

RESEARCH

Open Access



A deep learning model combining circulating tumor cells and radiological features in the multi-classification of mediastinal lesions in comparison with thoracic surgeons: a large-scale retrospective study

Feng Wang^{1†}, Minwei Bao^{2†}, Bo Tao², Fugui Yang³, Guangxue Wang⁴ and Lei Zhu^{2*}

Abstract

Background CT images and circulating tumor cells (CTCs) are indispensable for diagnosing the mediastinal lesions by providing radiological and intra-tumoral information. This study aimed to develop and validate a deep multimodal fusion network (DMFN) combining CTCs and CT images for the multi-classification of mediastinal lesions.

Methods In this retrospective diagnostic study, we enrolled 1074 patients with 1500 enhanced CT images and 1074 CTCs results between Jan 1, 2020, and Dec 31, 2023. Patients were divided into the training cohort ($n = 434$), validation cohort ($n = 288$), and test cohort ($n = 352$). The DMFN and monomodal convolutional neural network (CNN) models were developed and validated using the CT images and CTCs results. The diagnostic performances of DMFN and monomodal CNN models were based on the Paraffin-embedded pathologies from surgical tissues. The predictive abilities were compared with thoracic resident physicians, attending physicians, and chief physicians by the area under the receiver operating characteristic (ROC) curve, and diagnostic results were visualized in the heatmap.

Results For binary classification, the predictive performances of DMFN (AUC = 0.941, 95% CI 0.901–0.982) were better than the monomodal CNN model (AUC = 0.710, 95% CI 0.664–0.756). In addition, the DMFN model achieved better predictive performances than the thoracic chief physicians, attending physicians, and resident physicians ($P = 0.054$, 0.020, 0.016) respectively. For the multiclassification, the DMFN achieved encouraging predictive abilities (AUC = 0.884, 95%CI 0.837–0.931), significantly outperforming the monomodal CNN (AUC = 0.722, 95%CI 0.705–0.739), also better than the chief physicians (AUC = 0.787, 95%CI 0.714–0.862), attending physicians (AUC = 0.632, 95%CI 0.612–0.654), and resident physicians (AUC = 0.541, 95%CI 0.508–0.574).

Conclusions This study showed the feasibility and effectiveness of CNN model combining CT images and CTCs levels in predicting the diagnosis of mediastinal lesions. It could serve as a useful method to assist thoracic surgeons in improving diagnostic accuracy and has the potential to make management decisions.

Keywords Mediastinal tumor, Convolutional neural network, Diagnosis, Circulating tumor cell

[†]Feng Wang and Minwei Bao contributed equally to this work.

*Correspondence:

Lei Zhu

zhuleiZlei@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Mediastinal lesions are a broad-spectrum disease consisting of either benign or malignant tumors [1]. The diagnosis of mediastinal lesions should be established based on the pathology [2]. However, due to the relatively challenging nature of the region, the diagnosis and treatment have traditionally been mainly based on empirical experiences and clinical auxiliary examinations according to the National Comprehensive Cancer Network (NCCN) guidelines [3]. In addition, the therapy and prognosis vary among distinct kinds of mediastinal lesions. For example, complete thymectomy is the standard treatment for thymoma at any stage, except for stage I thymoma. For patients with stage I thymoma, thymomectomy might be the more reasonable surgical option [4, 5]. Therefore, a simple, accurate, and feasible method for differentiating mediastinal lesions is imperative, especially for the treatment strategy and clinical management.

While the histopathological examination is the gold diagnosis standard, the process is invasive, risky, and expensive. To avoid unnecessary operations, preoperative enhanced CT provides adequate quantitative information on the lesions, such as size, density, vessels, and adjacent structures. Owing to its better visibility, versatility, and timeliness, enhanced CT has been increasingly considered as an ideal diagnostic imaging method for mediastinal lesions, although it has been proven that CT combined with other methods could improve the accuracy of chest diseases at or above the level of thoracic surgeons [6, 7]. However, the accuracy of CT alone remains limited, particularly in the early detection of small or atypical lesions. In addition, the complexity of enhanced CT for primary surgeons to monitor and accurately identify different mediastinal lesion types could not be ignored.

Recently, liquid biopsy has drastically revolutionized the field of precision oncology, offering personalized molecular genomic data. Its non-invasive nature also paved a new era in thymoma [8, 9]. The combination of CT imaging and CTCs in a multimodal approach has shown promise in enhancing diagnostic accuracy [10, 11]. Studies have indicated that integrating radiological features with molecular biomarkers can improve cancer detection and prognostication. For example, radiomics, which extracts quantitative features from medical images, has been combined with CTC analysis in several studies to enhance the performance of cancer detection systems [12]. Despite these advances, most methods rely on either CT imaging or CTCs for diagnostic purposes. Additionally, CT imaging often requires advanced post-processing or fusion techniques to achieve high diagnostic accuracy, and CTC detection methods can be expensive and technically challenging [12, 13].

Recently, a growing number of studies have applied deep learning to classify thymoma and achieved excellent diagnostic accuracy [14, 15]. However, the integration of deep learning models with molecular data such as CTCs is still in its early stages. Additionally, these studies mainly focused on the specific disease (i.e., thymoma) and they were mostly used for the binary classification of mediastinal lesions (i.e., benign vs malignant). On the other hand, their diagnostic performance in examining the broad spectrum of mediastinal lesions was insufficient.

This study introduces a novel deep multimodal fusion network (DMFN) that integrates CTCs and CT imaging for the multi-classification of mediastinal tumors, providing a more comprehensive approach to diagnosis. Our model differs significantly from previous studies by combining radiological and molecular data, addressing the limitations of current diagnostic methods. Unlike traditional monomodal convolutional neural network (CNN) models, which rely on either CT images or CTCs alone, our approach combines both datasets to improve diagnostic accuracy. We demonstrate that the DMFN model significantly outperforms both monomodal CNN models and clinical experts, as evidenced by superior AUC scores and diagnostic performance. This work represents a significant step forward in the integration of multimodal data for cancer diagnosis and has the potential to transform the clinical management of mediastinal tumors.

Methods

Study cohort

This is a retrospective and single-center study. The inclusion criteria were as follows: (1) complete clinical basic information, chest-enhanced CT images, and circulating tumor cells within 2 weeks after admission. (2) Initial treatments including surgery, and neoadjuvant chemoradiotherapy were not received. (3) No distant metastasis or multiple primary cancers. (4) Patients aged 18 years or older. (5) Follow-up: Patients with sufficient follow-up data or surgical intervention records. Exclusion criteria: (1) Patients under 18 years old. (2) Inadequate imaging: patients with poor-quality CT images (e.g., significant motion artifacts, incomplete scans) were excluded. (3) Incomplete data: Patients lacking CTCs data or follow-up records or surgical records. (4) Non-mediastinal lesions: Patients with lesions located outside the mediastinal region were excluded. (5) Coexisting medical conditions: Patients with other major coexisting conditions (e.g., metastases to the mediastinum) that could complicate the diagnosis were excluded. This study was compliant with the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guidelines.

We initially identified 2000 patients from the doctor workstation system who had CT scans taken between January 1, 2020, and December 31, 2023. After the inclusion and exclusion criteria, a total of 1074 patients whose data met all requirements. The study flow chart is provided in Fig. 1.

Our study was approved by the ethics committee of Shanghai Pulmonary Hospital, School of Medicine, Tongji University. Informed consent was waived because of the retrospective nature.

CT examination and image preprocessing

All patients in the cohort underwent chest-enhanced CT using Siemens Somatom Definition AS scanners (Siemens Medical Systems, Erlangen, Germany) or Philips Brilliance 40 scanners (Philips Healthcare, Cleveland, USA) within 2 weeks after admission. After intravenous injection of iodine contrast agent with a 60–65-s delay, CT images were standardized with a window width of 350 and a window level of 50. Mathematical algorithms and detector advancements have helped to reduce noises and reconstruct CT images at 1-mm slice intervals. Two thoracic surgeons draw the regions of interest (ROI) in the reconstructed images by manually segmenting them. Through objective and quantitative image analysis and clinical experience, thoracic surgeons make differential diagnoses of the delineated lesions.

Plasma CTCs testing

Venous blood samples were collected from patients on the day of admission, and the blood samples were

stored in the CellSave tubes. Then, the samples were centrifugated at 800 g for 10 min. After centrifugation, the residual supernatant was removed. CTCs were detected through the Immunomagnetic bead sorting combined with the immunofluorescence staining method, and the data were obtained using the CellSearch Circulating Tumor Cell Kit (GENO, 20,163,400,061, China).

Pathological ground truth

The diagnosis of mediastinal nodules was confirmed based on histological pathology results [16]. Benign diseases included cysts, lipoma, benign thymoma, mature teratoma, cystic lymphangioma, hemangioma, fibroma, and neuroma. Malignant diseases included malignant thymoma, sarcoma, neuroblastoma, immature teratoma, lymphoma, and metastatic tumor. The frequencies of mediastinal nodules diagnosed in the training, validation, and test cohort were summarized in Table 1.

Data preprocessing

In the chest-enhanced CT image processing stage, we first used an easy fourth-order partial differential equation diffusion model in the noise reduction process to solve the problem of blurred edges and texture details. And then we extract the region of interest separately from the CT fault images. Due to the fact that the spacing between tomographic scans is usually larger than the pixel size of two-dimensional images, cubic voxel data with consistent directions are generated in three-dimensional space, which can outline the three-dimensional shape of the measured tissues. Based on the reconstructed 3D image,

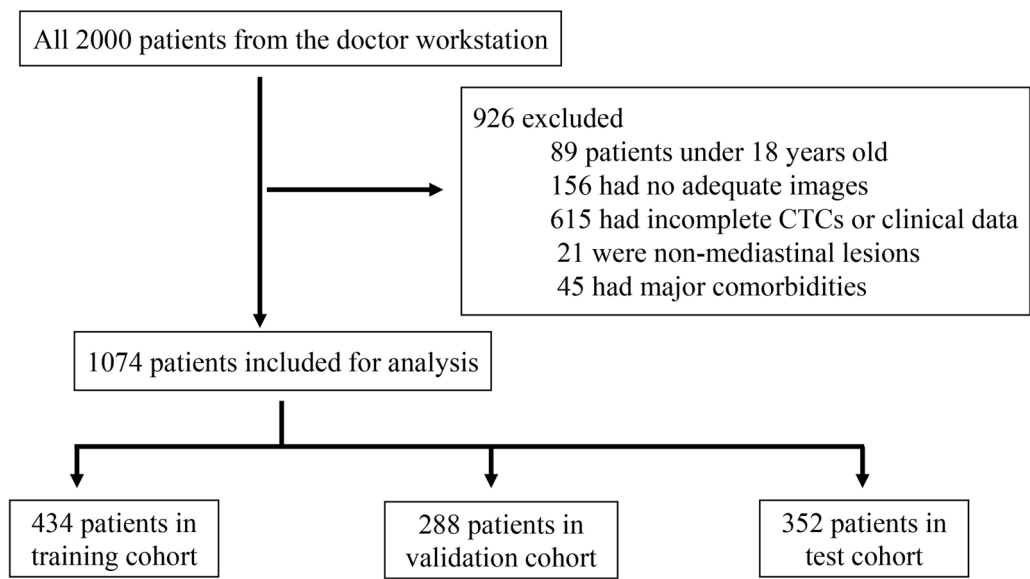


Fig. 1 The study flow chart of the study

Table 1 Summary of diagnosis in the training, validation, and test cohorts $n = 1074$

Characteristics	Training cohort $n = 434$	Validation cohort $n = 288$	Test cohort $n = 352$
Patient demographics			
Age, years (mean \pm SD)	53.81 \pm 12.42	52.41 \pm 13.63	53.76 \pm 12.88
Sex (n , %)			
Male	197(45.39%)	127(44.10%)	162(46.02%)
Female	237(54.61%)	161(55.90%)	190(53.98%)
Lesion localization			
Anterior	297(68.43%)	211(73.26%)	275(78.13%)
Middle	94(21.66%)	50(17.36%)	54(15.34%)
Posterior	43(9.91%)	27(9.38%)	23(6.53%)
CTCs level (pg/ml)	12.48 \pm 5.23	11.02 \pm 4.62	15.19 \pm 6.58

CTCs Circulating tumor cells

we use various geometric transformation operations to achieve measurement of different projection imaging methods and indicators such as volume, area, and thickness. We complete the reconstruction of different sections, display views in different directions, determine the nature of lesions, and proposed management plans.

Deep multimodal fusion network model development

To better help the machine classifier distinguish the characteristics of mediastinal lesions, we utilized the 3D-Slicer software to segment the mediastinal region. Next, we cropped and extracted the lesions from the CT images. Subsequently, these extracted images were prepared for the classification tasks based on the prior knowledge and the segmentation mask.

To integrate the internal and external information, we developed a deep multimodal fusion network (DMFN) model that combined the mediastinal lesions' appearance, CTC data, and CT images for deep feature fusion. The DMFN input contains various data from three modal, including clinical close-up manifestation, CT images, and 3D-Slicer images. The CT images use a 2D convolutional neural network (CNN) to process CT scan images. And this architecture consists of several convolutional layers followed by max-pooling layers to down-sample the image and retain important features. The CTC data were processed by the fully connected neural network, which consists of dense layers that learn the relationships between different levels of CTC-related features. DMFN will extract multi-layer data from the above aspects, and each layer of metrics will be extracted, fused, and preserved. The monomodal CNN typically processes a single modality of data, such as CT images, using convolutional layers for feature extraction, followed by pooling and fully connected layers for classification. We used

the monomodal CNN model to generate a similar training result for comparison to DMFN [17, 18].

In this study, we used the Regulated Networks (RegNet) to predict the pathology of the mediastinal lesions, which was incorporated as a backbone architecture within our DMFN model. It consists of convolutional blocks, each containing a set of convolutional filters followed by batch normalization and activation functions. Specifically, we utilized RegNet for feature extraction from CT images. Meanwhile, we used the neutral cross-entropy loss to mitigate the over-sharpness of entropy minimization. The architecture of the DMFN and RegNet models are summarized in Fig. 2. In the statistical process, the output outcomes were the top 3 prediction results. We used the top 1 as the optimal model and the performance was repeated 3 times. The statistical methods and algorithms explanations are cited from Prof. Xu J study [19].

Performance evaluation of CNN model and thoracic surgeons

Experienced thoracic surgeons were divided into three groups, including resident physicians ($n = 2$), attending physicians ($n = 2$), and chief physicians ($n = 2$). The resident physicians have 1–3 years of clinical experience. They mainly focused on the patients' manifestations and CT images, initial patient assessments, and routine monitoring, under the supervision of more experienced physicians and they are mainly responsible for the binary classification of mediastinal lesions, for example, benign vs malignant. The attending physicians usually have 5 or more years of clinical experience. They were responsible for image interpretation, detailed diagnostic assessments, and contributing to treatment planning, and were asked to verify the accuracy of classification and proposed therapy strategies, for example, surgery or not, follow-ups. The chief physicians are experts in the field

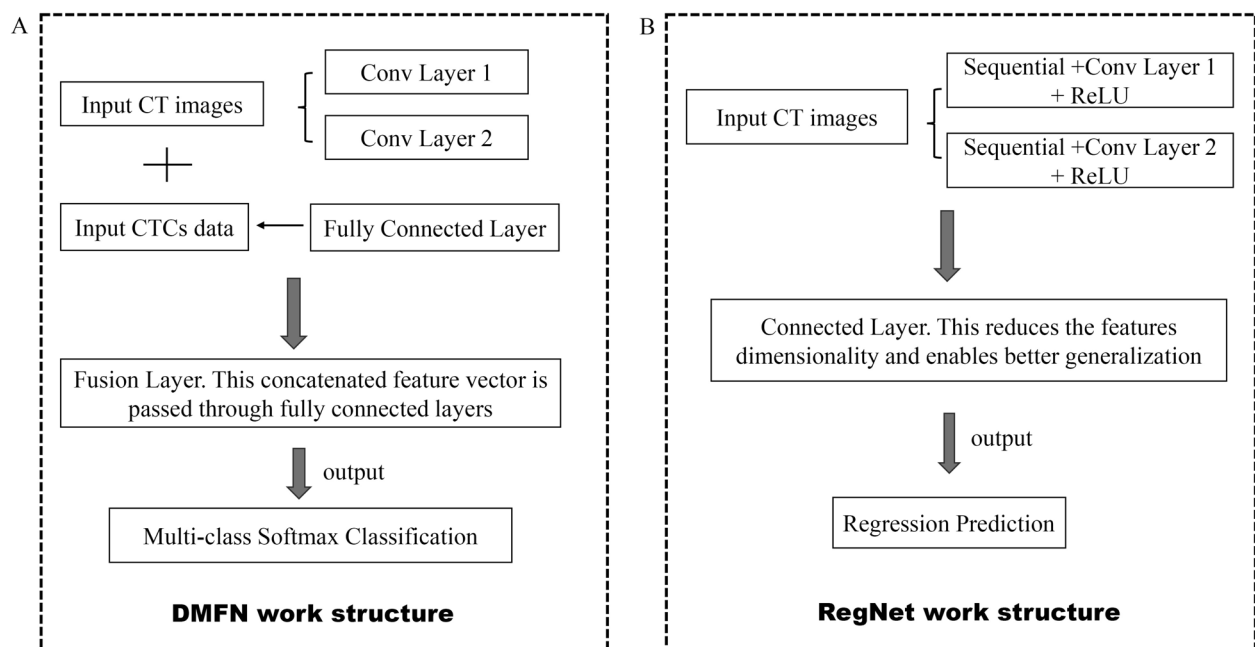


Fig. 2 The structures of DMFN and RegNet models. **A** DMFN model: Combines CT image features and CTC data through convolutional and fully connected layers, fusing them for multi-class classification via a softmax layer. **B** RegNet model: Extracts features from CT images using sequential convolutional layers with dimensionality reduction to enhance generalization and improve feature extraction efficiency

with 10 or more years of experience. They took responsibility for final clinical decisions, particularly in complex or ambiguous cases. They provided guidance and supervision for the entire medical team and ensured that the final treatment plans aligned with national clinical guidelines and best practices. Management decisions (e.g., surgery, follow up or conservative treatment) are established according to the clinical guidelines (NCCN and ESMO guidelines), multidisciplinary team, and individualized treatment. All physicians were asked to make management decisions and performed the work in a blinded manner to patients' information. Disputes and inconsistencies are resolved through discussion. At last, the results of diagnostic performance were compared between thoracic physicians and CNN.

Statistical analysis

Chi-square and Mann–Whitney U were used to compare categorical variables and continuous variables, respectively. Receiver operating characteristic (ROC) curves and the area under the ROC curves (AUC) were calculated to assess the diagnostic performances of CNN and thoracic physicians. Subsequently, the sensitivity, specificity, positive, predictive value (PPV), negative predictive value (NPV), and accuracy were also calculated to evaluate the diagnostic performances. The DeLong test and McNemar's test were used to compare the AUCs and sensitivity and specificity. Differences were considered

statistically significant at 2-sided $P < 0.05$. All the data were calculated using the R software (version 4.3.3).

Results

Patients' information and characteristics of mediastinal lesions

A total of 1074 patients were included in the study. The average age of the population was 53.18 ± 13.29 , and 501 (46.65%) were male. Among the patients, 434 patients were in the training cohort, and 288 patients were in the validation cohort. In addition, 352 patients were in the test cohort. This study produces 1500 enhanced CT slices, with 596 from the training cohort, 419 from the validation cohort, and 485 from the test cohort. The most frequent mediastinal lesions were thymoma (34.73%, 373), followed by cyst (28.49%, 306), Schwannoma (7.73%, 83), and thymic squamous cell carcinoma (5.49%, 59). Detailed information is shown in Table 2.

Diagnostic performance of CNN models

In the validation cohort, a total of 288 patients were collected for analysis. For the differential diagnosis of benign and malignant mediastinal lesions, the performance of DMFN combined with CT images and CTCs was superior to the monomodal CNN model both in AUC, sensitivity, and accuracy. The AUC of DMFN combined with CT images and clinical close-up manifestations and monomodal CNN were 0.941 (95%

Table 2 Basic information of different mediastinal lesions and CTCs levels in the whole cohort

Characteristics	Training cohort <i>n</i> = 434	Validation cohort <i>n</i> = 288	Test cohort <i>n</i> = 352	CTCs level (pg/ml)
Benign				
Cyst	162	54	90	10.38 ± 4.44
Granuloma	3	6	14	13.24 ± 5.57
Mature teratoma	8	7	3	9.85 ± 2.08
Schwannoma	32	17	34	11.97 ± 5.29
Fibroma	11	4	10	14.77 ± 5.52
Hemangioma	4	2	5	11.22 ± 2.83
Lipoma	10	3	6	12.73 ± 4.39
Malignant				
Type A thymoma	9	15	21	11.13 ± 4.32
Type AB thymoma	42	31	39	13.78 ± 6.51
Type B1 thymoma	15	29	26	14.84 ± 3.04
Type B2 thymoma	29	21	15	12.43 ± 6.23
Type B3 thymoma	13	10	13	14.04 ± 2.78
Mixed thymoma	13	19	17	13.50 ± 7.13
TSCC	28	13	18	15.89 ± 7.75
Immature teratoma	7	9	4	14.38 ± 5.42
MALT	4	5	3	21.66 ± 4.09
Carcinoid	2	4	5	9.86 ± 0.67
Lymphoma	18	22	14	16.25 ± 6.36
Metastatic tumor	5	3	4	11.38 ± 5.81
Others	19	14	11	16.33 ± 4.62

TSCC Thymic squamous cell carcinoma, MALT Mucosa-associated lymphoid tissue, Others Not specified lesions, CTCs Circulating tumor cells

confidence interval [CI] 0.901–0.982) and 0.710 (95% CI 0.664–0.756) respectively. The sensitivities were 0.809 (0.785–0.833) and 0.702 (0.676–0.728), and the accuracies were 0.927 (0.910–0.944) and 0.779 (0.743–0.815). In addition, the positive predictive value of correctly diagnosed benign mediastinal lesions is 0.885 (0.837–0.932) and 0.692 (0.671–0.713), suggesting the diagnostic accuracy of DMFN combined with CT images and clinical close-up manifestations was better than monomodal CNN (Table 3).

After conducting the benign lesions, we also compared the correct diagnostic performance of the malignant mediastinal lesions. A total of 190 malignant lesions were identified by the DMFN combined with CT images and CTCs and monomodal CNN in the test cohort. The accuracy of them were 0.905 (95%CI 0.861–0.949) and 0.793 (95%CI 0.744–0.842) respectively. Collectively, these data strongly indicated that the DMFN model has higher accurate diagnostic performances of malignant mediastinal lesions (Additional file 1: Table 1). Similar results could also be found in the test cohort (Fig. 3).

Multiclassification performance of CNN

The diagnostic parameters of multiclassification diagnosis in the training cohort were described in Fig. 4. The DMFN achieved the highest AUC (0.884, 95%CI 0.836–0.932) among all predictive methods, significantly outperforming the monomodal CNN (AUC = 0.722, 95%CI 0.681–0.763, $P = 0.030$) (Additional file 2: Fig. S1). For example, 20% of the benign thymoma were diagnosed as malignant mediastinal lesions by monomodal CNN mistakenly, but were correctly diagnosed as benign mediastinal lesions by DMFN subsequently. Furthermore, the DMFN demonstrated an excellent AUC of 0.935 in the validation cohort about the specific mediastinal lesions, obviously surpassing that of the monomodal CNN with the AUC = 0.782. Similarly, the DMFN significantly reduced the possibility of monomodal CNN mistakenly diagnosing certain malignant diseases as benign diseases in both the training cohort and validation cohort.

Diagnostic performance of thoracic surgeons

For the binary classifications, the diagnostic performances of chief physicians were significantly

Table 3 Diagnosis performance of CNN models and different thoracic surgeons in the validation cohort

Ratings	AUC (95%CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	PPV (95% CI)	NPV (95% CI)
Benign classification						
CNN model						
Monomodal CNN model	0.710(0.664–0.756)	0.702(0.676–0.728)	0.738(0.716–0.762)	0.779(0.743–0.815)	0.692(0.671–0.713)	0.723(0.695–0.751)
DMFN model	0.941(0.901–0.982)	0.809(0.785–0.833)	0.929(0.894–0.964)	0.927(0.910–0.944)	0.885(0.837–0.932)	0.939(0.884–0.993)
Thoracic surgeons						
Resident physicians	0.52(0.475–0.566)	0.513(0.502–0.524)	0.558(0.515–0.606)	0.69(0.666–0.723)	0.573(0.527–0.619)	0.538(0.818–0.561)
Attending physicians	0.688(0.649–0.714)	0.607(0.546–0.677)	0.72(0.682–0.763)	0.737(0.696–0.784)	0.693(0.635–0.754)	0.704(0.672–0.734)
Chief physicians	0.802(0.742–0.863)	0.761(0.720–0.803)	0.813(0.760–0.866)	0.830(0.805–0.856)	0.822(0.784–0.861)	0.811(0.752–0.872)
Management decision						
Resident physicians	0.636(0.568–0.711)	0.627(0.578–0.664)	0.684(0.657–0.707)	0.711(0.685–0.734)	0.643(0.626–0.630)	0.661(0.624–0.707)
Attending physicians	0.739(0.709–0.772)	0.705(0.685–0.733)	0.768(0.631–0.804)	0.839(0.801–0.870)	0.723(0.694–0.758)	0.718(0.689–0.742)
Chief physicians	0.921(0.882–0.946)	0.916(0.885–0.946)	0.928(0.895–0.953)	0.907(0.883–0.935)	0.919(0.881–0.942)	0.882(0.854–0.908)

AUC Area under the ROC curve, 95%CI 95% confidence interval, CNN Convolutional neural network, DMFN Deep multimodal fusion network, PPV Positive predictive value, NPV Negative predictive value

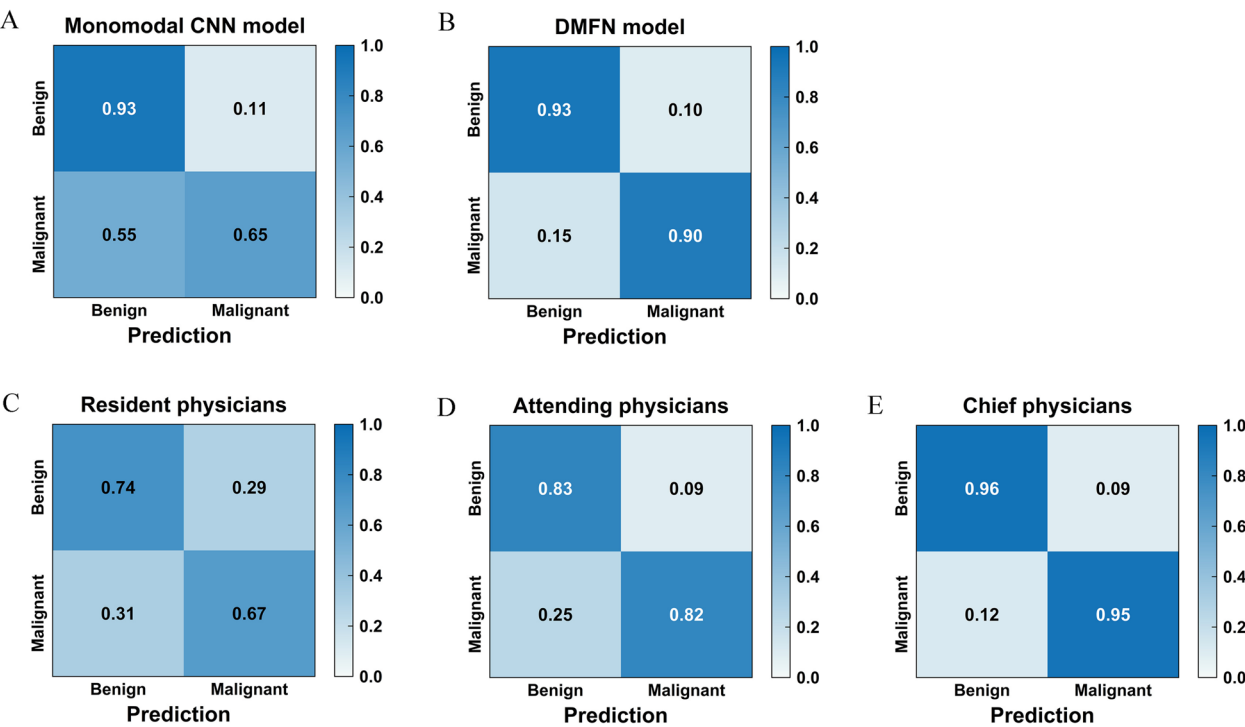


Fig. 3 Typical diagnostic performance matrices of binary classification in the test cohort. CNN: convolutional neural network; DMFN: deep multimodal fusion network

better than resident physicians and attending physicians in the test cohort, with the AUC of 0.802 (95%CI 0.746–0.858), 0.688 (95%CI 0.649–0.714), and 0.52 (95%CI 0.475–0.566). When combined with the CTCs,

the diagnostic performances of resident physicians significantly improved compared to not combining, with an AUC of 0.711 (95%CI 0.683–0.747) and 0.636 (95%CI 0.568–0.711). Additionally, the attending physicians

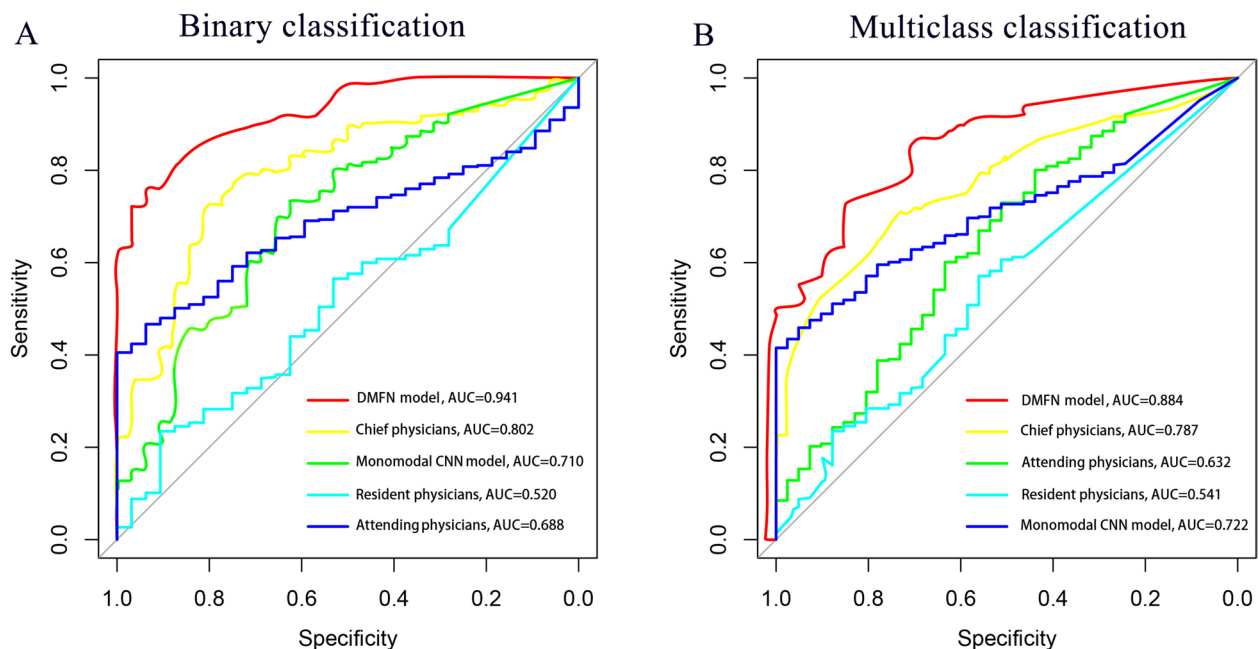


Fig. 4 ROC curves of CNN models and different thoracic surgeons in the binary and multiclass classifications. **A** ROC curves of CNN models and different thoracic surgeons in the binary classifications. **B** ROC curves of CNN models and different thoracic surgeons in the multiclass classifications. ROC: receiver operating characteristic; AUC: area under the ROC curve; CNN: convolutional neural network; DMFN: deep multimodal fusion network

significantly outperformed the diagnostic performance with an addition of CTCs, achieving an AUC of 0.805 (95%CI 0.773–0.840).

For the benign and malignant lesions, the diagnostic accuracy of resident physicians was lowest compared to resident physicians and attending physicians (Table 3). However, resident physicians showed a superior accuracy with the addition of CTCs, with the accuracy of 0.719 (95%CI 0.624–0.814) and 0.753 (95%CI 0.728–0.778).

Management decisions

For planning the optimal treatment for patients, all the thoracic surgeons were required to propose management strategies. As expected, the chief physicians made the correct management decisions with the highest accuracy of 0.907 (95%CI 0.883–0.935), followed by attending physicians 0.839 (95%CI 0.801–0.870) and resident physicians 0.711 (95%CI 0.685–0.734). The differences were statistically significant ($P=0.0425$, 0.0203) (Additional file 3: Table 2).

Diagnostic performance of CNN models vs. thoracic surgeons

For the classification of benign and malignant lesions in the test cohort, the DMFN model achieved a higher AUC than resident physicians, with an AUC of 0.932 (95%CI 0.871–0.993), higher than chief physicians (AUC=0.767,

95%CI 0.695–0.839). The DMFN achieved better predictive performances than the chief physicians, attending physicians and resident physicians ($P=0.054$, 0.020, 0.016) (Additional file 4: Fig. S2).

When predicting the malignant mediastinal lesions, the monomodal CNN model and attending physicians performed similar predictive abilities (AUC=0.665, 95%CI 0.619–0.712 vs AUC=0.723, 95%CI 0.649–0.797), both lower than the DMFN model (AUC=0.843, 95%CI 0.779–0.844). The predictive abilities were significantly different among the DMFN model and monomodal CNN model and physicians (Fig. 5).

Overall, for the classifications of the 14 types of mediastinal lesions, the DMFN model with an addition of CTCs data showed better performances than the chief physicians in the test cohort, and the difference reached statistical significance ($P=0.040$). The representative data are shown in Additional file 3: Table 2.

Discussion

In this study, we developed a DMFN to predict a broad spectrum of mediastinal lesions based on the CT images, clinical information, and CTCs levels, and we also compared its diagnostic performances with thoracic surgeons, especially for the overall classifications. The results demonstrated that the DMFN model significantly outperformed monomodal CNNs in multi-classification

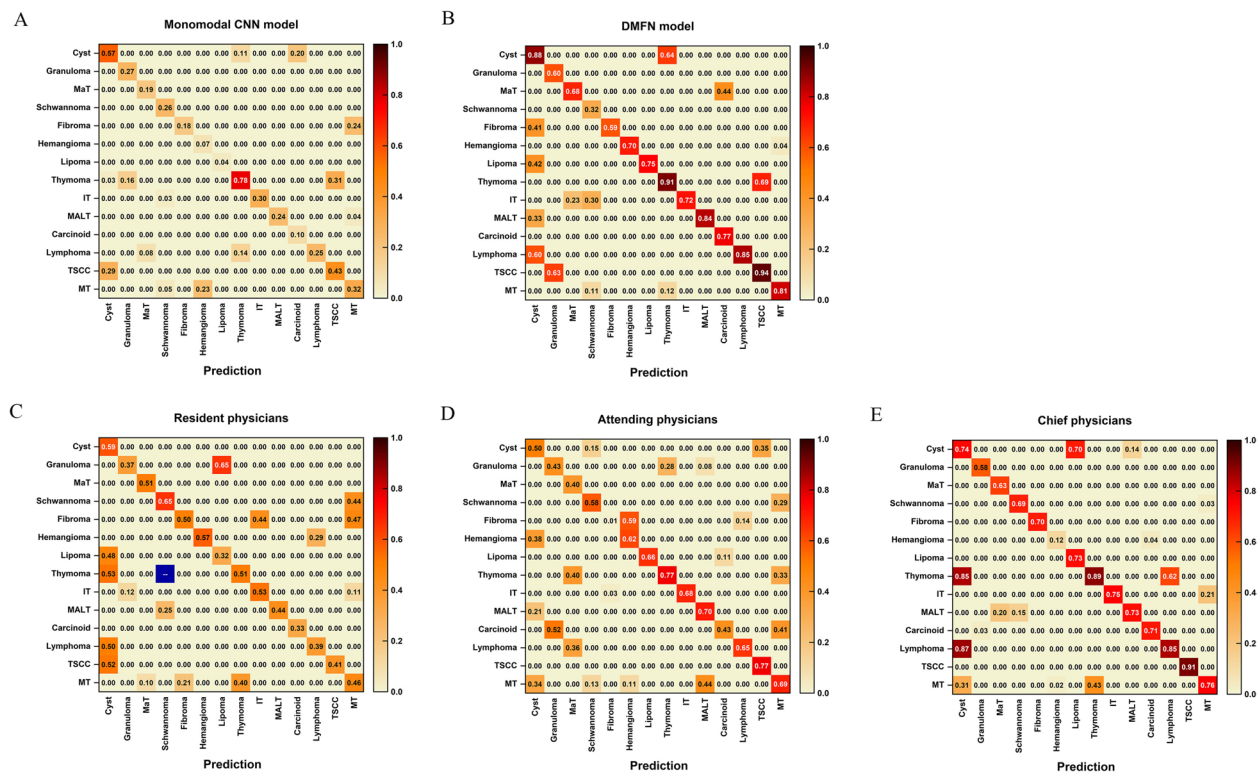


Fig. 5 Typical diagnostic performance matrices of multiclass classification in the test cohort. MaT: mature teratoma; IT: immature teratoma; TSCC: Thymic squamous cell carcinoma; MT: metastatic tumor; MALT: mucosa-associated lymphoid tissue

tasks, particularly for lesions with overlapping radiological appearances, such as benign and malignant thymomas. While CT is the preferred examination method, the CTCs could reveal the morphological or genetic information of primary tumor cells, supplementing key characteristics. Therefore, for the classification of mediastinal lesions, we suggest that DMFN combined with CTCs for optimal solutions.

Several studies have explored stratification and classification methods for mediastinal lesions using deep learning models, integrating clinical, radiomic, and deep features, and have demonstrated substantial potential for mediastinal tumor classification [20–24]. For example, Lin et al. developed a deep learning model combined with images to differentiate malignant and benign mediastinal lesions, achieving a diagnostic accuracy of 82% and an area under the curve of 0.8812 [23]. Similarly, Liu et al. demonstrated the utility of a 3D DenseNet model for detecting myasthenia gravis in thymoma patients using CT images, with an accuracy of 0.790 [24]. However, using a limited set of clinical factors may result in overestimating or underestimating model diagnostic performances and lead to a lack of generalizations to other diseases. Meanwhile, though effective, their model is slightly inferior to our model, indicating the added

benefit of incorporating CTCs data alongside radiological features. These studies collectively highlight the superior diagnostic performance achievable with multimodal fusion models, where combining imaging and molecular data not only enhances accuracy but also outperforms single-modality models.

The CNN model requires a large number of participants to effectively reflect the true underlying features and enhance the overall generalizability of the model [25]. Existing research has mainly been limited to insufficient data, leading to suboptimal performance or overfitting [26, 27]. By utilizing the backbone pretrained in the large data training cohort, we achieved improvements in both diagnostic abilities and accuracy compared to physicians in the real world. This is the first model to integrate CTCs with a large number of patients, and it also has the potential to generalize CNN to other types of thoracic diseases, assuming lesions share similar CTCs levels.

For binary classification tasks, the DMFN has superior performances than attending surgeons, but lower than the chief surgeons. This is maybe binary classifications were more common in clinical routine, and the CT images could provide more intuitive judgments. For the multi-classifications, the DMFN had advantages in the diagnosis of mediastinal lesions, and

demonstrated comparable or even superior performances compared to the chief surgeons. One reason is that some uncommon mediastinal lesions are resected with a rare incidence rate. Another reason may be that the DMFN with CTCs could easily identify mediastinal lesions through changes in CTC levels.

Since CTCs are gradually becoming popular in patients with mediastinal lesions [28, 29], our results suggested that DMFN with CTCs could be a suitable tool to improve diagnostic accuracy. According to the Ottaviano M's study, CTCs level is tightly associated with high mediastinal lesions disease burden and advanced stage [28]. In line with them, our study also supported that CTCs are higher in patients with malignant lesions than those with benign lesions. Given its low sensitivity, any steps to improve diagnosis should involve combining clinical data, such as CT images. While the DMFN demonstrates superior performance in complex multi-classification tasks, it may not offer significant advantages in simpler cases, such as mediastinal cysts, where physicians relying on CT images alone can achieve comparable results.

For clinical applicability, our DMFN model could be integrated into clinical practice as a secondary diagnostic tool to assist thoracic surgeons in pre-operative assessments. It can help classify mediastinal tumors based on CT images and CTC data, offering a non-invasive and automated diagnostic approach. However, the adoption of this model in clinical settings will require overcoming challenges such as ensuring data availability (e.g., high-quality CT scans and CTCs data), addressing the initial costs associated with model integration, and obtaining regulatory approvals from agencies like the FDA. Despite these challenges, the long-term benefits, including improved diagnostic accuracy, faster patient triage, and reduced need for invasive procedures, make this model a promising candidate for real-world clinical use. Once implemented, this system could significantly reduce diagnostic costs and improve patient outcomes, especially for those undergoing complex cancer treatments.

Some limitations should be considered. First, there are some inherent flaws in retrospective research. For example, we cannot control the external factors that affect CTCs. Second, the ability of DMFN model to generalize across large, diverse datasets further highlights its robustness, particularly when applied to cases with varying patient demographics and disease progression. Third, despite passing the validation of the test dataset, single-center research may still have selection bias. Multi-center studies are necessary to confirm our predictive models.

Conclusions

In this study, we proposed the DMFN model combined with CTCs and CT images, as a useful method to predict the classification of mediastinal lesions. And this model exhibited satisfactory diagnostic performance compared to thoracic surgeons. Therefore, this novel model may facilitate the diagnosis and help surgeons determine surgery strategy.

Abbreviations

AUC	Area under the ROC curves
CNN	Convolutional neural network
CTCs	Circulating tumor cells
DMFN	Deep multimodal fusion network
NPV	Negative predictive value
PPV	Positive predictive value
ROC	Receiver operating characteristic

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-025-04104-z>.

Additional file 1: Table 1 Diagnostic performance of the DMFN and mono-modal CNN models in identifying malignant mediastinal lesions.

Additional file 2: Fig. S1 Box plots comparing the diagnostic accuracies of thoracic surgeon and CNN model in the multiclass classification.

Additional file 3: Table 2 Diagnostic performance metrics and significance of group comparisons in the test cohort.

Additional file 4: Fig. S2 Box plots comparing the diagnostic accuracies of thoracic surgeon and CNN model in the binary classification.

Acknowledgements

We thank all the clinical and research teams involved in data collection, annotation, and model implementation processes in this study.

Authors' contributions

FW and MWB contributed to conception, design, data acquisition of the work. BT, FGY and GXW contributed to interpretation and data analysis. LZ designed the study and interpreted the results. All authors read and approved the final manuscript.

Funding

This work was supported by the Science and Technology Development Fund of Shanghai Pudong New Area (Grant No. PKJ2021-Y09).

Data availability

All the data displayed in the present manuscript are provided in the manuscript and additional files, and available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the ethics committee of Shanghai Pulmonary Hospital, School of Medicine, Tongji University (approval number: FK24-286). Informed consent was waived because of the retrospective nature.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Radiotherapy, School of Medicine, Shanghai Fourth People's Hospital, Tongji University, No. 1279, Sanmen Road, Shanghai 200081, China. ²Department of Thoracic Surgery, School of Medicine, Shanghai Pulmonary Hospital, Tongji University, 507 Zhengmin Road, Shanghai 200433, China. ³Department of Thoracic Surgery, School of Medicine, Shanghai East Hospital, Tongji University, 1800 Yuntai Road, Shanghai 200123, China. ⁴Research Center for Translational Medicine, School of Medicine, Shanghai East Hospital, Tongji University, 150 Jimo Road, Shanghai 200120, China.

Received: 6 June 2024 Accepted: 26 April 2025

Published online: 07 May 2025

References

- Knetki-Wróblewska M, Kowalski DM, Olszyna-Serementa M, Krzakowski M, Szolkowska M. Thymic epithelial tumors: do we know all the prognostic factors? *Thorac Cancer*. 2021;12(3):339–48. <https://doi.org/10.1111/1759-7714.13750>.
- Perrino M, Cordua N, De Vincenzo F, Borea F, Aliprandi M, Cecchi LG, et al. Thymic epithelial tumor and immune system: the role of immunotherapy. *Cancers (Basel)*. 2023;15(23):5574. <https://doi.org/10.3390/cancers15235574>.
- Ettinger DS, Riely GJ, Akerley W, Borghaei H, Chang AC, Cheney RT, et al. Thymomas and thymic carcinomas: clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2013;11(5):562–76. <https://doi.org/10.6004/jnccn.2013.0072>.
- Burt BM, Yao X, Shrager J, Antonicelli A, Padda S, Reiss J, et al. Determinants of complete resection of thymoma by minimally invasive and open thymectomy: analysis of an international registry. *J Thorac Oncol*. 2017;12(1):129–36. <https://doi.org/10.1016/j.jtho.2016.08.131>.
- Gu Z, Fu J, Shen Y, Wei Y, Tan L, Zhang P, et al. Thymectomy versus tumor resection for early-stage thymic malignancies: a Chinese alliance for research in thymomas retrospective database analysis. *J Thorac Dis*. 2016;8(4):680–6. <https://doi.org/10.21037/jtd.2016.03.16>.
- Tuan PA, Vien MV, Dong HV, Sibell M, La Civita E, Liotti A, Longo M, et al. MRI for determining thymoma in patients with myasthenia gravis. *Cancer Control*. 2019;26(1):1073274819865281. <https://doi.org/10.1177/1073274819865281>.
- Liu W, Wang W, Guo R, Zhang H, Guo M. Deep learning for risk stratification of thymoma pathological subtypes based on preoperative CT images. *BMC Cancer*. 2024;24(1):651. <https://doi.org/10.1186/s12885-024-12394-4>.
- Lagarde A, Le Collen L, Boulagnon C, Brixi H, Durlach A, Mougél G, et al. Early detection of relapse by ctDNA sequencing in a patient with metastatic thymic tumor and MEN1 mosaicism. *J Clin Endocrinol Metab*. 2022;107(10):e4154–8. <https://doi.org/10.1210/clinem/dgac454>.
- Ottaviano M, Giuliano M, Tortora M, La Civita E, Liotti A, Longo M, et al. A new horizon of liquid biopsy in thymic epithelial tumors: the potential utility of circulating cell-free DNA. *Front Oncol*. 2021;10:602153. <https://doi.org/10.3389/fonc.2020.602153>.
- Carlsson A, Nair VS, Luttgen MS, Keu KV, Horng G, Vasanawala M, et al. Circulating tumor microemboli diagnostics for patients with non-small-cell lung cancer. *J Thorac Oncol*. 2014;9(8):1111–9. <https://doi.org/10.1097/JTO.0000000000000235>.
- Kassam Z, Burgers K, Walsh JC, Lee TY, Leong HS, Fisher B. A prospective feasibility study evaluating the role of multimodality imaging and liquid biopsy for response assessment in locally advanced rectal carcinoma. *Abdom Radiol (NY)*. 2019;44(11):3641–51. <https://doi.org/10.1007/s00261-019-02135-8>.
- Nguyen TNA, Huang PS, Chu PY, Hsieh CH, Wu MH. Recent progress in enhanced cancer diagnosis, prognosis, and monitoring using a combined analysis of the number of circulating tumor cells (CTCs) and other clinical parameters. *Cancers (Basel)*. 2023;15(22):5372. <https://doi.org/10.3390/cancers15225372>.
- Ilie M, Hofman V, Long-Mira E, Selva E, Vignaud JM, Padovani B, et al. "Sentinel" circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease. *PLoS ONE*. 2014;9(10):e111597. <https://doi.org/10.1371/journal.pone.0111597>.
- Liu Z, Zhu Y, Yuan Y, Yang L, Wang K, Wang M, et al. 3D DenseNet deep learning based preoperative computed tomography for detecting myasthenia gravis in patients with thymoma. *Front Oncol*. 2021;11:631964. <https://doi.org/10.3389/fonc.2021.631964>.
- Yang L, Cai W, Yang X, Zhu H, Liu Z, Wu X, et al. Development of a deep learning model for classifying thymoma as Masaoka-Koga stage I or II via preoperative CT images. *Ann Transl Med*. 2020;8(6):287. <https://doi.org/10.21037/atm.2020.02.183>.
- World Health Organization. International statistical classification of diseases and related health problems (ICD). 2022. <https://www.who.int/standards/classifications/classification-of-diseases>. Accessed 1 Jan 2022.
- Sargül M, Ozyildirim BM, Avci M. Differential convolutional neural network. *Neural Netw*. 2019;116:279–87. <https://doi.org/10.1016/j.neunet.2019.04.025>.
- Wang P, Qiao J, Liu N. An improved convolutional neural network-based scene image recognition method. *Comput Intell Neurosci*. 2022;2022:3464984. <https://doi.org/10.1155/2022/3464984>.
- Xu J, Pan Y, Pan X, Hoi S, Yi Z, Xu Z. RegNet: self-regulated network for image classification. *IEEE Trans Neural Netw Learn Syst*. 2023;34(11):9562–7. <https://doi.org/10.1109/TNNLS.2022.3158966>.
- Yang Y, Cheng J, Peng Z, Yi L, Lin Z, He A, et al. Development and validation of contrast-enhanced ct-based deep transfer learning and combined clinical-radiomics model to discriminate thymomas and thymic cysts: a multicenter study. *Acad Radiol*. 2024;31(4):1615–28. <https://doi.org/10.1016/j.acra.2023.10.018>.
- Nakajo M, Takeda A, Katsuki A, Jinguiji M, Ohmura K, Tani A, et al. The efficacy of 18F-FDG-PET-based radiomic and deep-learning features using a machine-learning approach to predict the pathological risk subtypes of thymic epithelial tumors. *Br J Radiol*. 2022;95(1134):20211050. <https://doi.org/10.1259/bjr.20211050>.
- Han S, Oh JS, Kim YI, Seo SY, Lee GD, Park MJ, et al. Fully automatic quantitative measurement of 18F-FDG PET/CT in thymic epithelial tumors using a convolutional neural network. *Clin Nucl Med*. 2022;47(7):590–8. <https://doi.org/10.1097/RLU.00000000000004146>.
- Lin CK, Wu SH, Chua YW, Fan HJ, Cheng YC. TransEBUS: The interpretation of endobronchial ultrasound image using hybrid transformer for differentiating malignant and benign mediastinal lesions. *J Formos Med Assoc*. 2024;S0929–6646(24)00216-X. <https://doi.org/10.1016/j.jfma.2024.04.016>.
- Zhou Z, Guo Y, Tang R, Liang H, He J, Xu F. Privacy enhancing and generalizable deep learning with synthetic data for mediastinal neoplasm diagnosis. *NPJ Digit Med*. 2024;7(1):293. <https://doi.org/10.1038/s41746-024-01290-7>.
- Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. *J Med Imaging Radiat Oncol*. 2021;65(5):545–63. <https://doi.org/10.1111/1754-9485.13261>.
- Kayi Cangir A, Orhan K, Kahya Y, Özakıncı H, Kazak BB, Konuk Balci BM, et al. CT imaging-based machine learning model: a potential modality for predicting low-risk and high-risk groups of thymoma: "Impact of surgical modality choice." *World J Surg Oncol*. 2021;19(1):147. <https://doi.org/10.1186/s12957-021-02259-6>.
- Xiao G, Rong WC, Hu YC, Shi ZQ, Yang Y, Ren JL, et al. MRI radiomics analysis for predicting the pathologic classification and TNM staging of thymic epithelial tumors: a pilot study. *AJR Am J Roentgenol*. 2020;214(2):328–40. <https://doi.org/10.2214/AJR.19.21696>.
- Wu YH, Chao HS, Chiang CL, Luo YH, Chiu CH, Yen SH, et al. Personalized cancer avatars for patients with thymic malignancies: A pilot study with circulating tumor cell-derived organoids. *Thorac Cancer*. 2023;14(25):2591–600. <https://doi.org/10.1111/1759-7714.15039>.
- Ying J, Huang Y, Ye X, Zhang Y, Yao Q, Wang J, et al. Comprehensive study of clinicopathological and immune cell infiltration and lactate dehydrogenase expression in patients with thymic epithelial tumours. *Int Immunopharmacol*. 2024;126:111205. <https://doi.org/10.1016/j.intimp.2023.111205>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.