European Journal of Public Health and Environmental Research

Vol. 1 No.1 2025

Article **Open Access**

Innovative Application of Multimodal Medical Imaging in Complex Lesion Diagnosis Based on Deep Fusion Networks

Jisoo Park [1,*], Minjae Kim [1], Eunji Lee [1], Hyunwoo Choi [2] and Seungmin Oh [3]

[1] Department of Computer Science, Yonsei University, Seoul, South Korea
[2] Department of Artificial Intelligence, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea
[3] Department of Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea
* Correspondence: Jisoo Park, Department of Computer Science, Yonsei University, Seoul, South Korea

Received: 20 March 2025
Revised: 29 March 2025
Accepted: 13 April 2025
Published: 17 April 2025

Abstract: Multimodal medical imaging, which combines spatial and functional information, plays an important role in improving the accuracy of complex disease diagnosis. This study aims to address the diagnostic challenges of complex lesions by designing a deep fusion network that integrates channel attention and multi-scale feature extraction. An end-to-end model was built and tested on two public multimodal datasets: glioma and lung tumors. The experimental results show that, compared with existing multimodal fusion methods, the proposed approach achieves better performance in classification accuracy, area under the receiver operating characteristic curve (ROC-AUC), and Dice coefficient for image segmentation. This method provides a new solution for clinical decision support based on multi-source imaging.

Keywords: multimodal fusion; deep learning; medical image analysis; diagnostic support system; tumor detection

1. Introduction

Medical imaging has become a vital part of modern clinical diagnostics after more than a century of continuous development [1]. Since Wilhelm Röntgen discovered X-rays in 1895 and launched a new era of diagnostic imaging, the field has undergone steady technological advancement. Magnetic resonance imaging (MRI), known for its ability to distinguish soft tissues, offers sub-millimeter spatial resolution and allows clear visualization of gray and white matter structures in the brain [2]. Computed tomography (CT) has demonstrated excellent performance in lung disease diagnosis. Its scanning time has decreased from several minutes to a few seconds, greatly improving diagnostic efficiency. The detection rate of small pulmonary nodules has exceeded 90%. Positron emission tomography (PET) provides metabolic-level information and can detect abnormal activity months or even years before anatomical changes are visible, which is valuable in early tumor diagnosis [3]. However, single-modality imaging is limited by its lack of comprehensive information. When diagnosing complex lesions, it often fails to provide accurate and complete results. For example, in glioma diagnosis, MRI alone may not reliably determine the malignancy of tumors [4]. When combined with CT and PET, the diagnostic accuracy can improve by approximately 20%. Multimodal medical imaging fusion has emerged to address this challenge. By combining information from different imaging modalities, it provides both spatial and functional data that complement each other. In tumor

diagnosis, MRI offers detailed soft tissue contrast, CT is effective for imaging bones and lungs and PET reveals abnormal metabolic activity [5]. Fusing these modalities allows clinicians to evaluate the lesion's location, size, shape and biological behavior from multiple perspectives, which supports more accurate diagnosis and treatment planning [6].

From a clinical standpoint, multimodal fusion improves early detection rates and reduces the risk of misdiagnosis or missed diagnosis, helping patients receive timely treatment [7]. In research, it provides essential tools for exploring disease mechanisms, monitoring disease progression, and evaluating therapeutic outcomes. Therefore, studying multimodal imaging fusion has both practical relevance and wide application prospects. Globally, research on deep fusion networks for multimodal medical imaging started earlier and has progressed rapidly [8]. Many academic groups and healthcare institutions have focused on developing efficient fusion algorithms and network structures [9]. For example, the team led by Dr. Yinsheng Li at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, has conducted long-term work in medical imaging and artificial intelligence. By improving deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), they achieved effective fusion of MRI and CT data [10]. In brain disease diagnosis experiments, their fusion model increased diagnostic accuracy from 70% to 85% compared with single-modality approaches. Some advanced studies have already been applied in clinical settings. Large-scale clinical trials have confirmed their value. In one study involving 1,000 cardiovascular patients, multimodal fusion diagnosis improved treatment response rates from 65% to 78% [11]. In China, research in this area is also developing quickly. More universities and research institutes are investing in this field. Progress has been made in fusion strategies, model optimization and clinical integration [12]. Chinese researchers have proposed several new methods, such as attention-based fusion, which improves the model's ability to identify key information. In one comparative study, an attention-based fusion model achieved an ROC-AUC of 0.92 for lung tumor diagnosis, higher than the 0.85 from the model without attention mechanisms. In terms of architecture, some teams have developed end-to-end deep fusion networks with independent intellectual property rights, which improved diagnostic accuracy and efficiency [13]. Despite this progress, several core challenges remain. These include handling data heterogeneity across modalities, improving algorithm robustness and generalization and enhancing model interpretability to make the outputs more understandable and usable for clinicians [14].

This study aims to develop an effective and accurate auxiliary diagnostic method for complex lesions using deep multimodal fusion networks. The main goals are: To improve the diagnostic accuracy of complex lesions and reduce misdiagnosis and missed diagnosis; To enhance the model's ability to integrate and utilize multimodal imaging data; To explore effective model architectures and training strategies to improve performance and stability. To achieve these goals, this research combines theoretical analysis with experimental validation. First, related theories in multimodal medical imaging, deep fusion networks, and fusion strategies were reviewed to provide a strong theoretical basis. Based on this, an end-to-end model integrating channel attention and multi-scale feature extraction was designed. The model was trained and optimized using publicly available multimodal datasets on gliomas and lung tumors. Comparative experiments were conducted with existing fusion methods. Finally, the effectiveness and advantages of the proposed method were validated through detailed analysis of experimental results.

2. Methodology

The diagnostic model developed in this study adopts an end-to-end architecture based on a multimodal deep fusion network [15]. In the input layer, data from different imaging modalities such as MRI, CT and PET undergo preprocessing steps including normalization, cropping and registration to ensure data consistency [16]. In the fusion layer, a feature-level fusion strategy is employed. For each modality, a dedicated sub-network

is used to extract features. These features are then concatenated and passed through a channel attention mechanism, which reweights them to emphasize important information. The feature extraction layer processes the fused feature vectors using convolution and pooling operations with different kernel sizes [17]. This enables the extraction of high-level features at multiple scales. In the classification and segmentation layers, the extracted features are used for diagnostic tasks. A fully connected layer combined with a Softmax classifier is applied for lesion classification. For segmentation, networks such as U-Net are used to identify lesion regions, supporting the final diagnostic decision.

The model is trained on two multimodal datasets: one glioma dataset containing 200 patient cases, sourced from the National New Generation Artificial Intelligence Open Innovation Platform for Medical Imaging and one lung tumor dataset with 150 cases, derived from the IRENE model developed by the Macau University of Science and Technology, West China Hospital of Sichuan University, and the University of Hong Kong. During preprocessing, pixel values were normalized to the range [0, 1]. To reduce overfitting, data augmentation techniques were applied, including random rotation (± 15°), horizontal and vertical flipping (each with a probability of 0.5) and scaling (scaling range: 0.8 to 1.2). The Adam optimizer was used to update model parameters. The initial learning rate was set to 0.001, and an exponential decay strategy was applied, reducing the rate to 90% of its value every 50 epochs. Cross-entropy loss was used for the classification task, while Dice loss was used for the segmentation task [18]. The dataset was divided into training, validation, and test sets in a ratio of 70%, 15%, and 15%, respectively. The training set was used to update parameters, the validation set was used to monitor model behavior and avoid overfitting, and the test set was used to evaluate final model performance [19]. Model performance was further improved by tuning hyperparameters such as learning rate, network depth, and convolution kernel size.

3. Experimental Design

3.1. Datasets and Experimental Settings

This study used multimodal datasets for glioma and lung tumors to conduct experiments. The glioma dataset was collected from the National New Generation Artificial Intelligence Open Innovation Platform for Medical Imaging. It contains data from 200 patients, including MRI (T1, T2, FLAIR and contrast-enhanced T1-weighted), CT and PET scans. All patients were pathologically diagnosed, and the dataset includes detailed annotations on tumor type, grade and lesion location. In general, low-grade gliomas appear as hyperintense regions with clear boundaries on T2-weighted MRI, while high-grade gliomas show enhancement and edema on contrast-enhanced T1-weighted MRI. The lung tumor dataset was derived from the IRENE model, developed by the Faculty of Medicine at Macau University of Science and Technology, West China Hospital of Sichuan University, and the University of Hong Kong. It includes CT and PET images from 150 patients. CT scans show the structure and shape of lung tissues and tumors, while PET scans indicate metabolic activity [20]. All samples were confirmed through diagnostic and pathological analysis, with clear labels for benign and malignant tumors. Malignant tumors show high uptake on PET images, typically with elevated standardized uptake values (SUV). To evaluate the model, this study compared it with several existing multimodal fusion algorithms: The weighted average method directly fuses pixel values from different modalities but has limited ability to handle complex data. Its diagnostic accuracy for glioma was 78.5%; The PCA-based fusion algorithm reduces dimensionality and combines features from different modalities. It captured some useful information, and its ROC-AUC in lung tumor diagnosis was 0.831; The MM-CNN model applies a specific network design for joint learning of multimodal features. In glioma segmentation, it achieved a Dice coefficient of 0.834; In this study, model performance was evaluated using classification accuracy, ROC-AUC and Dice coefficient to assess its effectiveness in diagnosing complex lesions.

3.2. Experimental Results

To clearly demonstrate the performance of the proposed model and the comparison methods on the glioma and lung tumor datasets, the results are organized and presented in Table 1 and Table 2.

Table 1. Results on the glioma dataset.

Evaluation Metric	Proposed Model	Weighted Averaging	PCA-Based Fusion	MM-CNN
Classification Accuracy	93.6%	78.5%	82.3%	88.7%
ROC-AUC	0.952	0.813	0.845	0.901
Dice coefficient (segmentation)	0.885 (overall), 0.89 (low-grade lesions), and 0.88 (high-grade lesions)	0.752	0.786	0.834

Table 2. Experimental results on the lung tumor dataset.

Evaluation Metric	Proposed Model	Weighted Average	PCA Fusion	MM-CNN
Classification Accuracy	92.1%	76.8%	80.5%	86.9%
ROC-AUC	0.945	0.802	0.831	0.896
Dice coefficient (segmentation)	0.90 (regular tumors), 0.86 (irregular tumors)	0.735	0.771	0.827

3.3. Results Analysis

As shown in Table 1, in the glioma dataset experiment, the proposed multimodal deep fusion network achieved a classification accuracy of 93.6%. According to the results of the independent-sample t-test, the improvement over the comparison algorithms was statistically significant ($p < 0.05$). Based on the confusion matrix, the correct classification rate for low-grade gliomas was 95%. Although some cases were confused due to similar imaging features, the model accurately identified them by utilizing complementary information from multiple modalities [21]. For high-grade gliomas, the correct classification rate reached 92%. A small number of misclassifications were due to their regular shape and less distinct metabolic patterns. In terms of ROC-AUC, the model achieved 0.952, significantly higher than other methods. For instance, when the false positive rate was 10%, the true positive rate reached 90%, indicating that the model effectively balanced false negatives and false positives. For lesion segmentation, the overall Dice coefficient was 0.885. The Dice score for low-grade lesions reached 0.89. Although high-grade lesions have more complex boundaries, the Dice score still reached 0.88, showing clear improvement compared with other methods.

3.4. Analysis on Lung Tumor Dataset

In the lung tumor dataset experiment, as presented in Table 2, the model achieved a classification accuracy of 92.1%. A chi-square test confirmed that this result was significantly higher than those of the weighted averaging method, PCA-based fusion and MM-CNN ($p < 0.05$). For benign tumors, the correct classification rate reached 94%. By combining information from both PET and CT images, the model reduced the number of false positives, including cases such as inflammatory pseudotumor [22,23]. For malignant tumors, the correct classification rate was 91%. Although some early-stage micro-lung can-

cers were misclassified due to their atypical metabolic features, the model still significantly improved diagnostic accuracy [24]. In terms of ROC-AUC, the model scored 0.945, showing a clear advantage. At a false positive rate of 5%, the true positive rate reached 85%, much higher than those of the comparison methods. For lesion segmentation, the Dice coefficient was 0.90 for tumors with regular shapes and 0.86 for irregular ones. Both results exceeded those of the baseline methods and provided useful guidance for treatment planning.

3.5. Results Analysis

The experimental results indicate that the proposed multimodal deep fusion network shows clear advantages in diagnosing complex lesions. In both glioma and lung tumor tasks, the model achieved higher classification accuracy, ROC-AUC and segmentation Dice coefficients compared with other methods [25]. These improvements are mainly due to the use of a channel attention mechanism, which enhances features from the most relevant channels. For instance, when fusing MRI, CT and PET data, the mechanism assigns greater weights to channels carrying key diagnostic information — such as those related to tumor metabolism in PET or tissue structure in MRI — thus improving the model's ability to identify lesions [26]. Meanwhile, the multi-scale feature extraction component effectively captures lesion characteristics at different spatial resolutions. In glioma diagnosis, low-grade tumors tend to be smaller and structurally simpler, while high-grade tumors are usually larger with more complex internal patterns, such as necrosis or cystic areas [27]. The multi-scale strategy extracts feature across different levels, allowing the model to better understand the lesion structure and improve classification performance. In lung tumor diagnosis, this approach also proves useful. For small early-stage nodules, fine-scale features help detect subtle morphological or metabolic changes. For larger tumors, coarse-scale features capture global shape and the spatial relationship between the tumor and surrounding tissues, supporting more accurate diagnosis. The end-to-end architecture further enables the model to learn internal associations among different imaging modalities. During training, it can automatically integrate complementary information — such as soft-tissue contrast from MRI, anatomical and density information from CT and metabolic data from PET — to form a more comprehensive understanding of the lesion, thereby enhancing diagnostic effectiveness [28]. However, the model's performance still declines slightly when dealing with highly complex lesions, especially those involving comorbidities. For example, in lung tumors complicated by infection or pleural effusion, the resulting changes in imaging appearance — such as blurred boundaries or altered metabolism — may interfere with the model's classification and segmentation accuracy [29]. Future research may focus on refining the model structure. One potential direction is to incorporate enhanced attention mechanisms that consider not only channel-level information but also spatial relationships. Another is to develop improved fusion strategies, such as explicitly modeling interactions among different modalities during feature fusion, to improve performance in challenging clinical scenarios.

4. Conclusion

This study successfully developed an end-to-end multimodal deep fusion network that integrates a channel attention mechanism and multi-scale feature extraction, aiming to address the diagnostic challenges of complex lesions in medical imaging. Experimental results on publicly available glioma and lung tumor multimodal datasets showed that the proposed model significantly outperformed the weighted averaging method, PCA-based fusion algorithm, and MM-CNN in terms of classification accuracy, ROC-AUC, and Dice coefficient. In glioma diagnosis, the model achieved a classification accuracy of 93.6%, a ROC-AUC of 0.952, and an overall Dice coefficient of 0.885. In lung tumor diagnosis, the classification accuracy was 92.1%, the ROC-AUC reached 0.945, and the Dice coefficients

for tumors with regular and irregular shapes were 0.90 and 0.86, respectively. These results clearly demonstrate the effectiveness and superiority of the model in complex lesion diagnosis. The strength of this study lies in its ability to enhance focus on key features through the channel attention mechanism, to fully capture multi-scale characteristics of lesions through multi-scale feature extraction, and to support the learning of internal relationships among multimodal data through the end-to-end architecture. These improvements bring new momentum to the development of multimodal medical image fusion techniques and provide a novel and efficient solution for clinical multi-source image-assisted diagnosis. This approach has the potential to significantly improve early detection of complex diseases and reduce the risk of misdiagnosis and missed diagnosis.

References

1. S. Hussain, et al., "Modern diagnostic imaging technique applications and risk factors in the medical field: a review," *BioMed Res. Int.*, vol. 2022, no. 1, p. 5164970, 2022, doi: 10.1155/2022/5164970.
2. H. Wang, G. Zhang, Y. Zhao, F. Lai, W. Cui, J. Xue, and Y. Lin, "Rpf-eld: Regional prior fusion using early and late distillation for breast cancer recognition in ultrasound images," in *2024 IEEE Int. Conf. Bioinformatics Biomedicine (BIBM)*, 2024, pp. 2605–2612, doi: 10.1109/BIBM62325.2024.10821972.
3. N. S. Radder, S. Sonar, A. V. I. N. A. S. H. Nanivadekar, and S. B. Radder, "A step forward in elucidating the mystery in neuroimaging: PET-MRI fusion," in *Proc. European Congress of Radiology (ECR)*, 2019, doi: 10.26044/ecr2019/C-0697.
4. X. Wu, Y. Sun, and X. Liu, "Multi-class classification of breast cancer gene expression using PCA and XGBoost," *Preprints*, 2024, doi: 10.20944/preprints202410.1775.v1.
5. M. Raghavan, "Conventional modalities and novel, emerging imaging techniques for musculoskeletal tumors," *Cancer Control*, vol. 24, no. 2, pp. 161–171, 2017, doi: 10.1177/107327481702400208.
6. X. Shi, Y. Tao, and S. C. Lin, "Deep neural network-based prediction of B-cell epitopes for SARS-CoV and SARS-CoV-2: Enhancing vaccine design through machine learning," in *Proc. 2024 4th Int. Signal Process., Commun. Eng. Manag. Conf. (ISPCEM)*, Nov. 2024, pp. 259–263, doi: 10.1109/ISPCEM64498.2024.00050.
7. S. Kumar, S. Rani, S. Sharma, and H. Min, "Multimodality fusion aspects of medical diagnosis: A comprehensive review," *Bioengineering*, vol. 11, no. 12, p. 1233, 2024, doi: 10.3390/bioengineering11121233.
8. A. Kumar, "Deep learning for multi-modal medical imaging fusion: Enhancing diagnostic accuracy in complex disease detection," *Int. J. Eng. Technol. Res. Manag.*, vol. 6, no. 11, p. 183, 2022, doi: 10.5281/zenodo.15033792.
9. Y. Wang, Y. Wen, X. Wu, L. Wang, and H. Cai, "Assessing the role of adaptive digital platforms in personalized nutrition and chronic disease management," *World J. Innov. Mod. Technol.*, vol. 8, no. 1, Jan. 2025, doi: 10.53469/wjimt.2025.08(01).05.
10. Y. Wen, X. Wu, L. Wang, H. Cai, and Y. Wang, "Application of nanocarrier-based targeted drug delivery in the treatment of liver fibrosis and vascular diseases," *J. Med. Life Sci.*, vol. 1, no. 2, pp. 63–69, 2025, doi: 10.71222/bc9c9s73.
11. S. Amal, L. Safarnejad, J. A. Omiye, I. Ghanzouri, J. H. Cabot, and E. G. Ross, "Use of multi-modal data and machine learning to improve cardiovascular disease care," *Front. Cardiovasc. Med.*, vol. 9, p. 840262, 2022, doi: 10.3389/fcvm.2022.840262.
12. T. Zhang, B. Zhang, F. Zhao, and S. Zhang, "COVID-19 localization and recognition on chest radiographs based on Yolov5 and EfficientNet," in *2022 7th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, 2022, pp. 1827–1830, doi: 10.1109/ICSP54964.2022.9778327.
13. Y. Wang, Y. Wen, X. Wu, L. Wang, and H. Cai, "Modulation of gut microbiota and glucose homeostasis through high-fiber dietary intervention in type 2 diabetes management," *World J. Innov. Mod. Technol.*, vol. 7, no. 5, Oct. 2024, doi: 10.53469/wjimt.2024.07(06).04.
14. H. Ziang, J. Zhang, and L. Li, "Framework for lung CT image segmentation based on UNet++," 2025, *arXiv preprint arXiv:2501.02428*, doi: 10.48550/arXiv.2501.02428.
15. I. Domingues, G. Pereira, P. Martins, H. Duarte, J. Santos, and P. H. Abreu, "Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET," *Artif. Intell. Rev.*, vol. 53, pp. 4093–4160, 2020, doi: 10.1007/s10462-019-09789-w.
16. Y. Deng, W. Cui, and X. Liu, "Head Tumor Segmentation and Detection Based on ResUNet," in *Proc. 5th Int. Conf. Signal Process. Mach. Learn.*, 2024, doi: 10.54254/2755-2721/99/20251810.
17. Y. Wang, Y. Wen, X. Wu, and H. Cai, "Application of ultrasonic treatment to enhance antioxidant activity in leafy vegetables," *Int. J. Adv. Appl. Sci. Res.*, vol. 3, pp. 49–58, 2024, doi: 10.5281/zenodo.14275691.
18. S. M. Hosseini and M. S. Baghshah, "Dilated balanced cross entropy loss for medical image segmentation," 2024, *arXiv Preprint*, arXiv:2412.06045, doi: 10.48550/arXiv.2412.06045.
19. J. B. Qiao, et al., "Vitamin A-decorated biocompatible micelles for chemogene therapy of liver fibrosis," *J. Control. Release*, vol. 283, pp. 113–125, 2018, doi: 10.1016/j.jconrel.2018.05.032.

20. G. Qu, et al., "Self-assembled micelles based on N-octyl-N′-phthalyl-O-phosphoryl chitosan derivative as an effective oral carrier of paclitaxel," *Carbohydr. Polym.*, vol. 207, pp. 428–439, 2019, doi: 10.1016/j.carbpol.2018.11.099.
21. Y. Wang, L. Wang, Y. Wen, X. Wu, and H. Cai, "Precision-engineered nanocarriers for targeted treatment of liver fibrosis and vascular disorders," *World J. Innov. Mod. Technol.*, vol. 8, no. 1, Jan. 2025, doi: 10.53469/wjimt.2025.08(01).07.
22. J. Zhu, et al., "Multimodal nanoimmunotherapy engages neutrophils to eliminate *Staphylococcus aureus* infections," *Nat. Nanotechnol.*, vol. 19, no. 7, pp. 1032–1043, 2024, doi: 10.1038/s41565-024-01648-8.
23. Y. Wang, Y. Wen, X. Wu, and H. Cai, "Comprehensive evaluation of GLP1 receptor agonists in modulating inflammatory pathways and gut microbiota," *World J. Innov. Mod. Technol.*, vol. 7, no. 5, 2024, doi: 10.53469/wjimt.2024.07(06).23.
24. I. K. Lee, et al., "Micromolded honeycomb scaffold design to support the generation of a bilayered RPE and photoreceptor cell construct," *Bioact. Mater.*, vol. 30, pp. 142–153, 2023, doi: 10.1016/j.bioactmat.2023.07.019.
25. N. Yodsanit, et al., "Targeted PERK inhibition with biomimetic nanoclusters confers preventative and interventional benefits to elastase-induced abdominal aortic aneurysms," *Bioact. Mater.*, vol. 26, pp. 52–63, 2023, doi: 10.1016/j.bioactmat.2023.02.009.
26. I. Domingues, G. Pereira, P. Martins, H. Duarte, J. Santos, and P. H. Abreu, "Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET," *Artif. Intell. Rev.*, vol. 53, pp. 4093–4160, 2020, doi: 10.1007/s10462-019-09788-3.
27. Y. Wang, M. Shen, L. Wang, Y. Wen, and H. Cai, "Comparative modulation of immune responses and inflammation by n-6 and n-3 polyunsaturated fatty acids in oxylipin-mediated pathways," *World J. Innov. Mod. Technol.*, vol. 7, no. 4, Aug. 2024, doi: 10.53469/wjimt.2024.07(05).17.
28. A. M. Giantini-Larsen, S. Pannullo, and R. G. Juthani, "Challenges in the diagnosis and management of low-grade gliomas," *World Neurosurg.*, vol. 166, pp. 313–320, 2022, doi: 10.1016/j.wneu.2022.07.079.
29. S. C. Kotoulas, et al., "A thorough review of the clinical applications of artificial intelligence in lung cancer," *Cancers*, vol. 17, no. 5, p. 882, 2025, doi: 10.3390/cancers17050882.