REVIEW ARTICLE

A Recent Survey on Multi-modal Medical Image Fusion

Vandit Akhilesh Barola[1], **Prabhishek Singh**[1,*] **and Manoj Diwakar**[2]

[1] School of Computer Science Engineering and Technology, Bennett University, Greater Noida, Uttar Pradesh 201310, India
[2] Department of CSE, Graphic Era Deemed to be University, Dehradun, Uttarakhand 248002, India

Abstract

Fusion of multi-modal medical images has transformed healthcare by overcoming the limitations of single-modality imaging, where modalities such as CT, MRI, PET, and SPECT provide complementary information. This review systematically traces the evolution of multi-modal medical image fusion from conventional mathematical models to state-of-the-art artificial intelligence (AI) techniques. We examine the transition from classical approaches—such as multiscale transformations, wavelet decompositions, and sparse representation—to modern deep learning methods, including convolutional neural networks, generative adversarial networks, and transformer architectures. Key limitations of existing methods are highlighted, including limited interpretability, insufficient preservation of modality-specific details, poor cross-dataset generalization, and inadequate post-fusion refinement. The review categorizes fusion strategies into three hierarchical levels—pixel-level, feature-level, and decision-level—and analyzes their benefits and computational complexities. We provide a comparative evaluation of hybrid methods combining classical transform-based approaches with deep learning models, which show improved preservation of anatomical structures and functional information. Major clinical applications are discussed in oncology (tumor detection, staging), neurology (brain tumor localization, surgical planning), ophthalmology (disease characterization), and orthopedics (fracture detection). Challenges to clinical adoption—such as modality misalignment, information loss, high computational requirements, and lack of universal benchmarks—are examined alongside emerging concerns over data privacy and security in telehealth. We also analyze interpretability frameworks, modality-preservation techniques, cross-dataset generalization strategies, and post-fusion refinement methods. The review concludes with strategic recommendations to develop interpretable, real-time AI fusion models with robust clinical validation and standardized benchmarks to bridge the gap between theoretical advances and practical clinical integration, ultimately enhancing diagnostic accuracy and patient outcomes.

Keywords: biomedical informatics, smart healthcare, artificial intelligence in medicine, precision medicine, digital health.

Submitted: 26 July 2025
Accepted: 17 August 2025
Published: 07 November 2025

Vol. 1, **No.** 3, 2025.
10.62762/BISH.2025.414869

***Corresponding author:**
✉ Prabhishek Singh
prabhisheksingh1988@gmail.com

Citation
Barola, V. A., Singh, P., & Diwakar, M. (2025). A Recent Survey on Multi-modal Medical Image Fusion. *Biomedical Informatics and Smart Healthcare*, 1(3), 89–97.

1 Introduction

Technology development has advanced enough to transform various sectors, including healthcare. These advancements have significantly benefited medical imaging, improving diagnosis, treatment planning, and patient monitoring. Image fusion is a technology used to combine detailed information from multiple images of the same subject into a single image. The primary goal of image fusion is to enhance the overall quality and interpretability of the image. Fusion can be performed at various levels, such as the pixel level, feature level, and decision level [1]. This technique is extensively used in various fields, including remote sensing, surveillance, and medical imaging. Direct image registration, unlike iris segmentation, involves pixel-level fusion of raw image data, which provides high spatial and spectral information. The difference between feature-level and decision-level fusion lies in the approach taken to merge key features or the results of multiple classifiers or decision models, respectively [2]. The main benefits of image fusion include enhancing contrast, reducing redundancy, and improving feature representation to support analytical tasks. In medical imaging, image fusion is valuable for improving diagnosis, treatment planning, and disease monitoring by combining complementary information from two or more modalities. Therefore, image fusion has become increasingly critical, as it is becoming more automated and accurate through AI-based applications [3]. The example of medical image fusion is shown in Figure 1.

Figure 1. Example of multi-modal medical image fusion.

1.1 Rapid Obstacles in Present Methods

Although substantial achievements have been made in the field of multimodal medical image fusion, there are a number of major challenges that are yet to be addressed in the methods that exist today, hinder their practical use in clinical settings. Explainability still is a crucial issue because most fusion schemes on deep learning are black boxes and thereby clinicians perceive a hard time in comprehending how the decisions have been made and being biased about the fused outcome. This transparency is a great hindrance towards clinical implementation, where health experts need to have a clear notion of how various choices play a role in the end product fused image.

Imperfect preservation of the modality specific information is another egregious shortcoming of the existing fusion methodologies. Most of the current methods that have been developed have a tendency to suppress or submerge specialties that are peculiar to individual modalities of the imaging during the fusion process. As an example, the metabolic data of PET scans or the excellent anatomical details of MRI may get corrupted during the fusion with other modalities, which can result in loss of diagnostically important information.

Such is the case with generalization among heterogeneous datasets, such a problem is a big deal with existing fusion methods. The majority of the techniques are trained and validated against datasets that represent a specific imaging protocol, patient demographics or clinical condition and exhibit poor generalisation to alternative clinical practice or patient demographics. This shortcoming poses a major challenge to the practical implementation of fusion techniques in different health organizations.

The refinement processes after fusion used in existing literature are usually incomplete and only lay emphasis on the actual fusion process, which does not take into account refinement components of quality improvement and artifact correction of fused images. They depend on robust post-processing to be able to make sure that fused images can meet clinical standards of diagnostic accuracy and visual quality.

Medical imaging is a vital tool in diagnosing and managing many diseases. There are several imaging modalities to address different medical conditions. However, every modality has strengths and weaknesses, making it impossible to have a complete picture for proper diagnosis. For example, CT scans are regarded as best for bone structure but offer poor soft tissue contrast, while MRI offers high soft tissue contrast but no functional insight [4, 5]. To address this issue, multimodal medical image fusion is utilized, which combines information received from various imaging modalities, creating a single, information-rich image that greatly enhances diagnostic accuracy, treatment planning, and most importantly, efficiency. Although there are possibly hundreds of applications, one of the most important applications of multimodal fusion is in the detection of cancer, where PET-CT imaging is utilized extensively

[6, 7]. PET scans are often used to identify tumors in the early stages by marking metabolic activity, while CT scans provide anatomical details, allowing accurate localization of lesions. In the field of neurology, combining functional MRI (fMRI) and structural MRI allows brain activity mapping, which helps in pre-surgical planning for disabilities and conditions like brain tumors and epilepsy. Likewise, in ophthalmology, combining Color Fundus Photography (CFP) and Optical Coherence Tomography (OCT) increases the accuracy of diagnosing conditions like glaucoma and diabetic retinopathy, as it combines structural and vascular details of the retina. Also, combining X-rays with bone scans enhances the detection of faint fractures that are missed in typical X-ray images alone. Multimodal image fusion is typically categorized into three types [8, 9]. Pixel-level fusion merges raw intensity values from images directly, while feature-level fusion extracts and merges significant patterns from various modalities. Decision-level fusion merges outputs of distinct classifiers trained for individual imaging modalities. Recent advancements in deep learning techniques, such as Convolutional Neural Networks (CNNs) and Transformers, have greatly improved the quality of fused images by enhancing feature extraction and reducing artifacts. In addition, difference-enhancing fusion methods like the Non-Subsampled Contourlet Transform (NSCT) help retain finer details while reducing image noise [10]. A combination of the strengths of different imaging techniques, multimodal medical image fusion provides a more comprehensive and accurate image of a patient's condition. With each advancing study on fusion techniques, we await even greater precision, efficiency, and accessibility in medical imaging, ultimately benefiting both clinicians and patients [11].

1.2 Distinguishing Features of This Survey

Such a survey stands out among the established literature on this topic due to the systematic way of covering the major shortcomings described above in detail by analyzing interpretability frameworks, modalities-specific preservation strategies, cross-dataset generalization strategies, and sophisticated post-fusion refinement strategies. In comparison to other surveys conducted in the past, the key difference of this work is that it critically reviews the area of clinical applicability and deployment issues currently ill-defined in the existing literature with little exploration of an algorithmic evolution.

Multimodal medical image fusion is an emerging field focused on integrating different imaging modalities, such as CT, PET, MRI, and SPECT, to enhance diagnostic accuracy by combining structural and functional information. Current research is primarily aimed at deep learning-based methods, particularly Convolutional Neural Networks and Transformer models, which have shown enhanced performance in feature extraction and fusion [12, 13]. Despite this, their use in wider clinical applications is limited by associated difficulties such as high computational costs and the necessity of large, annotated datasets. Since the limitations of hybrid fusion methods were addressed to enhance spatial and spectral details, reduce memory issues, and minimize low-resolution artifacts, researchers are exploring hybrid fusion with deep learning-based models like Non-Subsampled Contourlet Transform (NSCT), Discrete Wavelet Transform (DWT), and Dual Tree Complex Wavelet Transform (DTCWT) to reduce artifacts. Furthermore, dealing with incomplete or imbalanced data is also one of the key challenges in medical image fusion, as real-world datasets often contain a single modality, yet multiple modalities are needed [14, 15]. In recent works, missing modalities are now being generated using GANs and hybrid data augmentation strategies, improving fusion reliability. The optimization of fusion rules is also another important aspect, as existing approaches use hand-crafted strategies that do not generalize well across different use cases. Recent improvements also include data-driven optimization, such as bio-inspired algorithms and neural network-based weighting models, to improve fusion efficiency and adaptability, as described in [10, 11]. The process of medical image fusion is shown in Figure 2.

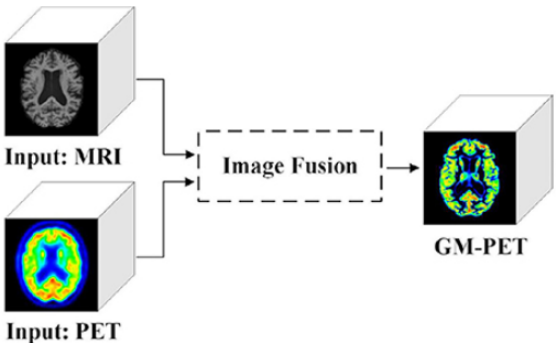

Figure 2. Process of multi-modal medical image fusion.

Despite significant progress in fusion accuracy and computing efficiency, clinical validation remains very difficult. In practice, few fusion approaches are

robustly validated with patient data and are therefore not directly applicable to clinical scenarios. Most are validated only on standard datasets. In addition, fusion methods are computationally intensive, and both training and real-time application require costly hardware limitations that hinder their use in healthcare environments with constrained resources [16]. Moreover, because different studies employ varying datasets and benchmarks, direct comparisons among them are not feasible, and there is no universal benchmark dataset for measuring the degree of fusion performance. Beyond diagnostics, researchers are also investigating telehealth and medical data security. One example of an emerging interface between fusion research and cybersecurity in healthcare is the WatMIF algorithm, which combines fusion methods with encryption and watermarking to enhance the security of medical image transmission [17]. Additionally, new methods are being explored to further improve edge clarity, noise removal, and spatial feature preservation, such as boundary-measured Pulse-Coupled Neural Networks (PCNN) and energy attribute fusion schemes. To bridge the gap between theoretical advancements and practical medical applications, future studies are expected to focus on developing deep learning models that are more efficient and computationally less demanding, standardizing benchmarking datasets, and ensuring clinical validation [18].

2 Related Work

Medical imaging has long been a cornerstone of diagnostic medicine, providing insights that guide clinical decision-making. However, no single imaging technique can capture the complete picture—MRI offers detailed soft tissue contrast, CT excels at detecting bone abnormalities, and PET reveals metabolic activity. Because of these differences, integrating multiple imaging modalities has become a necessity rather than a luxury. Multimodal image fusion allows for a more holistic view, combining anatomical and functional data to improve visualization and enhance diagnostic accuracy. This approach mitigates the limitations of standalone imaging methods and strengthens the reliability of medical assessments. In existing works, fusion techniques have been broadly classified into three categories, namely pixel-level, feature-level, and decision-level fusion. Early models made use of multiscale decomposition, fuzzy logic, and machine learning to improve fusion methodologies. Of these, wavelet transforms, and contourlet-based methods were popular since they enhanced image registration and segmentation. CNN and U-Net architectures rose in 2015 to refine these methods and provide a solid foundation upon which contemporary deep learning algorithms are based. Hybrid fusion models have been developed recently by integrating one traditional transform-based scheme with a prototype deep learning scheme for a good balance between accuracy and computational efficiency.

2.1 Interpretability Challenges in Existing Fusion Methods

That the deep learning-based fusion approaches are a black box has proven another critical obstacle to clinical application. Those surveys have failed to capture the interpretability crisis in the multimodal fusion methods, yet they dwell and focus on performance measures without considering the high relevance of explainable AI in medical practice. Recent attempts at various gradient-based visualization methods to visualize gradient flow on visualization, namely Grad-CAM and LIME generated (locally interpreted model explanations), have been implemented as fusion applications to output explanations of which areas and modalities show the most contribution to the overall fused output. Nevertheless, the methods are still no less limited in articulating clinically significant explanations.

The interpretability involving attention mechanisms has held potential to make the fusion decisions transparent. Fusion architectures using transformers also include self-attention and cross-attention mechanisms whose visualizations can be used to gain insights into inter-modality relationships. However, the clinical interpretation of these attention maps needs a great deal of expertise and such interpretation might not resonate with the diagnostic thought processes of medical professionals.

Various transform domain techniques have been tried for optimizing fusion results. The Non-Subsampled Contourlet Transform (NSCT) was introduced to improve spatial as well as functional feature retention with CT–MRI and MR T1/MR T2 fusion in some previous studies [8]. Another approach was to use two-scale image decomposition and sparse representation, retaining contrast sharpening and edge preservation [7]. Further, it was found that the Non-Subsampled Shearlet Transform (NSST), along with adaptive optimization techniques, provides improved computational efficiency with excellent fusion quality [9]. Taken together, these applications

improve spatial resolution and reduce redundancy simultaneously, making them useful in a variety of clinical applications. Due to the increasing complexity of fusion techniques, hybrid models combining various transform-based unsupervised methods have been developed to achieve high efficiency and mitigate the drawbacks of individual techniques.

2.2 Modality-Specific Information Preservation

Another outstanding flaw in the present body of fusion literature is that the specific features of modalities are not preserved throughout the fusion procedure sufficiently. Majority of traditional methods do not approach modalities as well as they look at each modality with the same level of sensitivity, whereas they can suppress any distinctive information peculiar to a certain imaging technique. As an example, the fine spatial information provided in an ultrasound image or the hotspots of the metabolic activity provided in a PET scan may be washed out when performing generic fusion with other modalities based on generic fusion rules.

High-quality preservation techniques have also been developed to alleviate this shortcoming e.g. feature extraction networks that are modality-aware and learn a specific representation of each imaging modality prior to fusion. The strategies utilize distinct encoder branches in different modalities, which helps them to maintain their essential properties, and is still able to permit each branch to be merged successfully at subsequent stages. Computational overhead and training complexity of such methods are, however, also major concerns.

In the era of deep learning, fusion models have gained a new form of transformation. According to deep learning-based fusion strategies, transformer-based architectures have emerged as a trend in multimodal imaging [2]. There are layer-level strategies that have been instrumental in addressing class imbalance and data scarcity problems in segmentation-specific fusion [3]. Deep learning networks have made it much easier, by automatically extracting hierarchical features, to produce more accurate fused images that are less prone to error in real-world settings. Deep learning models have successfully contributed to the fusion of medical images by utilizing large-scale datasets and high-performance computing resources to deliver higher resolution and better contrast, which are needed for detailed anatomical and functional analysis. Promising results have been demonstrated through hybrid approaches that combine traditional transform methods with deep learning. Some works have suggested hybrid NSCT, or dual-tree complex wavelet transform fusion models that enhance symmetry across fused MRI, CT, and PET datasets [6]. Others have used Multiscale Adaptive Transformer (MATR) networks based on adaptive convolution and attention to further improve global context representation [12]. Real-time processing with high precision has been achieved through the combined application of these methodologies for brain tumor detection, cardiac imaging, and neurodegenerative disease assessment. Hybrid models combining traditional mathematical frameworks and the latest deep learning algorithms provide greater adaptability to various imaging conditions, enhancing the robustness and reliability of fused outputs.

2.3 Cross-Dataset Generalization Issues

Generalisation of the multiple diverse datasets is also one of basic challenges that was not properly tackled in earlier fusion surveys. The most commonly used approaches perform well with respect to particular data sets but do not transfer successfully to images at another institution, imaging protocol, or patient population. The shortcoming of this is attributed to various changes such as the domain shift, differences in imaging parameters and anatomical differences of the population.

Domain adaptation methods were also tried, do better on the cross-dataset generalization, such like adversarial training methods which can learn domain-invariant feature representations, also known as meta-learning methods which can perform quick adaptation to new datasets. Nonetheless, these techniques tend to have a heavy computational cost and it is common that they do not scale well onto the underlying domain.

Recent advances in deep fusion networks have focused on adding attention mechanisms to select features more specifically. A deep learning framework based on a multi-branch and multiscale architecture, integrating unsupervised image segmentation, has been proposed to facilitate superior semantic feature extraction and structural retention [18]. Fusion models with Shearlet Transform and Enhanced Monarch Butterfly Optimization (EMBO), further classified using the Discrete Gravitational Search Algorithm (DGSA), have achieved higher accuracy than conventional CNN and SVM-based classifiers [14]. By increasing reliance on attention-based mechanisms, only specific features are enhanced, while the information extracted and

interpreted is also improved. Attention-driven fusion models tackle this problem by employing attention mechanisms to perform noise and artifact reduction in medical images, resulting in more precise diagnostic insights. As telehealth and cloud-based medical imaging gain popularity, security in fused medical data has become a new concern. Another study presented WatMIF, a fusion watermarking technique combining NSCT, RDWT, RSVD, and a Denoising Convolutional Neural Network (DnCNN), to improve encryption robustness and identity verification in cloud storage [17]. Such expansion of cybersecurity in medical imaging is expected to continue, with data integrity and patient privacy protected while still allowing access for remote diagnostic applications. Security in image fusion techniques not only ensures the privacy of patient data but also facilitates interoperability between different medical institutions and allows seamless integration of images across varied healthcare systems.

2.4 Post-Fusion Refinement Strategies

The method of post-fusion refinement is minimal in the current fusion literature and most of these methods only seek to find out how the fusion process occurs itself. Nevertheless, post-processing of fused images can be utilized to improve their quality significantly using advanced methods to de-noise them, enhance the contrast and get back lost small details, which could have been degenerated in the process of fusing.

As the further development, the three artifact suppression algorithms are proposed to detect and remove the distortions caused by fusion, contrast enhancement algorithms that are optimized pertinent to multimodal contents, and detail restoration algorithms to restore the high frequency information of each modality. Refinement in the form of machine learning has also proven to be quite promising, with trained networks being used to optimize the set parameters of post-processing in each type of fusion scenario.

Toward this end, multimodal image fusion will gradually gravitate toward adaptive, context-aware networks with better semantic integration. Transformer-based models and attention mechanisms have enabled real-time medical imaging solutions. Nevertheless, challenges such as data scarcity, high computational requirements, and cross-domain generalization persist. In the future, research should focus on self-supervised learning methods, optimal loss functions, and efficient models for multimodal fusion. Other keys to building trust include the development of interpretable AI models that will enable the use of fused images in clinical diagnostics. However, more research is required to implement new feature extraction techniques that minimize computational cost while maximizing accuracy to suit various clinical environments using fusion methods.

With deep learning-dependent integration, medical imaging enters a significant phase in the transition from transform-based fusion. A major step toward improvements in medical diagnostics and clinical decision-making is achieved by incorporating attention mechanisms, improved feature selection strategies, and enhanced security. The efficiency and accuracy of medical imaging will be dramatically enhanced by multifusion frameworks as deep learning evolves [18]. Today, the synergy between artificial intelligence, medical physics, and biomedical engineering continues to provide the means to further improve healthcare. Fused imaging is an integral tool for modern medicine. The future of medical imaging will be shaped by ongoing research and technological advancements in this field and will lead to more accurate assessments and better planning of new therapies, with a high probability of improved patient outcomes.

3 Advanced Methodological Frameworks

3.1 Interpretable-enhanced Fusion Architectures

To mitigate the problem of interpretability that was outlined in the current fusion strategies, interest in recent studies has included the development of transparent fusion architectures that offer clinically interpretable explanations to their decisions. RI methods Unexplainable AI (XAI) approaches dedicated to multimodal medical image fusion have already welded into the techniques of attention visualization, gradient-based explanation, and feature attribution that contribute to targets being understood by clinicians on the way modalities are involved in the result fused.

Fusion networks have been given a layer-wise relevance propagation (LRP) adaptation; this allows the decomposition of fusion decisions into modality-specific contributors. Such a strategy enables the medical practitioners to follow the way that certain anatomical structures or functional patterns affect the final fused image which heightens the sense of trust and medical acceptance.

3.2 Detection of Modality-Aware Preservation Techniques

Adapting modality based weighting devices have proved to be an efficient way of maintaining information based specific to modalities, when fusing. The techniques have an adaptive manner of varying the input of each modality depending on picture characteristics at the location of interest in the image; using them, fused output is guaranteed to retain special diagnostic pulse. Separation of imaging modalities before having to combine them is performed via multi-branch architectures with modality-specific feature extractors.

Measures defined over information, such as mutual information and entropy-based measures, are gaining attention as specification of what to maintain in information fusion to best preserve the modality-specific information present in data. These methods make sure that the process of fusion maximizes information retention at the cost of decreasing redundancy among modalities.

3.3 Generalization Enhancing Strategies

Meta-learning solutions to fusion proved to be of much value in solving the problem of the cross-dataset generalization. Such approaches can allow quickly adapting fusion models to use new imaging paradigms or learn new patient cohorts with little extra training data. Medical image fusion using few-shot learning specifically trained method enables the models to learn well in a variety of clinical settings.

Adversarial training and the option of contrastive learning have been pursued in domain-invariant feature learning, the objective of which is to design fusion models, whose performance would be stable across the datasets and a variety of imaging conditions. Such methods acquire robust representations which are less vulnerable to diversity in domain differences.

3.4 Post-Fusion Refinement First-Generation Bio

To overcome some limitations of single-pass fusion technology, multi-stage refinement networks have been developed. These architectures have individual refinement modules that emphasize certain features in image quality, such as artifact suppression, contrast enhancement and detail restoration. Generative adversarial networks (GANs) have also been useful in post-fusion refinement (where the challenge is to learn to produce high-quality fused images that preserve clinical fidelity).

Quality-conscious refinement algorithms employ image quality measurement parameters in making refinement processes in such a way that ensure end products are of clinical grade in terms of their quality to be used in diagnosis. These methods use reference-based and no-reference quality measures in combination to maximize performance of fusion.

4 Clinical Impact, Barriers, and Strategic Directions for Multimodal Medical Image Fusion

4.1 Clinical Sentence and Practice Implication

The theoretical benefits of advanced fusion technologies reach much further than their clinical implications. PET-CT fusion has brought revolutionary changes in the diagnosis, staging and treatment monitoring of cancer in oncology, since it allows earlier detection and hence better patient outcomes. There has been higher application of fMRI-structural MRI fusion in neurological applications, where risk assessment of any surgical operation has been easier to study and work with. Incorporation of CFP-OCT in ophthalmological diagnostics has greatly increased the precision of this diagnostic methodology, which facilitated earlier diagnostics of vision- threatening conditions. Multimodal fusion combinations are applications that can show that it is not only a technological innovation, but also a clinical utility that can best serve patients.

4.2 Barriers and Actual Challenges

While technological advancements are of great benefit, our discussion has led us to persistent challenges that are barriers to clinical adoption. Another major challenge is computational complexity, as most state-of-art techniques are characterized by excessive hardware and long latencies that cannot fit into clinical processes. There are no standardized metrics of evaluation and benchmark datasets that allow making an objective comparison of fusion strategies and a systematic approach to quality improvement. The gap between research validation with data curated by the researchers/company and actual clinical performance is large and therefore further strict clinical validation standards are required.

The limitations of interpretability have remained a significant obstacle to the adoption of these approaches in clinical practice because usually most deep learning-based fusion techniques work on a black box and do not output justifications that would be of clinical interest. Practical implementing of

cross-dataset in other clinical domains and with other populations is constrained by cross-dataset generalization issues.

4.3 Future Research Directions and Strategic Recommendations

These are the core weaknesses that the future of multimodal medical image fusion must seek to mitigate by adopting strategic research directions. Creation of new, lightweight, efficient architectures that can allow deployment in a real time clinical setting is a major priority. Explainable AI with clinically interpretable explanations of fusion decisions will be necessary to both establish credibility with clinicians and pass regulatory scrutiny. Operations standardized benchmarks and holistic clinical validation processes to support confidence measures of performance and support deriving regulatory pathways would have to be developed.

New developments in technologies like federated learning and edge have shown potential in the direction of distributed fusion and still proving privacy to patients. More-efficient post-fusion refinement techniques based on generating adversarial networks and quality-conscious optimization are significant future directions of research.

4.4 Concluding Observations

Multimodal medical image fusion is currently in a binary position between technological sophistication and clinical practicality and research at this stage should focus on procuring a good balance of both. Clinical translation potential is well borne out by the astonishing increases in fusion accuracy and ability to represent features that have been brought about by recent advances in deep learning.

The systematic investigation of the different frameworks of interpretability, as well as the modality-specific preservation, cross-dataset generalization measures, and high-level post-fusion refinement strategies in this survey helps to fill critically missing gaps in the literature discussing fusion and offers a prospective research agenda. Given the ongoing growth in artificial intelligence, and the higher level at which it is currently exercised in clinical validation, the democratization of fusion techniques is highly likely to revolutionize medical imaging and position multimodal fusion as the standard of precision medicine. The final degree of success of multimodal medical image fusion will not be gauged solely by the algorithmic complexity but by quantifiable contributions in truth as regards the outcome of diagnosis, treatment planning and patient outcomes in the real clinical scenario.

5 Conclusion

This exhaustive review has systematically laid out the fast developing field of multi-modal medical image fusion with the results indicating a field that is technologically very advanced and has tremendous clinical prospects. This analysis shows that combining complementary information in the various scans through different imaging modalities such as CT, MRI, PET, and SPECT is a paradigm shift to more accurate, comprehensive, and clinically actionable imaging. The key contribution of the survey presented is the rational step-by-step identification, and evaluation, of a set of major limitations that have been poorly resolved in prior fusion literature. Compared to the current surveys whose main focus is on algorithmic performance, the work is covering presumptively the topic of interpretability issues, modality-specific preservations questions, cross-dataset generalization issues, and the failure of post-fusion refinements. As we show, hybrid solutions that combine classic mathematical models with deep learning strategies perform at least as well as an implementation of a single strategy, and typically they better preserve anatomical structure as well as functional data. Combining transformer structures and attention mechanisms has solved many key problems of selective feature augmentation and global context modeling and generated fused images that are semantically more unified. Recent solutions to the security concerns regarding protecting patient data within more interconnected healthcare systems are security-minded fusion technologies such as watermarking, encryption and other information security methods.

Data Availability Statement

Not applicable.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Hermessi, H., Mourali, O., & Zagrouba, E. (2021). Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Processing, 183*, 108036. [Crossref]

[2] Li, Y., Daho, M. E. H., Conze, P. H., Zeghlache, R., Le Boité, H., Tadayoni, R., ... & Quellec, G. (2024). A review of deep learning-based information fusion techniques for multimodal medical image classification. *Computers in Biology and Medicine, 177*, 108635. [Crossref]

[3] Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. *Array, 3*, 100004. [Crossref]

[4] Huang, B., Yang, F., Yin, M., Mo, X., & Zhong, C. (2020). A review of multimodal medical image fusion techniques. *Computational and mathematical methods in medicine, 2020*(1), 8279342. [Crossref]

[5] Tirupal, T., Mohan, B. C., & Kumar, S. S. (2021). Multimodal medical image fusion techniques–a review. *Current Signal Transduction Therapy, 16*(2), 142-163. [Crossref]

[6] Alseelawi, N., Hazim, H. T., & Salim ALRikabi, H. T. (2022). A novel method of multimodal medical image fusion based on hybrid approach of NSCT and DTCWT. *International Journal of Online & Biomedical Engineering, 18*(3). [Crossref]

[7] Maqsood, S., & Javed, U. (2020). Multi-modal medical image fusion based on two-scale image decomposition and sparse representation. *Biomedical Signal Processing and Control, 57*, 101810. [Crossref]

[8] Bhatnagar, G., Wu, Q. J., & Liu, Z. (2015). A new contrast based multimodal medical image fusion framework. *Neurocomputing, 157*, 143-152. [Crossref]

[9] Jose, J., Gautam, N., Tiwari, M., Tiwari, T., Suresh, A., Sundararaj, V., & MR, R. (2021). An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion. *Biomedical Signal Processing and Control, 66*, 102480. [Crossref]

[10] Zhang, Y., Sidibé, D., Morel, O., & Mériaudeau, F. (2021). Deep multimodal fusion for semantic image segmentation: A survey. *Image and Vision Computing, 105*, 104042. [Crossref]

[11] Liu, Z., Yin, H., Chai, Y., & Yang, S. X. (2014). A novel approach for multimodal medical image fusion. *Expert systems with applications, 41*(16), 7425-7435. [Crossref]

[12] Tang, W., He, F., Liu, Y., & Duan, Y. (2022). MATR: Multimodal medical image fusion via multiscale adaptive transformer. *IEEE Transactions on Image Processing, 31*, 5134-5149. [Crossref]

[13] Bhavana, V., & Krishnappa, H. K. (2015). Multi-modality medical image fusion using discrete wavelet transform. *Procedia Computer Science, 70*, 625-631. [Crossref]

[14] Parvathy, V. S., Pothiraj, S., & Sampson, J. (2020). Optimal deep neural network model based multimodality fused medical image classification. *Physical Communication, 41*, 101119. [Crossref]

[15] He, C., Liu, Q., Li, H., & Wang, H. (2010). Multimodal medical image fusion based on IHS and PCA. *Procedia Engineering, 7*, 280-285. [Crossref]

[16] Tan, W., Tiwari, P., Pandey, H. M., Moreira, C., & Jaiswal, A. K. (2020). Multimodal medical image fusion algorithm in the era of big data. *Neural computing and applications*, 1-21. [Crossref]

[17] Singh, K. N., Singh, O. P., Singh, A. K., & Agrawal, A. K. (2024). Watmif: Multimodal medical image fusion-based watermarking for telehealth applications. *Cognitive Computation, 16*(4), 1947-1963. [Crossref]

[18] Lin, C., Chen, Y., Feng, S., & Huang, M. (2024). A multibranch and multiscale neural network based on semantic perception for multimodal medical image fusion. *Scientific Reports, 14*(1), 17609. [Crossref]

Vandit Akhilesh Barola is a Computer Science student (Batch 2022–2026) at Bennett University, specializing in AI and cybersecurity. With multiple publications in IEEE and Scopus-indexed journals, he focuses on GAN-based medical image fusion, ethical hacking, and secure system design. Vandit has hands-on experience in building intelligent, privacy-aware applications, along with full-stack web development using modern technologies. (Email: technovandit18@gmail.com)

Dr. Prabhishek Singh, Associate Professor at Bennett University and Senior IEEE & ACM Member, is among the World's Top 2% Scientists. With a Ph.D. in Computer Science and over 10 years of experience, he specializes in image processing and AI. He has authored 220+ papers and actively contributes as editor and reviewer across reputed journals. (Email: prabhisheksingh1988@gmail.com)

Prof. (Dr.) Manoj Diwakar is a Professor in the Department of Computer Science and Engineering at Graphic Era University, Dehradun, with over 20 years of academic and industrial experience. He holds a Ph.D. in Computer Science and is a professional member of IEEE and a life member of IAENG. His research focuses on image processing, medical imaging, and information security. Dr. Diwakar has published over 250 research papers in reputed journals, conferences, and book chapters.. (Email: manoj.diwakar@gmail.com)