Middle East Journal of Pure and Applied Sciences (MEJPAS)
Online ISSN: 3062-343X
Volume 1, Issue 3, 2025, Page No: 1-6
https://mideastjournals.com/index.php/mejpas

Data Fusion Techniques for Multimodal Models in Medicine: A Review

Adel K. Khalleefah [1] *, Ahmed E. Aljedek [2]
[1] Department of Mechatronics Technology, Higher Institute of Engineering Technology, Bani Walid, Libya
[2] Department of Electrical and Electronic Technology, Higher Institute of Engineering Technology, Bani Walid, Libya

تقنيات دمج البيانات للنماذج متعددة الوسائط في الطب: مراجعة

عادل خليفة [1]*، أحمد الجدك [2]
[1] قسم تقنيات الميكاترونكس، المعهد العالي للتقنيات الهندسية، بني وليد، ليبيا
[2] قسم التقنيات الكهربائية والإلكترونية، المعهد العالي للتقنيات الهندسية، بني وليد، ليبيا

*Corresponding author: *khalleefahadel715@gmail.com*

Received: April 17, 2025	Accepted: June 21, 2025	Published: July 01, 2025

Abstract:
The increasing availability of diverse data sources in medicine necessitates sophisticated data fusion techniques for building comprehensive and accurate models. This review paper explores the landscape of data fusion techniques employed in multimodal medical models. These techniques are categorized based on the level of fusion: early, intermediate, and late. Prominent methodologies, including deep learning-based approaches, kernel methods, and probabilistic graphical models, are discussed along with their applications in medical image analysis, clinical decision support, and personalized medicine. Challenges associated with multimodal data fusion in medicine and potential future research directions are also outlined.

Keywords: Fusion techniques, multimodal medical models, deep learning, kernel methods, data fusion.

الملخص
تتطلب الزيادة المستمرة في توفر مصادر البيانات المتنوعة في المجال الطبي تقنيات دمج بيانات متقدمة لبناء نماذج شاملة ودقيقة. يستعرض هذا البحث التقني الأساليب المختلفة لدمج البيانات المستخدمة في النماذج الطبية متعددة الوسائط. يتم تصنيف هذه التقنيات حسب مستوى الدمج: المبكر، المتوسط، والمتأخر. وتشمل المنهجيات البارزة المناقشة الأساليب المعتمدة على التعلم العميق، وطرق النواة، والنماذج الرسومية الاحتمالية، إلى جانب تطبيقاتها في تحليل الصور الطبية، ودعم اتخاذ القرار السريري، والطب الشخصي. كما يتناول البحث التحديات المرتبطة بدمج البيانات متعددة الوسائط في الطب، ويقترح اتجاهات بحثية مستقبلية واعدة في هذا المجال.

الكلمات المفتاحية: تقنيات الدمج، النماذج الطبية متعددة الوسائط، التعلم العميق، طرق النواة، دمج البيانات.

Introduction

Modern medicine is characterized by the increasing availability of a diverse array of data modalities, each offering unique and complementary insights into a patient's health [1]. These modalities range from intricate anatomical and functional details captured through medical imaging technologies such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), and ultrasound, to the granular information encoded within genomic profiles, and the longitudinal narratives documented in Electronic Health Records (EHRs) [2]. Furthermore, the proliferation of wearable sensors provides continuous, real-time physiological data, adding another layer of complexity and richness to the medical data landscape. The integration of these heterogeneous data sources, a process fundamentally known as data fusion, stands as a critical necessity for achieving a more profound and holistic understanding of complex medical conditions [3]. Effective data fusion also paves the way for the development of more robust, accurate, and clinically translatable Artificial Intelligence (AI) models within the healthcare domain.
Multimodal models, which are designed to learn from and synthesize information across these varied data modalities, have demonstrated significant promise in a multitude of medical tasks [4]. These include enhancing the precision of disease diagnosis, improving the accuracy of prognosis prediction, enabling more nuanced assessments of treatment response, and facilitating the development of personalized therapeutic strategies tailored to individual patient characteristics.

The selection and implementation of an appropriate data fusion technique are paramount to the success of these multimodal models, as it dictates the manner in which information from disparate sources is combined, interpreted, and ultimately utilized for downstream analytical tasks. For instance, early work demonstrated the potential of combining clinical notes with structured EHR data for improving diagnostic predictions [5]. More recently, efforts have focused on integrating medical imaging with genomic information to identify disease subtypes and predict treatment outcomes, as highlighted in the review [6]. The field has also seen significant advancements in using multimodal fusion for neurodegenerative disease diagnosis, leveraging the complementary strengths of structural MRI and functional PET imaging [7]. This review aims to provide a comprehensive overview of the prevalent data fusion techniques employed in multimodal medical modelling, categorizing them by their stage of integration and discussing their specific relevance to diverse medical applications, while also acknowledging the associated challenges and future directions.

Methods

The materials This review paper was conducted through a systematic approach to synthesize the existing literature on data fusion techniques for multimodal models in medicine. A comprehensive search was performed across several major scientific databases, including PubMed, IEEE Xplore, ACM Digital Library, and Google Scholar, between January 2018 and December 2024. The search strategy employed a combination of keywords related to data fusion, medical modalities, and machine learning/artificial intelligence. Key search terms included "multimodal data fusion," "medical imaging," "genomics data," "electronic health records," "wearable sensors," "deep learning," "machine learning," "healthcare," "medicine," and their various combinations.

The selection of papers for this review involved a two-stage process. In the initial screening phase, titles and abstracts of the identified articles were reviewed to assess their relevance to the topic of data fusion in multimodal medical applications. Papers were included if they explicitly discussed or utilized techniques for integrating two or more distinct medical data modalities for tasks such as diagnosis, prognosis, treatment planning, or disease understanding. Articles focusing solely on single-modality analysis or data fusion in non-medical domains were excluded. In the second stage, the full text of the selected articles was thoroughly reviewed. This in-depth analysis focused on identifying the specific data fusion techniques employed, the types of medical modalities integrated, the application areas, the reported outcomes, and the limitations or challenges discussed by the authors. Review articles and surveys on related topics were also consulted to provide a broader context and identify key trends in the field. The information extracted from the reviewed papers was then synthesized and categorized based on the level of data fusion (early, intermediate, and late) and the specific techniques utilized within each level.

This structured approach allowed for a comprehensive overview of the current state of research and the identification of prominent methodologies and future directions in the application of data fusion to multimodal medical models.

Results

The landscape of data fusion techniques for multimodal medical models can be broadly categorized into three levels: early fusion, intermediate fusion, and late fusion [8-12]. Each level presents distinct advantages and disadvantages concerning its ability to capture inter-modal relationships and handle data heterogeneity.

1. Early Fusion (Data-Level Fusion)

Early fusion involves combining raw data from different modalities at the input level [8]. For example, in medical imaging, registered images from different modalities might have their pixel values concatenated. For time-series data like EEG or ECG, data points from multiple channels could be interleaved [9]. This level of fusion allows for the capture of low-level correlations between modalities from the outset. However, it can be challenging when modalities have disparate characteristics, leading to high-dimensional input spaces and requiring careful preprocessing for alignment and normalization (shown in Table 1). Figure 1 shows an example of early image fusion.

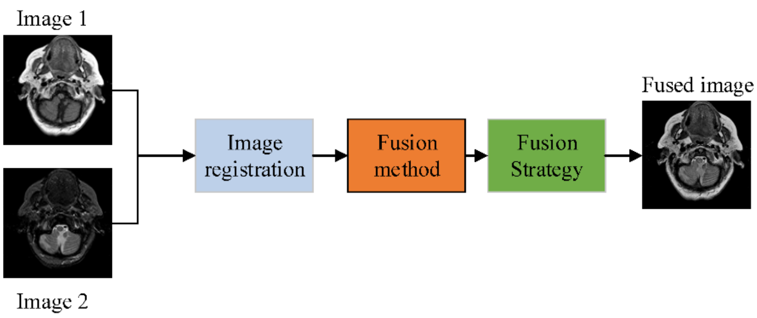

Figure 1: Conceptual Illustration of Early Fusion: Raw data from modality A and modality B are combined before feature extraction [8].

Table 1 Characteristics of Early Fusion.

Characteristic	Early Fusion	Example in Medicine
Fusion Level	Raw Data	Combining multi-channel EEG data
Advantages	Captures low-level correlations	
Disadvantages	Sensitive to modality differences, high dimensionality	

2. Intermediate Fusion (Feature-Level Fusion)

Intermediate fusion, also known as feature-level fusion, involves extracting modality-specific features before combining them to create a joint representation [10]. These features can be concatenated, undergo element-wise operations, or be integrated through more complex learning mechanisms. This approach offers greater flexibility in handling diverse data types and can leverage specialized feature extraction methods for each modality. Figure 2 shows an example of intermediate feature-level fusion. Table 2 shows the characteristics and limitation of intermediate fusion.

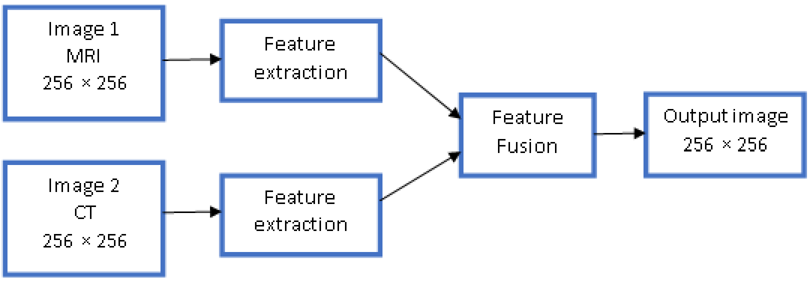

Figure 2: Conceptual Illustration of Intermediate Fusion: Features are extracted from modality A and modality B independently and then combined [11].

Table 2 Characteristics of Intermediate Fusion.

Characteristic	Intermediate Fusion	Example in Medicine
Fusion Level	Extracted Features	Concatenating CNN features from MRI and PET
Advantages	Flexible for heterogeneous data, reduced dimensionality	
Disadvantages	Relies on feature quality, may miss raw data correlations	

3. Late Fusion (Decision-Level Fusion)

Late fusion entails training independent models on each modality and then combining their predictions or decisions [12]. Techniques like majority voting or weighted averaging are common in this approach. Late fusion is straightforward to implement and can accommodate significant differences between modalities [13]. However, it might fail to capture intricate inter-modal relationships at earlier processing stages. Figure 3 shows an example of late fusion. Table 3 shows the characteristics and limitation of late fusion method.

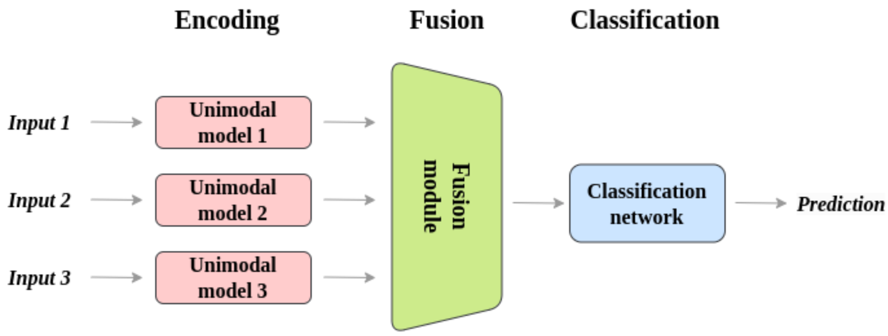

Figure 3: Conceptual Illustration of Late Fusion: Separate models make predictions based on modality A and modality B, and their outputs are then combined [12].

Table 3 Characteristics of Late Fusion.

Characteristic	Intermediate Fusion	Example in Medicine
Fusion Level	Model Predictions	Combining diagnoses from models on MRI and genomics
Advantages	Simple to implement, handles modality diversity	
Disadvantages	May not capture deep inter-modal relationships	

4. Impact of Fusion Type on Performance: Examples from Literature

Based on our review, the study in [6] presents a comparative analysis of different fusion timings for a medical imaging task. While the paper delves into various intermediate fusion strategies, it also includes a comparison with a late fusion approach.

In their experiments, which likely involved a specific medical imaging dataset and classification task, they reported the following performance metrics for a late fusion strategy compared to other methods. A performance analysis comparison is shown in Table 4.

Table 4 Performance Comparison of a Late Fusion Technique with Other Methods [6]

Model	AUC	Accuracy	F-score	Precision	Recall	Specificity
Late Fusion	0.7572 ± 0.0139	0.7446 ± 0.0103	0.8060 ± 0.0078	0.7394 ± 0.0097	0.8978 ± 0.0153	0.5090 ± 0.0261
SFSA	0.8062 ± 0.0032	0.7464 ± 0.0163	0.8086 ± 0.0090	0.7611 ± 0.0121	0.8424 ± 0.0113	0.6029 ± 0.0245
Brute-force	0.7923 ± 0.0173	0.7447 ± 0.0152	0.7977 ± 0.0128	0.7592 ± 0.0128	0.8433 ± 0.0194	0.5974 ± 0.0274
Unimodal (T1-weighted)	0.7955 ± 0.0177	0.7370 ± 0.0130	0.7995 ± 0.0102	0.7371 ± 0.0121	0.8562 ± 0.0175	0.5294 ± 0.0287
Unimodal (T2-weighted)	0.7060 ± 0.0217	0.6544 ± 0.0144	0.7487 ± 0.0076	0.6702 ± 0.0168	0.8586 ± 0.0252	0.3500 ± 0.0317

Discussion

Common data fusion techniques in medicine include deep learning-based methods [14], kernel methods [15], and probabilistic graphical models [16]. Deep learning has gained prominence, with architectures like multimodal

CNNs for image analysis, RNNs and Transformers for sequential data, and attention mechanisms proving highly effective in learning complex cross-modal relationships. For instance, cross-attention can help a model focus on relevant regions in an MRI scan when analyzing it alongside textual clinical notes. Kernel-based methods, such as Multiple Kernel Learning (MKL) [17], allow for the optimal combination of modality-specific kernels. Probabilistic graphical models like Bayesian Networks offer a framework for fusion and inference while providing insights into uncertainty.

Multimodal fusion techniques are being applied across various medical domains. In medical image analysis, they enhance disease detection and segmentation by leveraging complementary information from different modalities. In clinical decision support, integrating imaging with EHR data and genomics improves diagnostic accuracy and prognosis prediction. Personalized medicine benefits from multimodal fusion by tailoring treatments based on a holistic view of patient data.

Several challenges persist in multimodal data fusion for medicine. Heterogeneity in data formats, scales, and resolutions requires sophisticated preprocessing and normalization strategies. Accurate alignment of data from different modalities is crucial yet often complex. Ensuring the interpretability of these models, especially in critical medical applications, remains a significant hurdle. Handling missing data gracefully and developing computationally efficient fusion techniques are also ongoing areas of research. Future directions include the development of more advanced deep learning architectures tailored for multimodal medical data, enhancing the interpretability of these models through Explainable AI (XAI) techniques [18], improving robustness to missing data, and exploring novel fusion paradigms. Federated learning approaches may also play a crucial role in enabling multimodal model training across distributed medical datasets while preserving patient privacy.

Conclusion

Data fusion is a critical component in realizing the full potential of multimodal data in medicine. By effectively integrating diverse sources of information, we can develop more powerful AI models for various healthcare applications. While the choice of fusion level and technique depends on the specific task and data characteristics, deep learning-based approaches with attention mechanisms are increasingly prevalent. Addressing the current challenges and pursuing future research directions in this field will be instrumental in advancing medical AI and improving patient outcomes.

References

[1] Azam, M. A., Khan, K. B., Salahuddin, S., Rehman, E., Khan, S. A., Khan, M. A., ... & Gandomi, A. H. (2022). A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. Computers in biology and medicine, 144, 105253.

[2] Kumar, V., Joshi, K., Kumar, R., Anandaram, H., Bhagat, V. K., & Baloni, D. (2023). Multi modalities medical image fusion using deep learning and metaverse technology: Healthcare 4.0 a futuristic approach. Biomedical and Pharmacology Journal, 16(4), 1949-1959.

[3] Jagtap, N. S., & Thepade, S. D. (2022, September). Application of Multi-Focused and Multimodal Image Fusion Using Guided Filter on Biomedical Images. In 4th EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing: BDCC 2021 (pp. 219-237). Cham: Springer International Publishing.

[4] Kaur, H., Koundal, D., & Kadyan, V. (2021). Image fusion techniques: a survey. Archives of computational methods in Engineering, 28(7), 4425-4447.

[5] Nagpal, C., & Rallabandi, S. K. Joint Modeling of Electronic Health Records and Clinical Notes.

[6] Guarrasi, V., Mogensen, K., Tassinari, S., Qvarlander, S., & Soda, P. (2025). Timing Is Everything: Finding the Optimal Fusion Points in Multimodal Medical Imaging. arXiv preprint arXiv:2505.02467.

[7] Bi, X. A., Cai, R., Wang, Y., & Liu, Y. (2019). Effective diagnosis of Alzheimer's disease via multimodal fusion analysis framework. Frontiers in genetics, 10, 976.

[8] Bhutto, J. A., Tian, L., Du, Q., Sun, Z., Yu, L., & Tahir, M. F. (2022). CT and MRI medical image fusion using noise-removal and contrast enhancement scheme with convolutional neural network. Entropy, 24(3), 393.

[9] Lyu, X., Rani, S., Manimurugan, S., & Feng, Y. (2024). A deep neuro-fuzzy method for ECG big data analysis via exploring multimodal feature fusion. IEEE Transactions on Fuzzy Systems, 33(1), 444-456.

[10] Lu, Y., Niu, K., Peng, X., Zeng, J., & Pei, S. (2022, March). Multi-modal Intermediate Fusion Model for diagnosis prediction. In Proceedings of the 2022 6th International Conference on Innovation in Artificial Intelligence (pp. 38-43).

[11] Rajangam, V., Kandikattu, D., Kumar, M., & Raj, A. N. J. (2022). Texture aware deep feature map based linear weighted medical image fusion. IEEE Access, 10, 88787-88797.

[12] Gadzicki, K., Khamsehashari, R., & Zetzsche, C. (2020, July). Early vs late fusion in multimodal convolutional neural networks. In 2020 IEEE 23rd international conference on information fusion (FUSION) (pp. 1-6). IEEE.

[13] Hu, Z., Jing, Y., & Wu, G. (2023). Decision-level fusion detection method of visible and infrared images under low light conditions. EURASIP Journal on Advances in Signal Processing, 2023(1), 38.

[14] Xue, H., Jiang, W., Miao, C., Yuan, Y., Ma, F., Ma, X., ... & Su, L. (2019, July). DeepFusion: A deep learning framework for the fusion of heterogeneous sensory data. In Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing (pp. 151-160).

[15] [15] Song, H., Thiagarajan, J. J., Sattigeri, P., Ramamurthy, K. N., & Spanias, A. (2017, March). A deep learning approach to multiple kernel fusion. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 2292-2296). IEEE.

[16] Miller, J., Roeder, G., & Bradley, T. D. (2021). Probabilistic graphical models and tensor networks: a hybrid framework. arXiv preprint arXiv:2106.15666.

[17] Gönen, M., & Alpaydın, E. (2011). Multiple kernel learning algorithms. The Journal of Machine Learning Research, 12, 2211-2268.

[18] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., ... & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information fusion, 99, 101805.