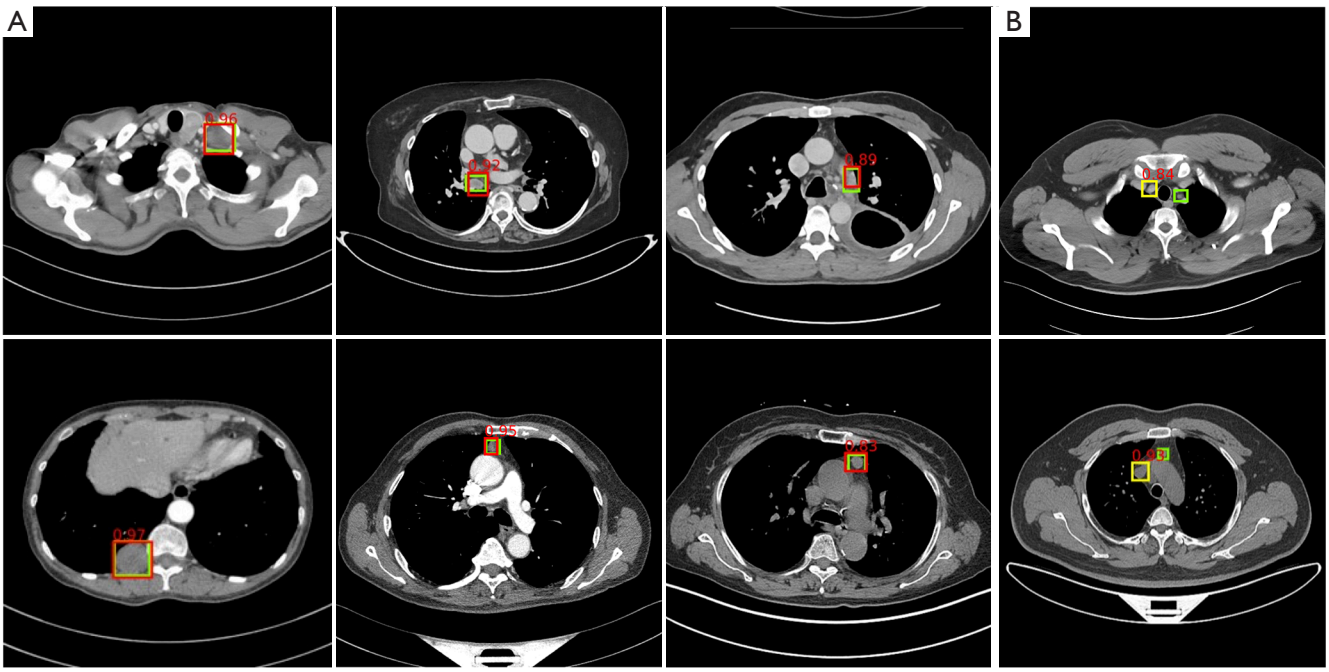
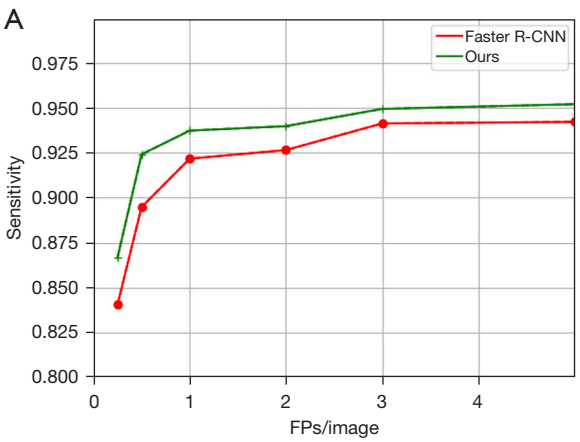


Figure 3 Representative images of CT scanning, detection, and annotation. (A) Representative images of true-positive predictions (first row: DeepLesion; second row: MICCAI 2022 MELA). Green boxes correspond to ground truths, red boxes correspond to true positives, and yellow boxes indicate false positives. The number in red is the confidence score for the lesion detection. (B) Representative images of false-positive predictions (first row: DeepLesion; second row: MICCAI 2022 MELA). Green boxes correspond to ground truths, and yellow boxes correspond to false positives. CT, computed tomography; MICCAI, Medical Image Computing and Computer-Assisted Intervention; MELA, Mediastinal Lesion Analysis.

A



B

Threshold	FPs/image	Sensitivity (%)
0.02	39.30	97.52
0.03	10.25	96.41
0.04	3.89	95.12
0.05	1.02	93.95
0.06	0.11	81.07

Figure 4 Sensitivity comparison between the faster R-CNN model and the proposed model. (A) The FROC curve for lesion detection with the faster R-CNN model and our proposed model in the validation set. (B) Sensitivity evaluation under various thresholds. R-CNN, region-based convolutional neural network; FP, false positive; FROC, free-response receiver operating characteristic.

Table 2 The sensitivity comparison between the faster R-CNN and the proposed model

Model	FPs per image						
	0.25	0.5	1	2	3	4	CPM
Faster R-CNN, 3 slices	84.05	89.48	92.17	92.65	94.13	94.22	91.12
Proposed model (with channel-aware attention), 3 slices	86.67	92.41†	93.73†	93.98†	94.94†	95.20†	92.82†

†, the best score. R-CNN, region-based convolutional neural network; FP, false positive; CPM, competition performance metric.

model. In addition, we observed that the sensitivity was higher than that of the faster R-CNN model and was still over 92.4% at a cost as low as 0.5 FPs/image.

It is considerably challenging to balance the tradeoff between sensitivity and FPs. As shown in *Figure 4B*, the value of 0.05 seems to be ideal cut-off for mediastinal lesion detection model in this study, ensuring an outstanding sensitivity (93.95%) with a comparably low FP rate (1.02 FPs/image).

Discussion

The emergence of AI, particularly through the utilization of advanced algorithms, represents a substantial evolution in diagnostic radiology. The incorporation of AI brings a significant advantage by reducing the workload of radiologists, thus enabling them to devote more attention to complex cases and delicate aspects of patient care. In this study, we have designed a self-attention network-based algorithm capable of accurately detecting and localizing multiple mediastinal lesions on chest radiographs. Additionally, we have evaluated its diagnostic accuracy. With a high FROC score and precise lesion detection, the model demonstrates the potential of AI to facilitate early diagnosis and enhance treatment, ultimately improving patient outcomes. Our model achieved a remarkable performance improvement of nearly 2% at the CPM level in detecting mediastinal lesions compared to the faster R-CNN model. While this numerical improvement may appear small, its practical significance in clinical applications is considerable. This advancement suggests that a broader range of patients could benefit from timely and precise early diagnoses, potentially leading to better treatment results. Moreover, improved performance decreases the chances of misdiagnoses and overlooked conditions, essential for patient satisfaction and alleviating the healthcare system's burden. This underscores the profound impact of even incremental technological progress on clinical practice. There are two strengths in our proposed model. First, our model was pre-trained with the DeepLesion dataset and then trained with the MICCAI 2022 MELA Challenge dataset, which helped to improve the performance. An attention mechanism is needed to make the network focus on the most salient feature maps to determine the feature space on the z-axis with limited slices. The self-attention (25) method is useful for capturing the rich contextual relationships in the feature space. Second, our proposed model was adopted. CAAB, which facilitated the localization of most of the mediastinal lesions with high precision.

The model's integration into clinical settings is poised to enhance the radiological interpretation of CT scans, potentially increasing diagnostic accuracy and reduce radiologists' workload. By automating initial screenings and enhancing the detection of diverse lesion characteristics, the model improves accuracy and supports more customized diagnosis and management plans. This technological advancement is especially valuable in remote and resource-limited environments, where it is a crucial support tool, aiding healthcare providers in validating diagnostic impressions and making informed decisions. Designed to augment, not replace, radiological expertise, the model enhances workflow and allows radiologists to focus on complex diagnostic tasks. Crucial to its success is the seamless integration of this AI into clinical workflows, featuring adaptable and user-friendly systems that align with clinical needs and improve patient care efficiency and quality.

The model showed high confidence in identifying large lesions with distinct features such as strong contrast and clear borders. However, the retrospective nature of our validation dataset introduces potential selection bias, limiting the model's generalizability. More critically, excluding normal imaging from the training dataset impairs the model's ability to distinguish between healthy tissue and lesions, likely increasing the rate of false positives and reducing clinical utility.

The model's sensitivity in detecting mediastinal lesions is significantly affected by lesion size and contrast with

surrounding tissues. The model demonstrates high confidence in identifying lesions larger than 10 mm. However, for smaller lesions, sensitivity decreases due to the complex anatomy of the mediastinum and the subtle presentation of these lesions. Clinically, our analysis indicates that many detected lesions are significant and can influence patient management. The model effectively identifies critical pathological conditions, such as malignant tumors and enlarged lymph nodes, essential for staging and treatment planning. Conversely, while the model accurately detects benign conditions like simple cysts, these findings may not substantially alter clinical management. Nonetheless, the model's ability to exclude serious conditions provides reassurance and can reduce the need for invasive diagnostics.

Furthermore, the study acknowledges inherent biases in the training data, which could lead to diagnostic inaccuracies and potentially raise ethical and legal concerns, especially in misdiagnosis cases. Dependence on biased AI systems might diminish medical practitioners' diagnostic skills over time, highlighting the need for a balanced approach that integrates technology while preserving professional expertise. Ethical considerations, particularly regarding privacy and AI transparency, are crucial as AI becomes increasingly prevalent in clinical settings. Addressing these issues requires robust information technology (IT) infrastructure, continual radiologist training, and clear guidelines for AI integration into clinical decisions.

Although our AI model shows promise in detecting mediastinal lesions, its limitations underscore the need for thoughtful future development. This includes enriching the dataset to include more diverse and rare lesion types, addressing inherent biases, enhancing the interpretability of AI decisions, and evaluating the impact of contrast-enhanced imaging on diagnostic performance. We will also implement rigorous model evaluation metrics and validation protocols to ensure the improvements translate into clinically meaningful outcomes.

Future research will expand the training and validation datasets to improve diagnostic precision. This expansion will incorporate diverse lesion characteristics such as density, borders, and contrast enhancement. By doing so, the AI can more accurately classify lesions into clinically relevant categories, enhancing diagnostic specificity. Furthermore, improving our model's ability to distinguish between clinically significant and insignificant lesions will maximize its practical utility in clinical settings. Continuous refinement of AI algorithms, guided by radiologist feedback, is essential. This process will help align the models with clinical needs and enable learning from real-world data.

Moreover, we plan to further integrate advanced deep learning architectures, such as transformers and graph neural networks, to enhance our model's detection and classification capabilities. These state-of-the-art models have demonstrated superior performance in various computer vision tasks and promise to improve the robustness and accuracy of medical image analysis.

Our future goal is to further develop a reliable, ethical, and effective diagnostic model that augments radiologists' expertise, thus improving patient outcomes while preserving the essential human element in healthcare. By systematically addressing these challenges and incorporating advanced models and evaluation methods, we aspire to create a robust diagnostic tool that meets the evolving demands of clinical practice.

Conclusions

In conclusion, our proposed AI model showed excellent performance in the detection of mediastinal lesions on CT images and has the potential to drastically reduce the workload of radiologists, improve their performance, and speed up the reporting time in real-world situations.

Acknowledgments

Funding: This study was funded by the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences Program (grant No. 2021-I2M-1-057).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-679/rc

Peer Review File: Available at: https://jtd.amegroups.com/article/view/10.21037/jtd-24-679/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-679/coif). D.W. is from Hangzhou Healink Technology. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all

aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective study was approved by Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China review board (IRB number: IMICAMS/02/24/HREC, approval date: 30/01/2024), and informed consent was obtained from all individual participants.

References

1. Shabb NS, Fahl M, Shabb B, et al. Fine-needle aspiration of the mediastinum: a clinical, radiologic, cytologic, and histologic study of 42 cases. Diagn Cytopathol 1998;19:428-36.
2. Duwe BV, Sterman DH, Musani AI. Tumors of the mediastinum. Chest 2005;128:2893-909.
3. Shaheen MZ, Sardar K, Murtaza HG, et al. CT guided trans-thoracic fine needle aspiration/biopsy of mediastinal and hilar mass lesions: An experience of pulmonology department at a tertiary care teaching hospital. Pak J Chest Med 2008;12:26-38.
4. Takahashi K, Al-Janabi NJ. Computed tomography and magnetic resonance imaging of mediastinal tumors. J Magn Reson Imaging 2010;32:1325-39.
5. Juanpere S, Cañete N, Ortuño P, et al. A diagnostic approach to the mediastinal masses. Insights Imaging 2013;4:29-52.
6. Thacker PG, Mahani MG, Heider A, et al. Imaging Evaluation of Mediastinal Masses in Children and Adults: Practical Diagnostic Approach Based on A New Classification System. J Thorac Imaging 2015;30:247-67.
7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
8. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology 2017;284:574-82.
9. Nam JG, Park S, Hwang EJ, et al. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. Radiology 2019;290:218-28.
10. Hwang EJ, Park S, Jin KN, et al. Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. Clin Infect Dis 2019;69:739-47.
11. Park S, Lee SM, Kim N, et al. Application of deep learning-based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. Eur Radiol 2019;29:5341-8.
12. Hwang EJ, Hong JH, Lee KH, et al. Deep learning algorithm for surveillance of pneumothorax after lung biopsy: a multicenter diagnostic cohort study. Eur Radiol 2020;30:3660-71.
13. Kim Y, Park JY, Hwang EJ, et al. Applications of artificial intelligence in the thorax: a narrative review focusing on thoracic radiology. J Thorac Dis 2021;13:6943-62.
14. Fanni SC, Marcucci A, Volpi F, et al. Artificial Intelligence-Based Software with CE Mark for Chest X-ray Interpretation: Opportunities and Challenges. Diagnostics (Basel) 2023;13:2020.
15. Silva M, Picozzi G, Sverzellati N, et al. Low-dose CT for lung cancer screening: position paper from the Italian college of thoracic radiology. Radiol Med 2022;127:543-59.
16. Dickson JL, Horst C, Nair A, et al. Hesitancy around low-dose CT screening for lung cancer. Ann Oncol 2022;33:34-41.
17. Henschke C, Huber R, Jiang L, et al. Perspective on Management of Low-Dose Computed Tomography Findings on Low-Dose Computed Tomography Examinations for Lung Cancer Screening. From the International Association for the Study of Lung Cancer Early Detection and Screening Committee. J Thorac Oncol 2024;19:565-80.
18. Yan K, Wang X, Lu L, et al. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J Med Imaging (Bellingham) 2018;5:036501.
19. Chen C. 2022. MICCAI 2022 mela challenge: Mediastinal lesion analysis. Available online: https://mela.grand-challenge.org/Dataset/
20. Yan K, Wang X, Lu L, et al. DeepLesion: Automated

Deep Mining, Categorization and Detection of Significant Radiology Image Findings using Large-Scale Clinical Lesion Annotations. arXiv 2017. doi: 10.48550/arXiv.1710.01766.
21. Wang J, Ji X, Zhao M, et al. Size-adaptive mediastinal multilesion detection in chest CT images via deep learning and a benchmark dataset. Med Phys 2022;49:7222-36.
22. Lin TY, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI; 2017:936-44.
23. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. arXiv 2015. doi: 10.48550/arXiv.1512.03385.
24. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137-49.
25. Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need. Part of Advances in Neural Information Processing Systems 30 (NIPS 2017). Accessed: Nov 26, 2022. Available online: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Cite this article as: Wu S, Liu S, Zhong M, de Loos ER, Hartert M, Fuentes-Martín Á, Lenzini A, Wang D, Qian Q. Development and validation of a self-attention network-based algorithm to detect mediastinal lesions on computed tomography images. J Thorac Dis 2024;16(5):3306-3316. doi: 10.21037/jtd-24-679