frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 21 July 2025
DOI 10.3389/fmed.2025.1636008

OPEN ACCESS

EDITED BY
Liang Zhao,
Dalian University of Technology, China

REVIEWED BY
Shixiong Zhang,
Xidian University, China
Jinyang Huang,
Hefei University of Technology, China

*CORRESPONDENCE
Rongdao Sun
✉ sunrongdao1837@163.com

RECEIVED 27 May 2025
ACCEPTED 27 June 2025
PUBLISHED 21 July 2025

CITATION
Wang S, Wang S and Sun R (2025) DCAI: a dual cross-attention integration framework for benign-malignant classification of pulmonary nodules. *Front. Med.* 12:1636008. doi: 10.3389/fmed.2025.1636008

DCAI: a dual cross-attention integration framework for benign-malignant classification of pulmonary nodules

Shuling Wang[1], Suixue Wang[2] and Rongdao Sun[1]*

[1]Department of Neurology, Haikou Affiliated Hospital of Central South University Xiangya School of Medicine, Haikou, China, [2]School of Computer Science and Technology, Hainan University, Haikou, China

Lung cancer remains a leading cause of cancer-related mortality worldwide, and accurate early identification of malignant pulmonary nodules is critical for improving patient outcomes. Although artificial intelligence (AI) technology has shown promise in pulmonary nodule benign-malignant classification, existing methods struggle with modality heterogeneity and limited exploitation of complementary information across modalities. To address the above issues, we propose a novel multimodal framework, the Dual Cross-Attention Integration framework (DCAI), for benign-malignant classification of pulmonary nodules. Specifically, we first convert 3D nodules into multiple 2D images and obtain nodule features annotated by clinical experts. These features are encoded using Transformer models, and then a dual cross-attention module is proposed to dynamically align and interact with the complementary information between the different modalities. The fused representations from both modalities are then concatenated for benign-malignant prediction. We evaluate our proposed method on the LIDC-IDRI dataset, and experimental results demonstrate that DCAI outperforms several existing multimodal methods, highlighting the effectiveness of our approach in improving the accuracy of pulmonary nodule benign-malignant classification.

KEYWORDS

pulmonary nodule, benign-malignant classification, artificial intelligence, multimodal, cross-attention, transformer

1 Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide, with its five-year survival rate strongly dependent on early diagnosis. According to the World Health Organization (WHO), lung cancer accounted for ∼2.2 million new cases and 1.8 million deaths worldwide in 2020, accounting for 18% of all cancer deaths (1). Pulmonary nodules are critical radiographic indicators of early-stage lung cancer, and accurate differentiation between benign and malignant nodules is essential for clinical decision-making and patient prognosis. Low-dose computed tomography (LDCT), the primary screening modality, has significantly improved detection rates, but it also results in a high false-positive rate of up to 95% (2), leading to unnecessary invasive procedures (e.g., biopsy) and increased strain on healthcare resources. Conventional diagnosis relies heavily on radiologists' subjective assessment of nodule morphological features (e.g., spiculation, lobulation), which is limited by inter-observer variability (Cohen's k = 0.45–0.67) (3) and reduced sensitivity for small nodules (<6 mm) (4). Although adjunctive techniques such

as PET-CT and liquid biopsy can provide additional diagnostic information, their clinical utility is constrained by high costs, radiation exposure, and risks associated with invasive procedures (5).

Hence, the development of efficient and noninvasive methods for predicting the malignancy of pulmonary nodules has become a major research focus. Radiomics, which quantitatively analyzes features such as texture, shape, and heterogeneity of nodules, has significantly enhanced diagnostic objectivity. For instance, the American College of Radiology (ACR) Lung-RADS classification system standardizes the evaluation process, increasing early lung cancer diagnostic specificity to 85%. However, its positive predictive value for category 3–4 nodules remains below 35% (5). Recent advances in artificial intelligence (AI) offer new opportunities for automated benign-malignant classification of pulmonary nodules with multimodal data, such as CT images and structured features annotated by the clinician. Deep learning models (6–8) can extract high-level features from multimodal data beyond human visual perception, with studies showing that AI-assisted systems can achieve classification accuracies exceeding 90% for small nodules, thereby reducing overdiagnosis and improving resource allocation (9).

Despite progress in existing methods, multimodal fusion still faces three major challenges: (1) The modality heterogeneity between CT images (high-dimensional spatial data) and clinical features (low-dimensional structured data) complicates feature alignment. (2) Traditional fusion strategies, such as concatenation or weighted averaging, struggle to dynamically adjust for redundant or conflicting information. To address the above issues, we propose a dual cross-attention integration framework, named DCAI, to classify the benign-malignant of pulmonary nodules. Specifically, 3D nodule CT scans are first converted into multiple 2D slices, and expert-annotated clinical features are interpolated. Both modalities are encoded using Transformer models, and a dual cross-attention module is proposed to capture complementary information between them. The resulting fused representations are concatenated for benign-malignant prediction. Experimental results demonstrate the superior performance of DCAI over existing multimodal methods, highlighting its effectiveness in improving pulmonary nodule benign-malignant classification accuracy.

In summary, our contributions are as follows:

- We incorporate transformer models to deeply encode clinical structured data and CT images separately, enabling the learning of high-level features from both modalities.
- We propose a dual cross-attention module that dynamically regulate the flow of complementary information between imaging and clinical structured features, effectively addressing modality heterogeneity and alignment challenges.

2 Related work

In recent years, research on pulmonary nodule malignancy classification has evolved along two major technical pathways: unimodal and multimodal approaches. In unimodal methods, most studies focus on developing deep learning models based solely on CT imaging. For instance, Li et al. (10) a deep convolutional neural network for nodule classification that leverages automatic feature learning and exhibits strong generalization performance. Donga et al. (11) propose a machine learning-based framework using a modified gradient boosting method for classifying pulmonary nodules, which integrates CT image preprocessing, random walker segmentation, and feature extraction. Wang et al. (12) proposed a decision tree model based on nodule size and density thresholds for preliminary risk stratification in the Chinese population, achieving an AUC of 0.899 in the first phase of their C-Lung-RADS system. However, unimodal approaches face challenges in handling the complexity and variability of pulmonary nodules, including high intra- and inter-patient variability in shape, size, and texture. Small or early-stage nodules may lack distinct features in CT scans, leading to reduced sensitivity and accuracy in malignancy prediction.

Multimodal approaches further enhance performance by integrating imaging data with clinical information. A notable example is the work by Yao et al. (13), who introduced a machine learning framework combining dynamic PET/CT metabolic and hemodynamic features, such as time-activity curve decomposition (TAC), to improve diagnostic specificity. Recent innovations also explore cross-modal fusion, such as the RFSC network developed by Wang et al., which aligns low-dose CT and MRI images through unsupervised registration, achieving 89.9% classification accuracy while reducing radiation exposure (14). Yuan et al. (15) propose a multi-modal fusion multi-branch classification network, which integrates structured radiological features and 3D CT patch data using an effective attention mechanism to classify pulmonary nodules as benign or malignant. These multimodal frameworks address the limitations of unimodal methods by leveraging complementary data sources, thereby reducing false positives and optimizing resource allocation in clinical practice. Sun et al. (16) propose the Nodule-CLIP model, which leverages comparative learning to explore the relationship between CT images and lung nodule attributes, enhancing the model's ability to distinguish between benign and malignant nodules. Tang et al. (17) construct two models (SUDFNN and SUDFX) that integrate 3D CNN-extracted image features with radiologist-annotated structured features using softmax and XGBoost classifiers, respectively. Liu et al. (18) propose a multimodal deep learning network integrating ResNet imaging features, Word2Vec semantic data, and self-attention mechanisms, achieving high accuracy in differentiating benign/malignant pulmonary nodules.

3 Method

3.1 Data preprocessing

The experimental data were derived from two publicly available datasets: the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) (3) and the Lung Nodule Analysis 2016 (LUNA16) (19). The LIDC-IDRI dataset comprises 1,018 thoracic CT scans with XML annotations generated through a two-phase review protocol involving four board-certified radiologists. Annotated nodule characteristics include malignancy, subtlety, internal structure, calcification, sphericity,

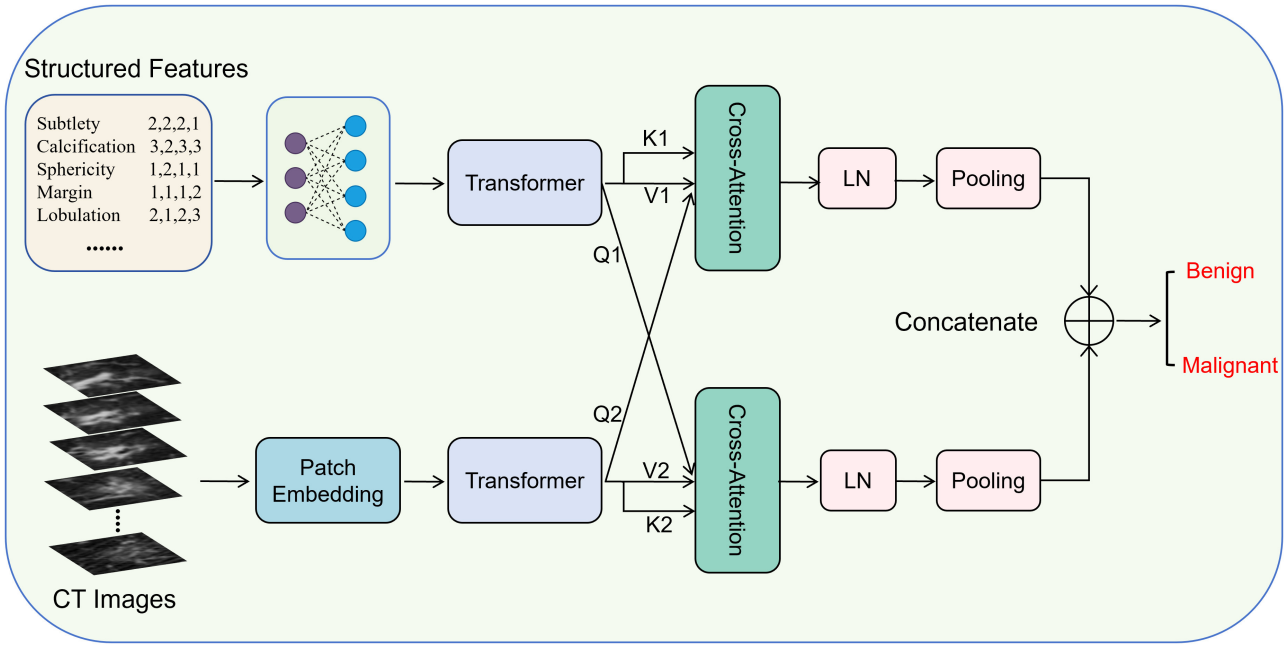

FIGURE 1
The overall framework of dual cross-attention integration (DCAI) .

margin, lobulation, speculation, texture, and diameter (the latter provided by LUNA16).

Following previous work (17), we consider the criteria required nodules $\geq$ 3 mm in diameter with consensus annotations from at least three radiologists. The malignancy score (ranging 1–5, higher values indicating increased malignancy likelihood) served as the classification target. For each nodule, multi-reader malignancy ratings are averaged and rounded to the nearest integer. Nodules with final scores of 1–2 are categorized as benign (n = 354), while those scoring 4–5 are classified as malignant (n = 330). Intermediate scores (3) are excluded to ensure diagnostic certainty. Missing annotations in other structured characteristics are addressed through radiologist-wise imputation. When only three radiologists provided annotations, the fourth radiologist's entry is populated using the rounded mean of existing annotations. CT image preprocessing involved isotropic resampling to $1 \times 1 \times 1 mm^3/voxel$ resolution, followed by extraction of $32 \times 32 \times 32$ voxel cubes centered on nodule coordinates. This yielded standardized 3D nodule volumes as unstructured imaging inputs. As a result, we obtain 684 samples, each comprising nine structured radiographic attributes, one 3D nodule volume, and a binary benign/malignant label.

3.2 Dual cross-attention integration framework

The overall framework of DCAI is illustrated in Figure 1. Given a nodule case x_i from the dataset $x = \{x_1, x_2, \ldots, x_N\}$, containing a 3D nodule volume (CT image) $c_i \in \mathbb{R}^{32\times32\times32}$ and nine structured features $s_i \in \mathbb{R}^{9\times4}$, we encode them separately into representations using two types of encoders, which are then aligned with two dual cross-attention modules and fused with concatenation.

Specifically, we employ a linear network built on two fully connected layers to map the original structured features to a new semantic space, and then learn the relationships within the structured features using a Transformer module. They are written as:

$$S_{emb} = \mathrm{ReLU}\left(\mathrm{W}_2(\mathrm{W}_1\mathrm{S} + \mathrm{b}_1) + \mathrm{b}_2\right) \quad (1)$$

where $\mathrm{W}_1 \in \mathbb{R}^{4\times32}$, $\mathrm{W}_2 \in \mathbb{R}^{32\times256}$, $\mathrm{b}_1 \in \mathbb{R}^{1\times32}$, and $\mathrm{b}_2 \in \mathbb{R}^{1\times256}$ are learnable weights. ReLU is the activation function.

$$S_{enc} = \text{Transformer-S}\,(S_{emb}) \quad (2)$$

where Transformer-S () is the Transformer encoder module, which consists of 6 Transformer blocks.

Meanwhile, 3D nodule volume (CT image) is first divided into 32 32 $\times$ 32 images, and then patch embedded with a lightweight CNN, followed by encoding with a Transformer (20, 21), which are written as:

$$C_{emb} = \text{LightweightCNN}\,(C) \quad (3)$$
$$C_{enc} = \text{Transformer-C}\,(C_{emb}) \quad (4)$$

where LightweightCNN denotes the lightweight CNN with a kernel size of 32×32 and 256 channels.

After that, the encoded representations of structured features and CT image are aligned and dynamically interact with two dual

TABLE 1 Performance comparison of DCAI and existing models on the LIDC-IDRI dataset.

Methods	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Nodule-CLIP (16)	0.934	0.925	0.939	0.930	0.932
Self-attention-based (18)	0.920	0.910	0.924	0.915	0.917
SUDFX32 (17)	0.942	0.926	0.955	0.930	0.940
DCAI (Ours)	**0.964**	**0.955**	**0.970**	**0.958**	**0.962**

TABLE 2 Ablation study of modality.

Modality	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Structured features	0.927	**0.967**	0.879	**0.972**	0.921
CT images	0.942	0.903	**0.985**	0.901	0.942
Multimodal	**0.964**	0.955	0.970	0.958	**0.962**

cross-attention modules (22, 23), which are written as:

$$S_{CA} = \text{CrossAttn}\left(W_{q,s}C_{enc}, W_{k,s}S_{enc}, W_{v,s}S_{enc}\right) = \text{Softmax}\left(\frac{W_{q,s}C_{enc}S_{enc}^{T}W_{k,s}^{T}}{\sqrt{d}}\right)W_{v,s}S_{enc} \quad (5)$$

$$C_{CA} = \text{CrossAttn}\left(W_{q,c}S_{enc}, W_{k,c}C_{enc}, W_{v,c}C_{enc}\right) = \text{Softmax}\left(\frac{W_{q,c}S_{enc}C_{enc}^{T}W_{k,c}^{T}}{\sqrt{d}}\right)W_{v,c}C_{enc} \quad (6)$$

where $W_{q,s}, W_{k,s}, W_{v,s}, W_{q,c}, W_{k,c}$, and $W_{v,c}$ are trainable weight matrices multiplied by the corresponding queries, keys, and values.

Then, the S_{CA} and C_{CA} are layer normalizated and average pooled, respectively. Next, the two unimodal representations are concatenated as the multimodal representations:

$$S_{final} = \text{AvgPool}\left(\text{LayerNorm}\left(S_{CA}\right)\right) \quad (7)$$

$$C_{final} = \text{AvgPool}\left(\text{LayerNorm}\left(C_{CA}\right)\right) \quad (8)$$

$$V = \left[S_{final} \oplus C_{final}\right] \quad (9)$$

Finally, the multimodal representations V are used to predict the probabilities of nodule samples. Specifically, the multimodal representations V are first fed into a linear network, followed by the application of a cross-entropy loss function to compute the final loss. The overall process is formulated as follows:

$$\mathcal{L} = \text{CE}\left(\text{Linear}\left(V\right), y^{true}\right) \quad (10)$$

where y^{true} is the ground-truth labeling and $\text{CE}(\cdot)$ denotes the cross-entropy loss function.

4 Experiments and results analysis

4.1 Experiment setup

The experiments are executed on a Linux-based system equipped with three NVIDIA A100 GPUs and the PyTorch framework. The DCAI model is trained for 30 epochs with a batch size of 100, using a learning rate of 0.0005 and a weight decay of 0.001 to regularize optimization. The Transformer architecture comprised six stacked encoder blocks, ensuring sufficient depth for feature abstraction.

We conduct all experiments using a rigorous five-fold cross-validation to ensure robust performance evaluation. The dataset is randomly partitioned into five distinct subsets, with each fold serving as the test set once, while the remaining four folds are utilized for model training. The final performance metrics are averaged across all five iterations to mitigate bias and enhance statistical reliability.

4.2 Evaluation metrics

We evaluate the performance of pulmonary nodule malignancy classification (malignant as positive, benign as negative) by various metrics, such as accuracy, precision, sensitivity, specificity, and F1-Score. They are calculated based on a confusion matrix:

$$\text{Confusion-Matrix} = \begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix} \quad (11)$$

where TN (true negative) and TP (true positive) denote correct predictions for benign and malignant cases, respectively; FP (False Positive) and FN (False Negative) represent misclassified benign and malignant cases.

The metrics of accuracy, precision, sensitivity, specificity, and F1-Score are written as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (12)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (13)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (15)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

TABLE 3 Ablation study of key components.

Methods	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Without transformer	0.927	0.900	0.955	0.901	0.926
Without cross-attention	0.898	0.882	0.909	0.887	0.896
DCAI (complete)	**0.964**	**0.955**	**0.970**	**0.958**	**0.962**

4.3 Comparison with existing multimodal methods

To validate the effectiveness of our method, we compare DCAI with three multimodal fusion approaches: Nodule-CLIP (16), Self-attention-based (18), and SUDFX32 (17). As shown in Table 1, DCAI achieves superior performance across all metrics, attaining an accuracy of 0.964, precision of 0.955, sensitivity of 0.970, specificity of 0.958, and F1-score of 0.962. Notably, DCAI outperforms SUDFX32, the previous best method, by 2.2% in accuracy and 2.2% in F1-score, demonstrating its robust capability in integrating multimodal features. The significant improvements in sensitivity (1.5% higher than SUDFX32) and specificity (2.8% higher) further highlight its balanced diagnostic reliability. These results validate the efficacy of our cross-modal alignment strategy, which effectively reduce inter-modal discrepancies while preserving discriminative features.

4.4 Ablation study

To investigate the performance of different modal input data and individual submodules, we conduct ablation experiments on Unimodal vs. Multimodal configurations and Key Components, respectively.

4.4.1 Unimodal and multimodal

Table 2 reveals distinct strengths of unimodal inputs: structured features excel in precision (0.967), indicating robust identification of benign cases, while CT images achieve superior sensitivity (0.985), effectively detecting malignant nodules. However, unimodal models exhibit limitations–structured features show lower sensitivity (0.879), and CT images underperform in specificity (0.901). Multimodal fusion balances these metrics, achieving optimal accuracy (0.964) and F1-score (0.962). This synergy highlights how combining clinical metadata with imaging data mitigates modality-specific biases, enhancing holistic diagnostic reliability for both benign and malignant cases.

4.4.2 Key components

Ablating key components (Table 3) demonstrates their critical roles. Removing the Transformer reduces specificity (0.901 vs. 0.958), suggesting its necessity for modeling global context to minimize false positives (benign misclassified as malignant). Disabling cross-attention causes significant drops in sensitivity (0.909 vs. 0.970) and precision (0.882 vs. 0.955), emphasizing its role in aligning multimodal features for accurate malignant detection. The complete DCAI model architecture achieves balanced performance, proving that both components are crucial for dynamic cross-modal interaction and discriminative feature preservation.

5 Conclusion

In this paper, we propose DCAI, a dual cross-attention integration framework for benign-malignant classification of pulmonary nodules. We design DCAI to address modality heterogeneity and feature alignment challenges in multimodal fusion. Leveraging Transformer-based encoders for clinical structured features and CT images, DCAI captures high-level semantic representations while dynamically aligning complementary information through the dual cross-attention module. Evaluated on the LIDC-IDRI dataset, DCAI achieves superior performance, outperforming existing methods significantly. Ablation studies confirm the necessity of both cross-attention and Transformer components. The experimental results indicate that our framework provides a robust, noninvasive solution to enhance early lung cancer diagnosis and reduce unnecessary interventions, demonstrating clinical potential for reliable malignancy characterization.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ShW: Conceptualization, Data curation, Validation, Writing – original draft, Investigation. SuW: Funding acquisition, Writing – original draft, Methodology, Software. RS: Validation, Writing – review & editing, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by Hainan Provincial Natural Science Foundation of China with No. 825CXTD608.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

2. Team NLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. (2011) 365:395–409. doi: 10.1056/NEJMoa1102873

3. Armato III SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys*. (2011) 38:915–31. doi: 10.1118/1.3528204

4. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP. et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer?: diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. (2013) 143:e93S–120S. doi: 10.1378/chest.12-2351

5. Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Ann Intern Med*. (2015) 162:485–91. doi: 10.7326/M14-2086

6. Zhao L, Huang P, Chen T, Fu C, Hu Q, Zhang Y. Multi-sentence complementarily generation for text-to-image synthesis. *IEEE Trans Multimed*. (2023) 26:8323–32. doi: 10.1109/TMM.2023.3297769

7. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x

8. Zhao L, Xie Q, Li Z, Wu S, Yang Y. dynamic graph guided progressive partial view-aligned clustering. *IEEE Trans Neural Netw Learn Syst*. (2024) 36:9370–82. doi: 10.1109/TNNLS.2024.3425457

9. Zhao L, Wang X, Liu Z, Wang Z, Chen Z. Learnable graph guided deep multi-view representation learning via information bottleneck. *IEEE Trans Circuits Syst Video Technol*. (2024) 35:3303–14. doi: 10.1109/TCSVT.2024.3509892

10. Li W, Cao P, Zhao D, Wang J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput Math Methods Med*. (2016) 2016:6215085. doi: 10.1155/2016/6215085

11. Donga HV, Karlapati JSAN, Desineedi HSS, Periasamy P, TR S. Effective framework for pulmonary nodule classification from CT images using the modified gradient boosting method. *Appl Sci*. (2022) 12:8264. doi: 10.3390/app12168264

12. Wang C, Shao J, He Y, Wu J, Liu X, Yang L, et al. Data-driven risk stratification and precision management of pulmonary nodules detected on chest computed tomography. *Nat Med*. (2024) 30:3184–95. doi: 10.1038/s41591-024-03211-3

13. Yao Z, Wang Y, Wu Y, Zhou J, Dang N, Wang M, et al. Leveraging machine learning with dynamic 18F-FDG PET/CT: integrating metabolic and flow features for lung cancer differential diagnosis. *Eur J Nucl Med Mol Imaging*. (2025) 1–13. doi: 10.1007/s00259-025-07231-0

14. Wang Y, Peng X, Yang S, Zhang Y, Dai R, Meng X, et al. RFSC: Multimodal medical image alignment fusion diagnostic classification network based on de discriminator image translation. *Biomed Signal Process Control*. (2025) 109:107905. doi: 10.1016/j.bspc.2025.107905

15. Yuan H, Wu Y, Dai M. Multi-modal feature fusion-based multi-branch classification network for pulmonary nodule malignancy suspiciousness diagnosis. *J Digit Imaging*. (2023) 36:617–26. doi: 10.1007/s10278-022-00747-z

16. Sun L, Zhang M, Lu Y, Zhu W, Yi Y, Yan F. Nodule-CLIP: lung nodule classification based on multi-modal contrastive learning. *Comput Biol Med*. (2024) 175:108505. doi: 10.1016/j.compbiomed.2024.108505

17. Tang N, Zhang R, Wei Z, Chen X, Li G, Song Q, et al. Improving the performance of lung nodule classification by fusing structured and unstructured data. *Inf Fusion*. (2022) 88:161–74. doi: 10.1016/j.inffus.2022.07.019

18. Liu G, Liu F, Mao X, Xie X, Sang J, Ma H, et al. Multimodal deep learning network for differentiating between benign and malignant pulmonary ground glass nodules. *Curr Med Imaging*. (2024) 20:e15734056301741. doi: 10.2174/0115734056301741240903072017

19. Setio AAA, Traverso A, De Bel T, Berens MS, Van Den Bogaard C, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal*. (2017) 42:1–13. doi: 10.1016/j.media.2017.06.015

20. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems*. Long Beach, CA: NeurIPS Foundation (2017). p. 30.

21. Wang S, Wang S, Liu Z, Zhang Q. A role distinguishing Bert model for medical dialogue system in sustainable smart city. *Sustain Energy Technol Assess*. (2023) 55:102896. doi: 10.1016/j.seta.2022.102896

22. Wang S, Hu X, Zhang Q. HC-MAE: hierarchical cross-attention masked autoencoder integrating histopathological images and multi-omics for cancer survival prediction. In: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Istanbul: IEEE (2023). p. 642–7. doi: 10.1109/BIBM58861.2023.10385635

23. Wang S, Zheng Z, Wang X, Zhang Q, Liu Z. A cloud-edge collaboration framework for cancer survival prediction to develop medical consumer electronic devices. *IEEE Trans Consum Electron*. (2024) 70:5251–8. doi: 10.1109/TCE.2024.3413732