Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 897-934 e-ISSN: 2656-8632

RESEARCH ARTICLE
OPEN ACCESS

(Correction)

Intelligent Fusion of Multi-Modal Medical Imaging: A Comprehensive Review of Methods, Challenges, and Clinical Integration

Majda Maatallah[1], Abdelmadjid Benmachiche[1], Khadija Rais[2], and Selma Touam[3]

[1] Laboratory of Computer Science and Applied Mathematics, Dept. of Computer Science, Faculty of Science and Technology, Chadli Bendjedid, University, El-Tarf, Algeria.
[2] Laboratory of mathematics, informatics and systems (LAMIS), Echahid Cheikh Larbi Tebessi University, Tebessa, Algeria.
[3] Laboratory of Physical-Chemistry of Materials, Dept. of Physics, Faculty of Science and Technology, Chadli Bendjedid University, El Tarf, Algeria.

Corresponding author: Khadija Rais (e-mail: khadija.rais@univ-tebessa.dz), **Author(s) Email**: Majda Maatallah (e-mail: maatallah-majda@univ-eltarf.dz), Abdelmadjid Benmachiche (e-mail: benmachiche-abdelmadjid@univ-eltarf.dz), Selma Touam (e-mail: touam-selma@univ-eltarf.dz)

Abstract **Multimodal Medical Imaging Fusion (MMIF) is defined as the incorporation of information from multiple imaging modalities in a way that is mutually supplementary, thereby addressing limitations associated with using a single imaging modality to evaluate a patient and increasing diagnostic accuracy. Further, this review provides a dedicated synthesis of deep learning architectures in MMIF, examining CNN-based hybrids, attention-enhanced transformers, GAN-driven unsupervised fusion, and emerging diffusion models. The state of the art in MMIF can be classified into three levels of fusion: (1) pixel level, fusion of raw pixel intensity values to preserve spatial detail; (2) feature level, features are derived from textures, edges, and region-of-interest (ROI) descriptors; (3) decision level, fusing independent outputs of each source using ensemble or rule-based methods to produce a single, integrated output from all sources, potentially improving interpretability of the integrated output. The use of AI algorithms improves fusion outcomes by yielding higher-quality results. However, clinicians' confidence in deep-learning-based models is limited due to their inability to generalise across multiple scanners, protocols, and medical systems. This analysis demonstrates that clinical AI systems must be developed with interpretability as a core attribute, to provide an explanation of how each modality is contributing to the final decision, and to establish a fusion policy that preserves the ability to make accurate diagnostic determinations based on fused images. In addition to developing more sophisticated algorithms, future developments in MMIF will require collaborative partnerships between developers and clinicians to develop fused images into reliable diagnostic tools to be used in precision medicine.**

Keywords **Multimodal medical image fusion (MMIF); deep learning; feature-level fusion; pixel-level fusion; decision-level fusion; medical imaging**

I. Introduction

In recent years, the COVID-19 pandemic accelerated the adoption of e-commerce, e-learning and e-health platforms, highlighting the importance of digital technologies in ensuring continuity of services and education. In addition, the changes associated with COVID-19 have varied greatly depending on the subject, and the dynamic assessment of lung damage, inflammatory conditions, and vascular issues has been critical to understanding disease progression. However, as the use of medical imaging has increased over the years, it has become clear that no single modality can provide a complete understanding of a patient's condition. MRI offers detailed views of soft tissues; however, it provides limited functional information. On the other hand, CT provides detailed anatomical information, but it is associated with exposure to radiation. PET/SPECT scans reveal areas of high metabolic activity, but they do not provide much structural detail [1].

Multimodal Medical Image Fusion (MMIF) has been developed to incorporate additional information from

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

different imaging modalities into a single composite image that offers more clinical information than either original image alone. The primary focus of MMIF is to develop imaging approaches that can create fused images to help physicians make more accurate diagnoses, better plan treatments, and establish a stronger basis for clinically sound decision-making. This type of combined imaging data is particularly important when studying diseases with distinct imaging characteristics, such as in oncology, where PET/CT fusion assists in accurately identifying and staging tumors, or in neurology, where MRI/PET fusion helps detect early changes in neurodegenerative disorders, as indicated by the correlation between amyloid deposition and structural atrophy [2].

In the past several years, the way we scale manufacturing processes from handcrafted approaches to data-driven methods has been significantly transformed, shifting from traditional techniques such as IHS and wavelet transforms to more advanced approaches based on deep learning, including convolutional neural networks (CNNs), generative adversarial networks (GANs), and vision transformers. These deep learning models significantly enhance intelligent systems by enabling automatic feature extraction, learning complex nonlinear patterns, and generalizing across diverse data. Unlike traditional methods, they improve accuracy, adaptability, and scalability, making them essential for advanced decision-making and automation. They enable the development of end-to-end learning systems for complex fusion maps that provide highly accurate texture details and a combined view of two or more images [3]. These developments have been greatly supported by the growing interest in precision medicine, where fused images are used not only to assist in diagnosis but also to provide data for CAD systems that support physicians in planning patient treatment [1], [2]. New types of taxonomies today encompass a wide range of methodologies, and there are now five distinct architectural paradigms: encoder-decoder, attention-based models, graph neural networks, generative models, and constraint-based approaches, which provide a more nuanced view than previous categorizations (early vs. late fusion) [4], [5].

While there has been some progress toward commercializing the MMIF method, there are still critical barriers to translating MMIF methods into clinical use. Barola et al. [6] point out that the majority of deep fusion models are treated as “black boxes,” resulting in a lack of trust from the clinical community because there is little or no explanation of how decisions are made. Mirzaei et al. [7] identified a persistent problem of loss of modality-specific information. Specifically, during fusion, important diagnostic features such as PET hotspots or MRI lesion boundaries are often lost, which may lead to false-negative diagnoses in the fused outputs. Saleh et al. [8] warn that most methods are validated on small, curated datasets such as the Whole Brain Atlas, and do not perform consistently in real-world clinical environments; imaging protocols are not uniform, scanners vary significantly, and patient populations are also heterogeneous. Haribabu et al. [9] identify significant computational complexity and the lack of standardized benchmarks as two major barriers to clinical adoption, noting that current state-of-the-art models have image processing times of 25+ seconds, which is too slow for many time-sensitive clinical applications. According to Diwakar et al. [10], post-fusion refinement is seldom performed; therefore, most raw fusion outputs do not satisfy clinical requirements due to the presence of artifacts or contrast inconsistencies that diminish diagnostic value, yet little research has been done on post-processing steps such as denoising, edge enhancement, and quality improvement.

Large, multimodal, and clinically annotated datasets are difficult to obtain due to several factors, including stringent privacy regulations (e.g., GDPR and HIPAA), as well as limited availability of data for building robust, generalizable models. Several publicly available data repositories, such as the “Cancer Imaging Archive” (TCIA), the “Open Access Series of Imaging Studies” (OASIS), the “Alzheimer’s Disease Neuroimaging Initiative” (ADNI), and the “Whole Brain Atlas” (AANLIB), provide important benchmarks for evaluation; however, standardized metrics across these repositories (e.g., SSIM, MI, Entropy, SCD) have not yet been established [3]. The “missing modality” issue, which refers to cases where one or more imaging sources are unavailable, remains understudied despite being a frequent occurrence in routine clinical practice [4].

Emerging research directions present exciting possibilities for addressing these challenges. They can also help identify new ways to collect data across similar cohorts from multiple sites using data-driven techniques such as multimodal canonical correlation analysis (mCCA) and independent component analysis (ICA). These techniques help reduce model bias while identifying correspondences between information across multiple physiological processes based on structure, function, and/or diffusion using different imaging modalities (i.e., how these modalities provide complementary information about the same processes). In addition, with large-scale N-way data fusion developments (e.g., BigFLICA, SuperBigFLICA), it is possible to integrate very large datasets (greater than 47 imaging types), although harmonization across the many different ways data are

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

acquired and measured remains a major challenge. Semi-blind fusion methods with reference signals (e.g., cognitive scores, severity scores, polygenic risk) may enable the discovery of jointly multimodal biomarkers with increased clinical relevance; however, depending on their validation strategy, these approaches may still lead to overfitting on specific datasets [11]. Cross-modality neuroimage synthesis may also be a viable and complementary approach to semi-blind fusion, addressing modality incompleteness by learning mappings between imaging modalities; however, emerging generative methods will require extensive validation to ensure clinical reliability and to avoid synthetic artifacts that may influence diagnosis [12].

For oncology, integrating biomedical informatics into multimodal fusion approaches is an essential component of cancer patient management. Evidence derived from Electronic Health Records (EHR), molecular profiles, digital pathology, and radiographic images must be combined to provide complementary information for identifying biomarkers and stratifying patients for treatment. Several factors must be considered when selecting a multimodal fusion strategy at the early, intermediate, or late stages of the multi-source data integration process: early fusion maximizes feature interaction at the expense of heterogeneity, whereas late fusion permits modality independence and accommodates missing modalities, at the cost of potentially overlooking cross-modality synergy. Furthermore, the use of interpretability methods can increase clinician trust in AI-enhanced workflows [13]. Specific examples include Gradient-weighted Class Activation Mapping (Grad-CAM) for image tiles and Shapley Additive exPlanations (SHAP) for molecular features. However, several implementation challenges remain, including ensuring real-time processing under limited computational resources; developing lightweight fusion algorithms through feature optimization and model pruning; and ensuring robustness under variations in illumination and sensor noise [14]. A comparison of existing work shows that CNN-based methods typically achieve better quantitative results than other approaches for multimodal fusion tasks. Although CNN-based methods are more effective at achieving quantitatively accurate results, they generally exhibit lower fusion symmetry, and their performance generalizability across different scan protocols is significantly lower than that of other methods (with more than a 30% reduction in cross-institutional evaluations) [15].

Medical image fusion can be classified, and further categorized by acquisition method, into multi-view, multi-modal, multi-time, and multi-focus approaches, which provide imaging solutions for a variety of clinical applications, ranging from surgical navigation to continuous disease monitoring [16]. Additionally, a systematic taxonomy can classify fusion methods according to input modality complexity as follows: short-wavelength (bi-spectral), mid-wavelength (multi-spectral), and long-wavelength (hyperspectral), as well as by fusion level (pixel level for spatial consistency, feature level for semantic consistency, and decision level for flexible classification), with each offering unique benefits and challenges for clinical implementation [14]. Individual predictive models that incorporate both fused images and behavioural, social, environmental, and genetic factors can provide valuable benchmarks for improving psychiatric diagnosis accuracy and personalizing treatment; however, such models must be prospectively validated across multiple clinical sites [5], [11]. to effectively assess the combined utility of these components [14]. Ultimately, to make advances in MMIF routinely applicable in precision oncology, thorough independent validation and adherence to FAIR principles for scientific and clinical reproducibility are essential [13].

The latest developments in deep learning-based multimodal image fusion frameworks, along with the increasing availability of benchmark performance results, have highlighted a significant research gap: the persistent divide between algorithmic innovation and clinical application. Most of the available literature on multimodal image fusion primarily focuses on classifying different modalities within taxonomic frameworks and introducing various quantitative image quality metrics (e.g., SSIM, PSNR, mutual information); however, little attention has been given to preserving modality-specific diagnostic signatures during fusion, evaluating how well algorithms generalize across different scanner protocols and patient populations, or standardizing interpretability and uncertainty quantification methods to build trust in clinical settings. In addition, there remains considerable variability across evaluation frameworks in relating fusion quality to actual diagnostic accuracy, inter-observer variability, and real-world clinical integration. This translational gap is further exacerbated by poorly structured validation practices, inadequate handling of missing or spatially misaligned modalities, and the absence of clinical auditing processes for post-fusion refinement pipelines. To address these challenges, a systematic synthesis of evaluation measures that go beyond technical benchmarks is required, with a clear focus on clinical applicability, robustness, and readiness for translation into practice, which is the central aim of this review.

The purpose of this article is to provide a comprehensive report on MMIF study methods by fusion level, pixel, feature, and decision, along with an extensive comparative analysis of these studies. To this end, it highlights recent developments in MMIF and identifies major trends in hybrid and unsupervised

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

learning, as well as the trade-offs among performance, scalability, and clinical feasibility. Furthermore, the article synthesizes current authoritative surveys on these topics to identify emerging research directions, such as explainable AI, federated learning, and real-time fusion, thereby providing a research roadmap that outlines priority areas for future work, including interpretability, modality preservation, cross-dataset generalizability, and integration into real-world clinical settings.

The remainder of this article is organized as follows: Section II provides background on traditional methods, deep learning in MMIF, and the critical preprocessing steps required for effective fusion. IIISection III details a comprehensive analysis of MMIF techniques categorized by fusion level: pixel-level, feature-level, and decision-level, respectively, including comparative tables of state-of-the-art methods. Section IV presents the datasets and evaluation metrics commonly used in the field. Section V offers a critical discussion on key challenges, including modality-specific information preservation, interpretability, and artifact reduction. Finally, Section VI concludes the review and outlines future research directions for clinical integration.

II. Background

MMIF is an emerging technology in today's healthcare systems that combines complementary information from multiple imaging modalities to improve diagnostic accuracy, assist in developing treatment plans, and enhance clinical decision support [17]. Traditional medical image fusion methods can be grouped into two categories: spatial-domain methods and transform-domain methods [18]. Spatial domain methods have a drawback: they typically cause distortion and spatial artifacts in the images being analysed, as these methods operate on actual pixel values (e.g. PCA and HIS) [17]. Transform domain methods, on the other hand, decompose an image into frequency components prior to performing the image fusion operation, and this provides a higher degree of structure preservation; however, many parameters need to be adjusted to best suit the application through complicated trial-and-error procedures (e.g., DWT, NSCT, NSST) [19].

There are many limitations associated with traditional fusion methods despite their successful and widespread use. The spatial-domain (or direct pixel) approach is computationally efficient but does not account for multi-scale contextual relationships between pixels; it also introduces spectral distortion and spatial artifacts due to simplistic pixel-wise operations [18]. Transform-domain techniques, on the other hand, preserve structural information at different resolutions through multi-resolution decomposition, but they require extensive manual parameter tuning, including the selection of decomposition levels, filter banks, and fusion rules (e.g., max-selection or weighted averaging). These processes are often empirical, modality-specific, and not optimally designed for preserving important diagnostic information [20]. In addition, such methods typically treat feature extraction and fusion rule design as two distinct processes, whereas they could be jointly optimized, thereby limiting adaptability to different anatomical structures and pathological variations across modalities. Furthermore, traditional methods lack the capability to automatically learn discriminative features or to extract them in a way that promotes generalization to unseen data, which has driven the shift toward deep learning-based fusion frameworks. By using deep learning networks, particularly CNN, feature extraction and fusion strategies can be jointly optimized in an end-to-end manner, enabling adaptive preservation of both structural and functional image characteristics while reducing reliance on manual intervention and heuristic design.

A. Deep learning in MMIF

Deep learning has emerged as a transformative paradigm, offering powerful feature extraction, adaptive fusion strategies, and end-to-end learning. Early deep learning-based fusion methods leveraged the robust feature extraction capabilities of CNNs. CNNs are specialized deep learning architectures designed to process grid-like data, utilizing learnable filters to automatically detect spatial hierarchies and patterns within images. CNNs operate through stacked convolutional layers that apply learnable filters to extract spatial features at different levels of abstraction. Early layers capture low-level features such as edges and textures, while deeper layers encode high-level semantic representations.

In MMIF, CNNs are particularly effective for learning modality-specific representations and generating fused feature maps through operations such as concatenation, addition, or attention-based weighting. CNNs have become the backbone of many MMIF systems due to their exceptional ability to extract hierarchical features from medical images. The hybrid S2 optimal CNN proposed by researchers at Saudi Electronic University demonstrates how CNN architectures can effectively fuse multimodal medical images by decomposing them using a modified discrete wavelet transform (MDWT) and classifying them as malignant or benign. This approach achieved superior performance with standard deviation, average gradient, and fusion factor compared to traditional methods [20]. Kong et al. [21] proposed CELM, a Siamese network that combines a CNN with an Extreme Learning Machine (ELM) and employs a two-stage fusion process to integrate CT, MRI, and PET images. Building on this direction, Vanitha et al. [22]

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

further refined the NSST-PCNN paradigm by introducing spatial frequency as a motivation for adaptive parameter tuning in PA-PCNN (SF-PAPCNN). This enhancement improved the discrimination of salient high-frequency features, leading to better edge retention and fusion symmetry as validated by entropy, mutual information, and edge strength metrics. Goyal et al. [23] introduced a hybrid framework operating in the Non-Subsampled Contourlet Transform (NSCT) domain, where a Siamese CNN generated adaptive weight maps for salient features while a [fractional-order total generalized variation (FOTGV) model performed joint denoising, addressing the clinical need for reliable fusion even with degraded source data. Almasri and Alajlan [20] combined MDWT with a CNN optimized by a Hybrid Dynamic (HOD) algorithm; their framework not only fused images but also classified the output as malignant or benign. Similarly, Vanitha et al. [24] proposed an energy attribute-guided fusion framework in the NSST domain, where low-frequency coefficients were selected via an energy-based activity measure and high-frequency details were merged using a parameter-adaptive PCNN (PA-PCNN). This approach preserved diagnostically relevant structures while maintaining computational efficiency, demonstrating robustness across multiple multimodal datasets. Arora et al. [25] developed RegFusion, a two-tier framework where VGG-19 extracted deep features for registration using thin-plate spline interpolation, followed by Lifting Wavelet Transform (LWT) and Singular Value Decomposition (SVD) for fusion, significantly improving structural preservation and edge retention. Addressing frequency-specific feature preservation, Zuo et al. developed DMC-Fusion [26], a deep multi-cascade framework that employs Gaussian high-pass filtering and PSNR thresholding to separate high- and low-frequency components. Using parameter-adaptive PCNN for high-frequency fusion and l1-weighted averaging for low-frequency content, DMC-Fusion's classifier-guided synthesis ensures diagnostically relevant features are prioritized, validated through statistical significance testing on brain disease classification tasks. Similarly, [27] proposed a hybrid CNN-BiLSTM framework that combines a modified Tetrolet transform for multi-focus CT and MRI fusion, enabling improved feature extraction and tumor prediction while reducing computational complexity. More recently, hybrid strategies have further evolved by tightly integrating deep learning with traditional enhancement techniques. For instance, the RGF-DnCNN-GMM framework combines Rolling Guidance Filtering (RGF), deep learning-based denoising (DnCNN), and gradient-based adaptive fusion (GMM), effectively separating structural and detail components, suppressing noise, and enhancing edge clarity, thereby achieving superior visual quality and information preservation compared to both classical and CNN-based methods [28]. Complementing architectural innovations, meta-optimization techniques enhance the robustness of fusion. For instance, Parvathy et al. [29] integrated Bayesian optimization with shearlet-based deep networks to automatically tune fusion weights and classifier hyperparameters, achieving superior diagnostic accuracy in edge-deployed medical IoT scenarios. In computational intelligence, optimization denotes the systematic process of adjusting decision variables, such as fusion weights, hyperparameters, or structural configurations, to extremize a defined objective function (e.g., diagnostic accuracy, structural fidelity, or convergence efficiency) while satisfying problem-specific constraints. When analytical gradients are unavailable or the search landscape is multimodal and non-convex, metaheuristic strategies (e.g., evolutionary, swarm-based, or bio-inspired algorithms) provide robust mechanisms to balance exploration and exploitation, thereby mitigating premature convergence to local optima. Complementing architectural innovations, meta-optimization techniques enhance fusion robustness. For instance, Dinh [30] employed the Marine Predators Algorithm (MPA) to optimize base-layer fusion parameters within a three-scale decomposition framework, while a Kirsch compass-based local energy function preserved edge details in high-frequency components. This bio-inspired strategy achieved superior contrast and structural fidelity compared to fixed-rule fusion, demonstrating the potential of evolutionary algorithms in adaptive medical image fusion.

To overcome the limitation of capturing long-range dependencies, attention mechanisms and Transformers have been widely adopted. Transformers have recently been adapted for medical image fusion due to their superior global modeling capabilities through self-attention mechanisms. Transformers rely on a self-attention mechanism that computes pairwise relationships between all input features, allowing the model to capture long-range dependencies. Given feature embeddings, self-attention assigns weights to different regions based on their relevance, enabling global context modelling. This is particularly advantageous in multimodal fusion, where relationships between distant anatomical structures and complementary modalities must be preserved. Unlike CNNs, which are inherently local due to convolutional kernels, Transformers provide a global receptive field, improving the integration of heterogeneous information. The Swin Transformer architecture uses hierarchical shifted windows to capture long-range dependencies in multimodal images [31]. Yu et al. [32] introduced an unsupervised

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

hybrid architecture integrating CNN and Vision Transformer modules, coupled with a complementarity information fidelity loss to preserve modality-unique features. This work laid the groundwork for later dual-branch designs, such as DFENet proposed by Li et al. [33], a dual-branch architecture integrating CNN and Vision Transformer modules to capture both local textures and global intensity distributions, further enhanced by a Global Semantic Feature Aggregation Module (GSFAM). Building on this idea, MRSCFusion [34] combines a residual Swin Transformer for long-range contextual modeling with a multiscale CNN for local feature extraction within an unsupervised autoencoder framework. Building on efficient long-range modeling, Zou et al. [35] introduced FocalNetFuse, which replaces standard self-attention with Focal Modulation Networks (FMNs) to capture global semantics with reduced computational cost. Coupled with Invertible Neural Networks (INNs) for detail preservation and a mutual information-driven decomposition strategy, FocalNetFuse achieves dual-level (global + local) fusion that excels in both infrared-visible and medical imaging tasks, demonstrating superior entropy and structural metrics over prior Transformer-based methods. In parallel, Ding et al. [36] introduced FTransCNN, which integrates CNN and Transformer features through a fuzzy logic-based Choquet integral to reduce feature heterogeneity, particularly in ambiguous tissue regions. Another line of work focuses on designing efficient attention mechanisms. AMMNet [37] incorporates an ECA-based attention module that jointly models channel and spatial dependencies. Complementary to these approaches, Zhou et al. [38] proposed TIEF, leveraging multi-head and spatial attention for both intra- and inter-modality feature integration in glioma diagnosis. Extending this paradigm, a manifold structure modeling method [39] introduces a gated hybrid attention mechanism to dynamically balance modality contributions. From a different perspective, SAFusion [40] employs a mixture-of-experts architecture to adapt fusion strategies across different imaging scenarios such as CT-MRI and PET-MRI. To further enhance fusion performance, Kamara et al. [41] developed FAMAFuse, integrating spatial attention residual modules and multiscale Gaussian attention to capture both local and global contextual information. A related approach, DRIFA-Net [42] incorporates dual attention mechanisms along with Monte Carlo dropout to improve robustness and enable uncertainty estimation. More advanced designs continue to explore the synergy between attention and multimodal fusion for more reliable and adaptive medical image analysis.

GANs consist of two competing components: a generator that produces fused images and a discriminator that evaluates their realism compared to ground truth data. The two networks are trained in an adversarial manner, where the generator aims to minimize the discrepancy between generated and real images, while the discriminator aims to maximize it. This adversarial objective encourages the generation of high-fidelity images that preserve both structural and textural details, making GANs particularly suitable for unsupervised multimodal fusion tasks. FusionGAN was among the first to apply GANs to medical imaging, improving target saliency by over 30% compared to conventional methods [31]. GANs have enabled unsupervised, high-fidelity fusion. Safari et al. [43] proposed MedFusionGAN for CT-MRI fusion, achieving high spatial and structural quality. Complementary to these approaches, U-Patch GAN improves the preservation of high-frequency detail through self-supervised and perceptual learning [44]. Extending GAN-based fusion to multi-modality brain imaging, Anita et al. [45] proposed MIMO-TGAN, a triple-generator adversarial framework where specialized generators process MRI-CT and PET-MRI pairs before a fusion generator integrates outputs via four adaptive rules. With SCRAT denoising and MSR enhancement as preprocessing, MIMO-TGAN achieves 99.43% accuracy in brain abnormality detection while preserving high-frequency anatomical details, a notable advance for clinical decision support. *Building on this idea*, IG-GAN enhances cross-modal consistency via dual-stream interactive and hierarchical fusion mechanisms [46], while residual attention-based GANs further improve texture and spatial feature extraction [47]. Zhou et al. [48] introduced DIFusion, an invertible framework ensuring lossless detail preservation. *From a different perspective*, structured constraint-based GANs address deformable misalignment and improve structural consistency in multimodal fusion [49].

Building on this paradigm, diffusion-based models have emerged as a robust alternative, leveraging strong generative priors to enable stable, high-quality fusion. A related approach, EPDiff integrates diffusion with multimodal attention to enhance anomaly detection and reconstruction [50], while AD-Diff combines diffusion-generated modalities with multimodal learning to improve diagnostic prediction [51]. Beyond *these methods*, GANs are also used for data augmentation and feature-level fusion to boost downstream classification performance [52], and more advanced designs such as MAC-GAN incorporate textual clinical knowledge to guide multimodal fusion [53]. A major trend is the fusion of imaging with heterogeneous data types to improve predictive accuracy and clinical relevance:

a. Genomics and Pathology: Kang et al. [54] developed a Transformer-based model fusing multi-sequence MRI with whole slide image (WSI)

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

features (extracted via TF-IDF) to predict neoadjuvant therapy response in head and neck cancer. Akbari et al. [55] similarly used a Transformer to integrate MRI and WSI for breast cancer detection, leveraging a virtual patient approach for cross-patient generalization.

b. Tabular and Clinical Data: Duenias et al. [56] proposed HyperFusion, a hypernetwork where a secondary network generates weights for an image backbone conditioned on tabular attributes (e.g., demographics, biomarkers). Yu et al. [57] introduced AMF-MedIT, featuring a Mamba-based tabular encoder (FT-Mamba) and an Adaptive Modulation and Fusion module to balance modality contributions under data-scarce conditions. Zeng et al. [58] developed C2HFusion, a clinical context-driven hierarchical framework integrating multi-sequence MRI, structured clinical variables, and unstructured radiology reports via a Mixture-of-Clinical-Experts (MoCE) module for personalized prognosis in pancreatic cancer. Ye et al. [59] proposed NMD-FusionNet for liver cancer, combining enhanced filtering, multimodal fusion of multi-phase CT, and a dual-path segmentation network.

c. Textual and Semantic Guidance: Xiang et al. [60] created SMFusion, a semantic-preserving framework that uses a multimodal image-text dataset and a cross-attention alignment module to incorporate expert-level diagnostic descriptions, ensuring fused images are both visually superior and clinically interpretable. Dai et al. [61] presented PCL-MFP, which uses prompt-level contrastive learning to align visual features with semantic context from multi-modal large language models, creating context-rich multimodal prompts for diagnosis. Rao et al. [62] proposed CAMF-SkinNet, integrating visual features (MedSigLIP), dermatology-specific embeddings (Google Derm Foundation), and medical captions (MedGemma-4B) via hierarchical cross-attention for skin disease classification.

Real-world clinical data often suffer from misalignment, missing modalities, and noise. Guo et al. [63] proposed SMAFusion, which incorporates a spatial registration module using a spatial evaluator network (U-Net) to mitigate the impact of unregistered image pairs, alongside a local-global multi-scale adaptive fusion module. Addressing the critical impact of spatial misalignment, Azam et al. [64] proposed a Multi-Resolution Rigid Registration (MRR) pipeline coupled with DWT-PCAv fusion. By iteratively aligning CT and MRI brain images across resolution pyramids before wavelet-based coefficient fusion, their method reduces registration-induced artifacts and improves mutual information retention, demonstrating that preprocessing quality directly governs downstream fusion efficacy, a finding later reinforced by SMAFusion's spatial evaluator network [63]. Wu et al. [65] introduced OmniFuse, a general framework with a missing modality imputation module using cascade residual autoencoders, a dynamically weighted fusion mechanism, and a traceable laziness activation strategy to handle missing, imbalanced, or noisy multimodal data, and validated it on MIMIC-III/IV and a proprietary lung cancer dataset. Pathak et al. [66] developed HYXmer, which synergistically combines CNN with Swin Transformer and a PCA-based high-gradient feature selection block to reduce redundancy (~6%) while preserving diagnostically relevant information. DL-based fusion has been tailored to a wide range of clinical tasks:

a. Lung Cancer: Sangeetha et al. [67] presented an MFDNN that fuses CT, MRI, PET, genomic data, and clinical records, achieving 92.5% accuracy and outperforming established methods.

b. Alzheimer's Disease: Cheng et al. [68] proposed a fusion network with a Multi-Scale Context-Aware Network for MRI, a Metabolic Abnormality Focus Network for FDG-PET, and a Cross-Modal Feature Constraint Fusion Module, achieving 87.02% accuracy on ADNI/AIBL datasets.

c. Breast Cancer: Xu et al. [69] introduced MSFT-Net for breast tumor classification using ultrasound, superb microvascular imaging, and strain elastography videos, with Spatio-Temporal Decoupling Attention and Sparse Cross-Attention to handle modality heterogeneity and noise. Huang et al. [70] proposed UltraMamba for breast lesion segmentation in multimodal ultrasound (B-mode, shear wave velocity, shear wave time) using Cross-Modal Knowledge Interaction and Region-Aware Feature Excitation modules.

d. Knee Osteoarthritis: Li et al. [71] developed SymUnet-DynCFC, combining a symmetric U-Net for T1W and T2W MRI with a Dynamic Confidence Fuzzy Control decision-level fusion mechanism, achieving superior cartilage segmentation.

e. Retinal Diseases: Dong et al. [72] evaluated a ResNet-based multimodal framework combining macular OCT, optic disc OCT, and fundus photographs to predict best-corrected visual acuity across six retinal diseases, achieving an MAE of 2.865 on the test cohort.

f. Ultrasound Segmentation: Chen et al. [73] fused B-mode ultrasound with parametric images (Nakagami, entropy) derived from raw RF data using a channel-aware attention module, improving segmentation of low-contrast thyroid and breast lesions.

g. Rectal Cancer: Lu et al. [74] presented MCAB-GFEResNet, which integrates

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

multiparametric MRI, whole-slide biopsy images, and clinical biomarkers (CEA) with a Multimodal Clue Attention Bridge and Global Feature Enhancement module, achieving 97.21% accuracy for predicting neoadjuvant chemoradiotherapy response using only pre-treatment data.

h. Kidney Tumor Segmentation: Ravikumaran et al. [75] developed a deep CNN-RBM architecture that leverages CNN for spatial feature extraction and Restricted Boltzmann Machines for higher-order semantic modeling. Validated on the KiTS19 dataset, their method achieved superior segmentation accuracy over the standard 3D U-Net, highlighting the value of hybrid deep-probabilistic models for organ-specific pathology localization.

Several recent works aim for broad applicability across modalities and tasks. Lu et al. [76] proposed UMF-SegNet, featuring a Dual-stream Adapter for feature alignment and a Dual-stream Cross-selective Transformer (DCFormer) to filter irrelevant cross-modal information, and validated it on the BraTS, MSD Colon Cancer, and PST900 datasets. Wang et al. [77] introduced AdaSFFuse, a task-generalized framework with an Adaptive Approximate Wavelet Transform (AdaWAT) and Spatial-Frequency Mamba Blocks, demonstrating strong performance across infrared-visible, multi-focus, multi-exposure, and medical fusion tasks. Zhang et al. [78] proposed FDGNet, an end-to-end unsupervised network that models fusion as feature-weighted guided learning, using pair feature differences to generate interactive weights and a hybrid loss function to prevent luminance degradation. Liu et al. [40] developed SAFFusion, a saliency-aware frequency fusion network combining a Mamba-UNet with an embedded contourlet transform and a saliency-aware adaptive loss to prioritize diagnostically critical regions.

B. Preprocessing Steps in MMIF

Image preprocessing involves techniques such as filtering, enhancement, and reconstruction to improve quality and extract data from digital signals. It specifically employs operations like convolution filters, noise reduction, and redundancy removal to prepare images for further analysis. These steps are essential for reducing computational complexity and ensuring efficient subsequent processing [79]. Preprocessing constitutes the foundational stage of MMIf, ensuring that heterogeneous imaging modalities, such as CT, MRI, PET, and ultrasound, are standardized, enhanced, and spatially aligned before fusion execution. Based on comprehensive evidence from recent literature ([1], [15], [80], [81], [82]), the MMIF preprocessing pipeline systematically addresses four critical objectives:

a. noise suppression and artifact reduction, employing techniques such as convolutional autoencoders, bilateral filtering, wavelet thresholding, and median-mean hybrid filters to mitigate modality-specific artifacts like speckle noise in ultrasound or beam-hardening in CT.
b. intensity normalization and contrast enhancement, utilizing z-score normalization, min-max scaling, CLAHE (Contrast Limited Adaptive Histogram Equalization), and the RAVEL method to harmonize pixel intensity distributions across modalities while preserving diagnostically relevant tissue contrasts.
c. spatial resolution standardization and format conversion, involving bilinear or B-spline interpolation for resizing, DICOM-to-NIfTI conversion for neuroimaging compatibility, and consistent voxel spacing to enable pixel-wise correspondence.
d. precise image registration, which aligns anatomical structures across modalities through a hierarchical strategy, initial rigid or affine transformations for global coarse alignment followed by non-rigid B-spline deformable registration optimized via mutual information maximization or the Davidon-Fletcher-Powell (DFP) algorithm, with landmark-based initialization using anatomical fiducials (e.g., anterior/posterior commissures, corpus callosum) to ensure sub-voxel accuracy.

Critically, preprocessing must be modality-aware: MRI benefits from skull-stripping and bias-field correction (N4ITK), PET requires attenuation and scatter correction with SUV normalization, while CT demands Hounsfield Unit calibration.

The integration of hybrid preprocessing combinations, such as CLAHE + Butterworth filtering for frequency-domain smoothing or unsharp masking + bilateral filtering for edge-preserving denoising, has demonstrated superior diagnostic performance across diverse datasets, underscoring that no single technique universally suffices; rather, adaptive, task-specific preprocessing pipelines are essential for maximizing fusion fidelity and downstream clinical utility.

Table 1 serves as a practical reference framework for implementing preprocessing pipelines in MMIF. By systematically organizing eight critical preprocessing stages, from noise reduction and intensity normalization to registration validation and quality assessment, the table enables researchers and clinicians to select modality-appropriate techniques while maintaining awareness of methodological trade-offs. Each row delineates the specific purpose, recommended algorithms, modality-specific considerations, and supporting literature for every step, facilitating reproducible and clinically relevant fusion workflows. Importantly, the table emphasizes that

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

preprocessing is not a uniform process; rather, optimal configurations must be tailored to the imaging modalities involved (e.g., MRI-PET vs. CT-MRI), the anatomical region of interest, and the intended clinical application.

Preprocessing is not a generic "one-size-fits-all" stage; optimal pipelines must be tailored to the specific modality pair (e.g., MRI-PET vs. CT-MRI), clinical task (e.g., tumor segmentation vs. surgical planning), and computational constraints. Hybrid approaches, combining spatial-domain enhancement with transform-domain denoising, consistently outperform single-method strategies, achieving up to 87.5% effectiveness across diverse medical imaging datasets [80]. Furthermore, rigorous registration validation using mutual information and landmark error metrics is non-negotiable, as misalignment propagates irrecoverable errors into the fused output [1], [81].

III. MMIF Techniques

MMIF techniques can be organized into three principal levels of processing: pixel-level, feature-level, and decision-level fusion. This hierarchical structure reflects the progression from low-level data integration to high-level semantic interpretation [83].

A. Pixel-level fusion techniques

Pixel-level fusion combines raw image intensities, typically after alignment, to maximize similarity and preserve spatial details [83]. Pixel-Level fusion is the lowest level of image fusion, where the fusion process directly operates on the pixel intensities of the source images to generate a single fused image. This approach requires strict pixel-wise registration between input images and aims to preserve the maximum amount of original information, including fine details, edges, and textures, which are crucial for subsequent analysis and clinical diagnosis [18]. Unlike feature-level or decision-level fusion, pixel-level fusion combines data at the most fundamental level, making it particularly suitable for applications such as medical imaging where information loss must be minimized [18], [84]. The general principle can be expressed as a process where the fused image $u(x)$ is derived from two registered source images $u_{CT}(x)$ and $u_{MR}(x)$ using a fusion rule ϕ, as shown in (1) [85]:

$$u(x) = \phi(u_{CT}(x), u_{MR}(x)) \quad (1)$$

Table 1. Comprehensive Preprocessing Steps in MMIF

Step	Purpose	Key Techniques/Methods	Modality-Specific Considerations
1. Noise Reduction [1], [15], [80], [82]	Suppress acquisition artifacts and improve signal-to-noise ratio	• Convolutional autoencoder • Bilateral filter • Median-mean hybrid filter • Wavelet thresholding (bior1.3, soft threshold=0.5) • Anisotropic diffusion	• Ultrasound: speckle noise reduction • CT: beam-hardening artifact correction • MRI: Rician noise modeling
2. Intensity Normalization [1], [80], [82]	Standardize pixel value ranges across modalities for consistent fusion	• Z-score normalization • Min-max scaling [0,1] • RAVEL method (Removal of Artificial Voxel Effect by Linear Regression) • Histogram matching	• PET: SUV normalization • MRI: bias-field correction (N4ITK) • CT: Hounsfield Unit calibration
3. Contrast Enhancement [1], [80], [81]	Improve visibility of anatomical structures and pathological features	• CLAHE (clipLimit=2.0, tileGridSize=8×8) • Unsharp masking (α=1.5) • Histogram equalization • Rolling guidance filtering	• MRI: T1/T2/FLAIR standardization • Low-contrast modalities benefit most
4. Resolution and Format Standardization [15], [80], [82]	Ensure spatial compatibility and interoperability	• Bilinear/B-spline interpolation • Resizing to target dimensions (e.g., 240×240 or 380×380 for EfficientNet) • DICOM → NIfTI conversion • PNG/JPG export with quality=0.95	• Neuroimaging: NIfTI format preferred • Deep learning models require fixed input sizes
5. Region of Interest (ROI) Extraction [1], [81], [82]	Focus processing on diagnostically relevant anatomy	• Skull-stripping (brain imaging) • Masking of non-anatomical structures (bed frames, artifacts) • Automated segmentation pre-fusion	• Brain: AC-PC alignment, ventricle masking • Oncology: tumor boundary preservation

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

The weighted averaging method is proposed to achieve optimal pixel-level fusion, where the fused image is obtained as a linear combination of the source images. For instance, the fused image can be defined using an alpha image $\alpha(x)$ that determines the contribution of each source pixel, as indicated in (2) [86]:

$$u(x) = \alpha(x) \cdot u_{CT}(x) + (1 - \alpha(x)) \cdot u_{MR}(x), \forall \alpha(x) \in [0,1] \quad (2)$$

Another widely used formulation treats pixel-level fusion as an energy minimization problem, where the goal is to find the fused image that balances similarity to both source images while maintaining smoothness. The energy functional is presented in (3) [86].

Table 1. Comprehensive Preprocessing Steps in MMIF (continued).

Step	Purpose	Key Techniques/Methods	Modality-Specific Considerations
6. Image Registration [1], [81], [82]	Spatially align corresponding anatomical structures across modalities	Coarse (Global): • Rigid/affine transformation • Landmark-based initialization (AC, PC, corpus callosum) Fine (Local): • Non-rigid B-spline deformation • Mutual information maximization • DFP optimization algorithm • Deep learning: U-Net + Spatial Transformer Networks	• PET-MRI: large inter-modality gaps require robust similarity metrics • Dynamic organs: respiratory/cardiac gating may be needed
7. Post-Registration Interpolation [81], [82]	Resample transformed images without introducing artifacts	• Cubic B-spline interpolation kernel • Partial Volume (PV) / Novel PV (NPV) interpolation • Edge-preserving resampling	• Critical for pixel-level fusion accuracy • Avoids aliasing in high-frequency components
8. Quality Validation [1], [80], [82]	Verify preprocessing integrity before fusion	• Visual inspection by radiologists • Quantitative metrics: SSIM, PSNR, entropy, mutual information • Landmark registration error <1 voxel	• Essential for clinical deployment • Automated QC pipelines recommended for scalability

$$u(x) = \arg\min_{u \in \Omega} \left\{ \int_{\Omega} [w_1(x)(u(x) - u_{CT}(x))^2 + w_2(x)(u(x) - u_{MR}(x))^2]\, dx + 2\lambda \int_{\Omega} |\nabla u(x)|\, dx \right\} \quad (3)$$

where $w_1$and $w_2$are weight maps derived from image gradients, and the smoothness term $2\lambda \int_{\Omega} |\nabla u(x)|\, dx$ enforces natural transitions between pixels [86]. In transform-domain pixel-level fusion, such as with the DWT or the stationary wavelet transform (SWT), the fusion is performed on the decomposed coefficients. The fused image is obtained by applying an inverse transform to the combined coefficients, as expressed in (4) [85]:

$$I_{fused}(x,y) = \omega^{-1}(\phi(\omega(I_1(x,y)), \omega(I_2(x,y)))) \quad (4)$$

where ω represents the wavelet transform and $\phi$is the fusion rule applied to the transform coefficients [85].

For high- and low-frequency coefficient fusion in multiscale transform methods, different rules are applied. For low-frequency coefficients, an average rule is commonly used, as given in (5) [85]:

$$B_F(x,y) = \frac{B_1(x,y) + B_2(x,y)}{2} \quad (5)$$

where $B_F(x,y)$ represents the fused low-frequency coefficients, and $B_1$and $B_2$are the low-frequency coefficients of the source images. For high-frequency coefficients, region energy-based fusion is employed, as defined in (6) [85]:

$$E_k^l(x,y) = \sum_{m,n \in S} \left(C_k^l(x+m, y+n)\right)^2 \quad (6)$$

where C_k^l represents the high-frequency coefficients at the l-th decomposition level in the k-th direction, and S is the local window [85].

A novel multimodal fusion approach combines variable-order fractional-order calculus with multi-

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

objective Darwinian PSO to optimize fusion weights and convergence dynamics, employing a gradient compass to enhance spatial detail [87]. Complementing this optimization-driven strategy, Shehanaz et al. employ DWT with Particle Swarm Optimization (PSO) to determine optimum weighted average fusion coefficients, demonstrating robustness against Gaussian and speckle noise in MRI-CT/PET/SPECT fusion [88]. Similarly, Alzahrani proposes a Modified DWT framework optimized by the Arithmetic Optimization Algorithm (AOA) alongside bilateral filtering, achieving superior structural similarity and entropy by adaptively selecting fusion rule parameters [89]. The versatility of such bio-inspired optimizers, PSO, for example, allows for the treatment of large-dimensional nonlinear problems, which suggests its strong potential for adapting fusion weights in MMIF to improve robustness against modal heterogeneity and noise. Building on deep learning, U-Patch GAN introduces an end-to-end pixel-level framework for brain image fusion, leveraging a U-Net generator and PatchGAN discriminator to preserve functional and structural details through spectral normalization and advanced loss functions [44]. Similarly, KDE-GAN integrates knowledge distillation with explainable GANs for data-efficient CT-MR fusion, operating at both pixel and feature levels by optimizing perceptual loss and training dynamics [90]. MedFusionGAN extends this direction with an unsupervised pixel-level framework that fuses MRI and CT via adversarial learning and perceptual loss, generating high-resolution outputs [43]. Khan et al. combine CNN-based denoising with the Non-Subsampled Contourlet Transform (NSCT), employing Directional Total Variation and Linear Spectral Clustering for low-frequency components, and the Sum Modified Laplacian for high-frequency details, to minimize pseudo-Gibbs artifacts [91]. Ramesh and Kumar operate in the YUV color space, where the luminance (Y) component undergoes SVD decomposition into structural/detail components. Low frequencies fused via energy weighting; high frequencies enhanced through lightweight VGG19 feature modulation while preserving the original PET chrominance (U/V) channels for metabolic color fidelity. Achieves PSNR 33.94 dB with sub-second runtime [92]. Further enhancing detail preservation, a method based on anisotropic diffusion and cross-bilateral filtering adaptively weights significant pixel regions to suppress noise and preserve edges, decomposing images into base and detail layers for edge-aware fusion [93]. Wei et al. introduces explicit Pixel-level Structure-Aware (PSA) filter that leverages spatial semantic relationships and interval gradients to identify salient structures. Uses multi-scale patch decomposition (MSPD) for salient structure fusion and

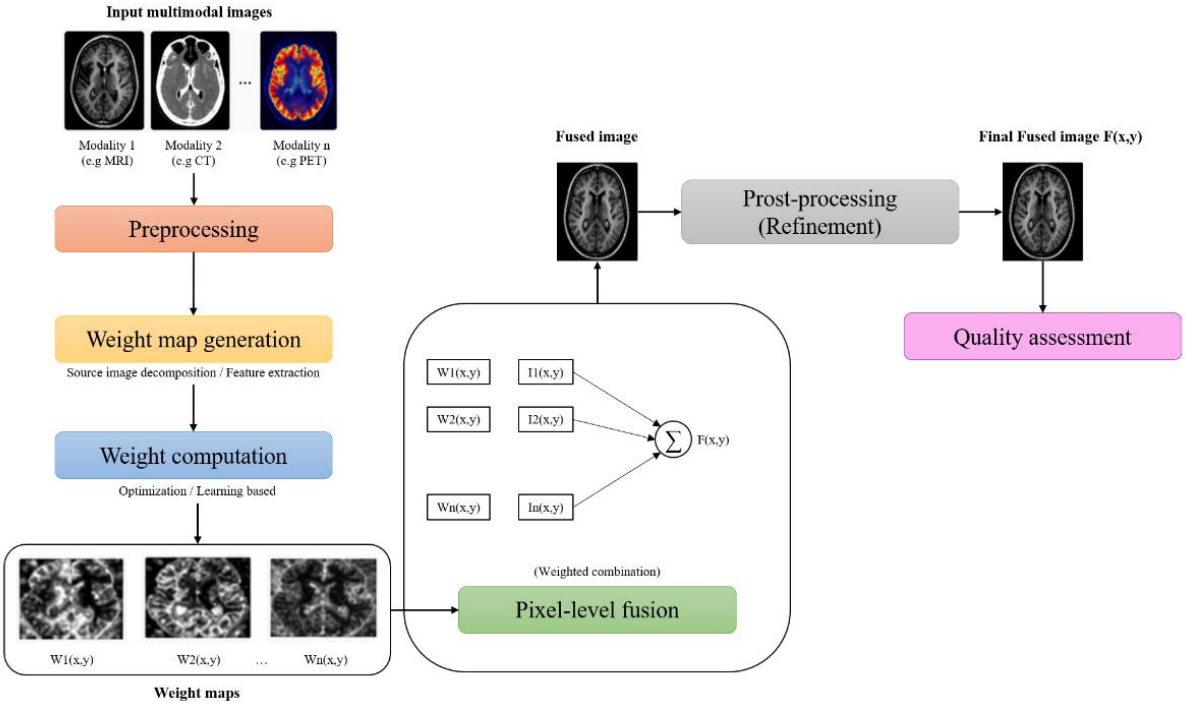

Fig. 1. Pixel-Level fusion framework.

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

weighted local energy (WNE) for non-salient regions. Preserves boundaries and contrast intensity, which are critical for skull base tumor diagnosis [94]. PPMF-Net advances pixel-level fusion through a hierarchical deep architecture that directly processes image intensities, incorporating dynamic edge enhancement, nonlinear feature interaction, and Transformer-based attention to retain structural and high-frequency details from PET and MRI [31]. Arora et al. propose a dual-stage framework combining VGG-19-based registration with LWT-SVD fusion. LWT decomposes images into approximation/detail coefficients; low frequencies are fused via averaging, high frequencies via SVD [25]. The primary advantages of pixel-level fusion include the preservation of rich, detailed information and the provision of accurate input for further image processing tasks. However, challenges include high computational requirements, sensitivity to registration inaccuracies, and potential information redundancy [18], [84].

Pixel-level fusion techniques demonstrate substantial progress in preserving both structural and functional information across multimodal medical images, with consistent improvements over traditional transform-based methods (Table 2). OWAF-PSO achieved strong quantitative performance, reaching SSIM values up to 0.9457, PSNR up to 37.82 dB, and MI up to 3.7865 across MRI-CT/PET/SPECT fusion tasks. It notably outperformed classical approaches such as DWT, FFT, and IHS, while maintaining robustness against Gaussian and speckle noise and preserving edge integrity for enhanced diagnostic clarity [88]. Similarly, the AOA-optimized MDWT framework proposed by Alzahrani et al. reported the highest entropy (6.785-7.918 bits/pixel) and SSIM scores (83.53-95.55%) among the compared methods. The integration of bilateral filtering with multi-resolution decomposition effectively suppresses noise while preserving anatomical boundaries, making it particularly suitable for treatment planning applications [89]. The CNN-enhanced NSCT framework with DTV-LSC and SML fusion introduced by Khan et al. further advances pixel-level fusion by achieving superior MI (4.64-4.97), SD (80.85-81.78), and $Q_{AB/F}$ (2.70-2.74) values across 181 multimodal image pairs. Its ability to minimize pseudo-Gibbs artifacts and preserve fine texture details results in top expert visual scores, supporting complex pathology assessment [91]. Optimization-driven approaches also demonstrate strong efficiency-performance trade-offs. VF-MODPSO-GC achieved state-of-the-art optimization metrics (IGD: 6.48×10^{-3}, HV: 1.81×10^{1}) alongside high image quality (Accuracy: 0.895), while reducing fusion time to just 0.085 seconds, over four times faster than competing methods, making it suitable for real-time intraoperative CT-MRI fusion on standard hardware [87]. Deep learning-based methods, particularly GAN architectures, consistently outperform traditional techniques across multiple metrics. U-Patch GAN achieved improvements of 27-53% across key metrics (e.g., $Q_G, Q_S, Q_{AB/F}, Q_{VIF}$) while effectively preserving both structural texture and functional color information across multiple modality pairs, including continuous-slice brain imaging [44]. KDE-GAN further demonstrates that high-quality fusion can be achieved even with limited datasets, outperforming nine state-of-the-art methods across five quantitative metrics and maintaining strong subjective visual quality [90]. MedFusionGAN extends these capabilities through an unsupervised framework, outperforming 15 competing methods on 6 of 9 evaluation metrics while preserving CT bone structures and MRI soft-tissue contrast without introducing artifacts. It achieves excellent tumor delineation performance (Dice: 0.96±0.020.96 \pm 0.020.96±0.02) with fast inference (~1.9 s per fusion), supporting time-sensitive applications such as radiotherapy planning and image-guided adaptive radiotherapy (IGART) [43]. Hybrid model designs further improve both perceptual quality and computational practicality. The SVD-VGG approach achieves PSNR values of 32.13-33.94 dB and SSIM of 0.927-0.935, while maintaining low perceptual error (LPIPS: 0.085-0.102). Its sub-second CPU inference (~0.56 s) and preservation of PET chrominance make it particularly suitable for resource-constrained clinical environments [92]. Lightweight fusion strategies such as AD+CBF also demonstrate strong performance, achieving the highest scores across multiple metrics (e.g., $Q^{x}_{pq/f}$=0.7233, API = 58.33, SD = 73.93, AG = 12.08) while producing high-contrast, artifact-free images that preserve both structural and metabolic information. Despite its simplicity, it remains competitive with more complex models, though further perceptual optimization is needed [93]. Transformer-based and high-capacity architectures continue to push performance boundaries. PPMF-Net achieves state-of-the-art results on PET-MRI fusion (e.g., SF: 38.27, SD: 96.55, MS-SSIM: 1.23), while demonstrating strong generalization to other modality pairs, enhancing lesion boundary clarity for clinical decision support [31]. Finally, clinically validated frameworks such as RegFusion (LWT-SVD) demonstrate the practical viability of fusion methods, achieving consistent improvements across RMSE, SSIM, PSNR, CC, and entropy metrics with relatively fast inference (1-4 s). Its validation on real-world clinical datasets highlights its applicability for brain tumor diagnosis and image-guided treatment planning [25]. Fig. 1 illustrates the complete pipeline of pixel-level fusion, which operates directly on image intensities. The process begins with:

a. Input multimodal images, such as MRI, CT, and PET, each providing complementary anatomical or functional information.

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

b. The images are then subjected to preprocessing, which includes noise reduction (e.g., filtering), intensity normalization (e.g., histogram matching), and spatial registration to ensure accurate pixel-wise alignment across modalities.

c. In the weight map generation, each input image is analyzed to estimate its local importance. This involves extracting spatial features such as gradients, edges, local variance, and entropy. These features are then used to compute spatially adaptive weight maps using optimization algorithms (e.g., PSO, AOA), deep learning models (e.g., CNNs, GANs), or hybrid strategies. The resulting weight maps determine the contribution of each modality at every pixel location.

Table 2. Pixel-Level Fusion Techniques.

Technique	Evaluation Metrics	Datasets	Imaging Protocols and Privacy	Computational Complexity	Real-World Impact
OWAF-PSO [88]	• MR-CT: SSIM 0.9457, PSNR 35.65 dB, MI 3.79, ($Q_{AB/F}$) 0.878 • MR-PET: SSIM 0.9218, PSNR 37.82 dB, MI 3.50 • MR-SPECT: SSIM 0.9154, RMSE 0.0261	• AANLIB: 20 MR-SPECT, 20 MR-PET, 18 MR-CT • 256×256	• Public, de-identified • Pre-registered images	• PSO (~50 iterations) + DWT • ~300 s per fusion • Faster than GA	• Improves brain pathology visualization • Noise-robust • Offline diagnostic use
MMIF-MDWTAOA [89]	• MRI-CT: Entropy 6.79-7.57, SSIM 83.5-94.1% • MRI-PET: Entropy 6.96-7.92, SSIM 84.8-95.6%	• Harvard dataset • MRI-CT, MRI-PET pairs • 256×256	• Public, anonymized • Pre-aligned, normalized	• AOA optimization + MDWT • Bilateral filtering preprocessing	• Enhances tumor boundary delineation • Improves contrast for treatment planning
DTV-LSC-NSCT [91]	• MI: 4.64-4.97 • SD: 80.85-81.78 • ($Q_{AB/F}$): 2.70-2.74 • SF: 28.73-29.81 • Visual score: highest	• 181 multimodal pairs • CT-MRI, PET-MRI, SPECT-MRI	• Public datasets •Registered, normalized •Patch augmentation	• CNN + NSCT decomposition • Iterative DTV optimization • Not real-time	• Superior edge/texture preservation • Reduces artifacts
VF-MODPSO-GC [87]	• IGD: 6.48×10^{-3} • HV: 18.1 • Accuracy: 0.895 • F1: 0.677 • High SSIM, MI, ($Q_{AB/F}$)	• Whole Brain Atlas • 14 CT-MRI pairs • 256×256	• Public, CC-BY licensed • Pre-registered	• ~0.085 s per fusion • Low complexity (non-adaptive fusion)	• Real-time neurosurgical navigation • Deployable on low-resource hardware
U-Patch GAN [44]	• (Q_G): +40.5% • (Q_S): +27.1% • ($Q_{AB/F}$): +53.5% • Top-2: VIF, MI	• 275 pairs (augmented to 4,650) • CT / MRI / PET / SPECT	• Public, de-identified • YUV conversion	• Dual discriminator GAN • RTX 2080Ti • Moderate training cost	• Continuous-slice fusion • Preserves structure + color • Needs real-time optimization
KDE-GAN [90]	• Top: SF, SSIM, ($Q_{AB/F}$), NMI, (Q_{Niec}) • Strong visual quality	• 400 CT-MR pairs • Whole Brain Atlas	• Public, de-identified	• Knowledge distillation reduces complexity • Explainable early stopping	• Works with limited data • Preserves diagnostic features

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

d. The framework then performs pixel-level fusion, where corresponding pixels from all modalities are combined using a weighted averaging rule, producing an initial fused image.
e. The post-processing/refinement enhances visual quality through contrast enhancement (e.g., CLAHE), edge-preserving smoothing (e.g., bilateral filtering), and artifact suppression.
f. The final output, shown in the final fused image, preserves both structural and functional details. The quality assessment evaluates the fusion performance using metrics such as SSIM, PSNR, mutual information, entropy, and visual fidelity, ensuring both quantitative and perceptual effectiveness.

Table 2. Pixel-Level Fusion Techniques (continued).

Technique	Evaluation Metrics	Datasets	Imaging Protocols and Privacy	Computational Complexity	Real-World Impact
MedFusionGAN [43]	• Top-1: ENT, SD, MG, SF, MI, ($Q_{AB/F}$) • Dice: 0.96 ± 0.02 • HD: 1.22 mm	• GLIS-RT (TCIA): 230 patients • 11,246 train / 2,276 test slices	• Public, CC-BY 4.0 • Augmented dataset	• 2×RTX 3090 • Training: 117 min • Inference: 1.9 s	• Radiotherapy planning (IGART, SRS) • Fast, end-to-end fusion
YUV based SVD-VGG [92]	• PSNR: 32.13-33.94 dB • SSIM: 0.927-0.935 • CC: 0.970-0.971 • LPIPS: 0.085-0.102 • Entropy ≈ 4.9	• 269 training + 188 testing pairs • MRI-PET	• Public, de-identified • Synthetic noise added	• SVD: (O(M^3)) • VGG19 forward pass • ~0.56 s (CPU)	• CPU-friendly deployment • Preserves PET metabolic info
AD+CBF [93]	• Edge transfer: 0.7233 • API: 58.33 • SD: 73.93 • AG: 12.08	• 4 multimodal pairs • CT-MRI, MRI-PET	• Public, de-identified	• ~5.86 s per fusion •Non-iterative method	• High contrast, low artifacts • Suitable for low-resource settings
PPMF-Net [31]	• SF: 38.27 • SD: 96.55 • SCD: 1.62 • MS-SSIM: 1.23	• 245 train / 24 test pairs • PET-MRI	• Public, CC-BY • Pre-registered	• Transformer + CNN • RTX 3090 Ti • High computational cost	• Improves lesion boundaries • Supports surgical planning
LWT-SVD [25]	• RMSE ↓: 0.2886 → 0.1054 • SSIM ↑: 0.7238 → 0.9312 • PSNR ↑: 58.9 → 68.8 dB • CC ↑: 0.928 → 0.997 • Entropy ↑: 5.84 → 10.93	• Public + clinical datasets • CT/MRI/PET	• Clinical data anonymized • IEC-approved • Informed consent obtained	• 1.03-3.93 s per image • Moderate-high complexity	• Clinically validated • Supports tumor diagnosis and radiotherapy

B. Feature-level fusion techniques

Feature-level fusion involves extracting and integrating modality-specific features, such as regions or edges, often using segmentation and statistical methods [83]. Feature-level fusion represents an intermediate level of image fusion, in which the fusion process operates not on raw pixel intensities but on attributes or "features" extracted from the source images. Instead of combining data at the most granular level, this approach first identifies salient characteristics, such as edges, textures, shapes, or deep semantic information, and then fuses these extracted representations to generate a composite feature set, from which the final fused image is reconstructed. This strategy reduces sensitivity to pixel-level misregistration and noise, as fusion decisions are guided by meaningful structural or functional properties rather than absolute intensity values [95], [96].

Unlike pixel-level fusion, which requires strict point-to-point correspondence, feature-level fusion allows for a more intelligent integration of complementary information by focusing on "objects and features within the scope" of each modality, making it particularly

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

advantageous for multimodal medical applications where source images exhibit different spatial resolutions or contrast mechanisms [95], [97]. The general principle can be expressed as a two-stage process: first, feature extraction operators $\mathcal{F}$ are applied to each source image to obtain feature representations; second, a fusion rule $\phi_{feature}$ combines these representations into a unified feature map. The fused image $u(x)$ is then reconstructed from this combined feature set, as shown in (7) [97], [98]:

$$u(x) = \mathcal{R}\left(\phi_{feature}(\mathcal{F}(I_{CT}(x)), \mathcal{F}(I_{MR}(x)))\right) \quad (7)$$

where $\mathcal{F}$ denotes the feature extraction operator (e.g., convolutional layers of a deep neural network), $\phi_{feature}$ is the feature-level fusion rule, and $\mathcal{R}$ is the reconstruction operator [97]. Several formulations have been proposed to achieve feature-level fusion in medical imaging. One common approach employs pre-trained deep convolutional neural networks (CNNs) as feature extractors. For instance, the VGG-16 network can be used to extract multi-layer deep features from the principal components of source images. These features are then fused using a weighted-average strategy, as expressed in (8) [97]:

$$F_{fused}^{l}(x,y) = \sum_{i=1}^{N} w_i \cdot F_i^l(x,y), w_i = \frac{\exp\left(A_i^l(x,y)\right)}{\sum_{j=1}^{N} \exp\left(A_j^l(x,y)\right)} \quad (8)$$

where F_i^l represents the deep feature maps extracted from the l-th layer of the CNN for the source image i, and w_i is the normalized weight computed from an activity level measurement A_i^l (e.g., block-based average) [97]. Another formulation leverages multi-level feature extraction within a cross-scale transformer architecture. In this approach, features are extracted from multiple residual blocks to capture both shallow (fine-grained) and deep (semantic) information. The fusion process is decoupled across these levels, as shown in (9) [98]:

$$F_{fused}^{level} = \text{CSTF}(F_A^{level}, F_B^{level}), \text{for } level \in \{shallow, middle, deep\} \quad (9)$$

where F_A^{level} and F_B^{level} are the multi-level features extracted from the Edge-Augmented Modules of each source image, and CSTF denotes the Cross-Scale Transformer Fusion Module that captures multi-scale contextual information [98]. In feature-level fusion, the extracted features can also be decomposed into complementary components. For example, low-rank representation (LRR) can separate source images into principal components (global structure) and salient components (local details). The deep features of the principal components are fused via weighted averaging, while the salient components are combined using a simple sum rule, as described in (10) and (11) [97]:

$$F_{principal}^{fused} = \text{WeightedAvg}(\text{CNN}(P_{CT}), \text{CNN}(P_{MR})) \quad (10)$$

$$S_{fused} = S_{CT} + S_{MR} \quad (11)$$

where P denotes the principal (low-rank) component and S denotes the salient component [97]. A key distinction of feature-level fusion is the use of "non-end-to-end" architectures, where the deep learning network is applied only during the feature processing stage prior to fusion. The process follows a sequential pipeline: feature extraction → feature fusion → reconstruction, as formalized in (12) [96]:

$$I_{fused} = \mathcal{R}\left(\phi_{fusion}(\Phi_{DL}(I_1), \Phi_{DL}(I_2))\right) \quad (12)$$

where Φ_{DL} represents a deep learning network (e.g., a CNN or Transformer) used solely for feature extraction, distinct from end-to-end methods where the network learns the entire mapping from source images to the fused output [96].

At the feature level, a region-based fusion strategy employs superpixel segmentation (LSC) and feature importance assessment (LGF, SML), refining the decision map via genetic algorithm optimization to preserve texture and structure while adaptively tuning fusion weights [99]. The MATR method leverages a multiscale adaptive Transformer architecture that integrates adaptive convolutions with self-attention to capture local textures and global semantics, improving contextual representation and fusion quality [100]. Another approach utilizes cross-modal learning, attention mechanisms, and entropy-guided feature selection to enhance glioma-relevant information from MMRI and SPECT/CT, enabling dynamic learning of discriminative representations and spatially-aware fusion for improved segmentation [38]. A dual-branch network with dilated convolutions and multi-scale decomposition captures both fine textures and global structures, and is augmented with attention mechanisms to boost fusion performance across standard metrics [36]. Prathipa and Ramadevi employ PCA for dimensionality reduction, Grey Wolf Optimization for weight optimization, and RNN for unsupervised feature categorization. Uses eigenvalue/eigenvector correlation analysis for fusion decisions [101]. FDGNet introduces difference-guided learning in an unsupervised framework, using a hybrid loss function to adaptively preserve luminance and texture, effectively mitigating degradation issues common in traditional multimodal image fusion [78]. AMMNet presents an end-to-end architecture combining multi-scale CNNs, DenseNet, and MobileNetV3 with attention modules, adaptively preserving contours and textures through decomposition and attention-guided selection [37]. To handle complex edge and texture information from CT

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

and MR, a method integrates Gabor-based representations with multiple G-CNNs and a fuzzy neural network, managing ambiguity in feature integration while enhancing visual quality [102]. Another approach combines transform-based decomposition (FDCuT) with fuzzy entropy-based region fusion, optimized via a multi-objective meta-heuristic (A-EFO) to improve structural and perceptual fidelity across frequency sub-bands [103]. DRIFA-Net further strengthens feature integration with dual attention modules, promoting robust generalization

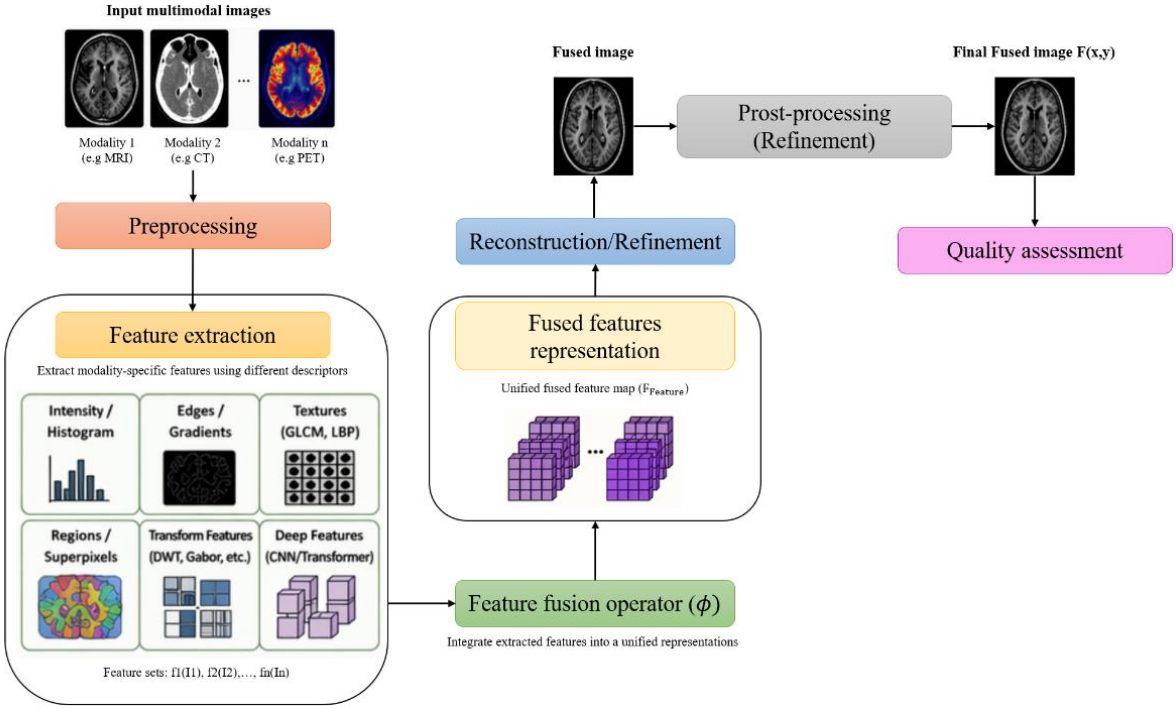

Fig. 2. Feature-Level fusion framework.

across medical modalities and improving uncertainty estimation for broader disease applicability [42]. Wang et al. use tensor chain decomposition to reduce high-dimensional multimodal data into compact latent representations. Implements two-stage attention fusion: intra-latent subspace fusion (ILSF) followed by cross-modal selective fusion (CMSF), with frequency filtering for noise suppression. Focuses on classification rather than pure fusion [104].

The primary advantages of feature-level fusion include enhanced robustness to misregistration, reduced sensitivity to noise, and the ability to intelligently integrate information from modalities with different resolutions or contrast characteristics. By operating on meaningful attributes rather than raw pixels, feature-level fusion can preserve semantically important structures (e.g., brain folds, tumor boundaries) while suppressing irrelevant details [97], [98]. However, challenges include the computational cost of feature extraction, the need for careful selection or design of feature extractors, and the potential loss of fine-grained information during the feature representation stage [95], [96]. Feature-level fusion approaches demonstrate strong capability in preserving complementary multimodal information by integrating representations prior to decision-making (Table 3). The LSC+GA region-based fusion method achieved state-of-the-art average scores across multiple metrics, including AG (12.50), SF (27.08), MI (4.65), and $Q_{AB/F}$ (0.77), consistently outperforming seven competing methods in maintaining structural details and contrast while benefiting from the computational efficiency of linear-time LSC segmentation [99]. Similarly, MATR significantly outperformed both representative and state-of-the-art methods on the Harvard medical dataset in terms of visual quality and quantitative metrics, such as Q_{NCIE}, Q_P, and VIF. The method effectively preserves functional metabolic information from SPECT alongside anatomical structures from MRI, while also demonstrating strong generalization across PET-MRI and GFP-PC tasks without requiring fine-tuning [100]. In the context of transformer-based fusion, TIEF achieved superior performance over 11 state-of-the-art methods across multiple metrics, including EN, SF, AG, and $Q_{AB/F}$, and CI. It effectively enhances contrast between tumor subregions (e.g., edema, enhancing

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

tumor, necrotic core) while preserving fine-grained texture details. Its robustness is further validated through downstream segmentation using nnU-Net, achieving the highest Dice scores and lowest HD95 values [38]. Deep feature extraction strategies also show consistent improvements. M4FNet demonstrated average performance gains of 12.8% in SD, 4.1% in MI, 8.5% in $Q_{AB/F}$, and 9.7% in VIF compared to six state-of-the-art methods. By leveraging multi-receptive-field feature extraction and attention-aware fusion, the model effectively preserves both structural and textural characteristics across CT-MRI, MRI-PET, and MRI-SPECT modalities [36]. Likewise, FDGNet achieved substantial improvements over nine competing methods, with gains of 68.68% in NMI, 6.73% in and $Q_{AB/F}$, 12.52% in , and 18.33% in VIF^P. Its hybrid loss formulation and feature difference-guided weighting strategy enable effective preservation of luminance (CT), tissue texture (MRI), and functional information (PET/SPECT), while mitigating intensity degradation [78]. Lightweight architectures further highlight the practicality of feature-level fusion. AMMNet achieved superior fusion quality, with higher AG (13.39), EN (5.33), and SF (60.20) compared to eight state-of-the-art methods, while maintaining ultra-fast inference (~0.036 s per image), making it suitable for near real-time clinical applications [37]. Hybrid intelligent frameworks also contribute to improved fusion quality. The G-CNNs combined with a fuzzy neural network (FNN) achieved gains of 10-13% in MI, 10-20% in SF, 11-14% in SD, and 22-43% in $Q_{AB/F}$ over nine methods. This approach effectively preserves multimodal characteristics, including MR texture, CT structural details, and lesion boundaries, enhancing diagnostic accuracy in cerebrovascular and tumor-related conditions [102]. Finally, DRIFA-Net extends feature-level fusion to classification-driven frameworks, achieving state-of-the-art accuracy (95.6-99.7%) across five diverse medical imaging datasets. Its dual-attention mechanism and integrated uncertainty estimation significantly improve generalizability and reliability, supporting robust clinical deployment across multiple cancer types [42]. Fig. 2 presents the feature-level fusion pipeline, which integrates information at a higher level of abstraction. The process starts with:

a. Input multimodal images, followed by preprocessing, including registration, normalization, denoising, and optional background removal These steps ensure that subsequent feature extraction is consistent across modalities.
b. The core stage is the feature extraction, where modality-specific features are derived. These include low-level descriptors (e.g., intensity histograms, gradients), texture features region-based representations (e.g., superpixels), transform-domain coefficientsand high-level deep

Table 3. Feature-Level Fusion Techniques.

Technique	Evaluation Metrics	Datasets	Imaging Protocols and Privacy	Computational Complexity	Real-World Impact
Region-Based Fusion (LSC + GA) [99]	• AG: 12.50 (avg) • SF: 27.08 (avg) • MI: 4.65 (avg) • ($Q_{AB/F}$): 0.77 (avg) • Best AG/SF across all test pairs	• 8 multimodal pairs • CT-MRI, MRI-SPECT, MRI-PET • 256×256	• Public, de-identified datasets • Pre-registered images assumed	• LSC: linear time (~0.09 s) • GA: ~83-85 s (bottleneck) • ~100× faster than Ncuts	• Region-level fusion preserves boundaries • Noise-robust • Suitable for offline diagnosis
MATR [100]	• (Q_{NCIE}, Q_P): top scores • (Q_{MI}, Q_{TE}, Q_G, VIF): highest averages • LMI, MS-SSIM: competitive • (Q_{CV}): lowest (best)	• 354 SPECT-MRI pairs • + PET-MRI (10), GFP-PC (18) • 256×256	• Harvard public dataset • RGB→YUV alignment • Pre-registered	• Transformer-based fusion • RTX 3090, Adam (lr=0.001) • Efficient via shifted-window attention	• Preserves metabolic + anatomical info • Strong generalization

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

Table 3. Feature-Level Fusion Techniques.

Technique	Evaluation Metrics	Datasets	Imaging Protocols and Privacy	Computational Complexity	Real-World Impact
M4FNet [36]	• SD: +12.8% • MI: +4.1% • ($Q_{AB/F}$): +8.5% • VIF: +9.7%	• CT-MRI, MRI-PET, MRI-SPECT • Public datasets	• De-identified datasets • Standard alignment assumed	• DHDCB + DWT multi-scale fusion • Attention-aware modules • End-to-end training	• Enhances texture + structural details • Captures long-range dependencies
AMMNet [37]	• AG: 13.39 / 6.31 • EN: 5.33 / 5.18 • SF: 60.20 / 31.12 • MI: 2.59 / 3.20 • ($Q_{AB/F}$): 0.57 / 0.49	• AANLIB (200 pairs) • MRI-PET, MRI-CT • 256×256	• Public dataset • Pre-registered multimodal images	• Lightweight (MobileNetV3 + ECA-S) • 30 epochs, lr=1e-4 • ~30-40 ms inference	• Near real-time capability • Strong edge/texture preservation • Extensible to other fusion tasks
TIEF [38]	• EN ≈ 7.8, SF ≈ 35.2, AG ≈ 14.3 • SSIM ≈ 0.89 • ($Q_{AB/F}$) ≈ 0.79 • CI ≈ 1.42 • PSNR ≈ 28.5 dB • MI ≈ 3.9 • Dice: 0.88-0.93 • HD95 ≈ 3.2 mm	• BraTS2019 (n=335) • AANLIB (75 pairs) • CT-MRI generalization	• Public datasets • RGB→YUV alignment • Pre-registered	• Transformer + SDCL • 320 epochs, batch=32 • A100 GPU	• Accurate tumor delineation • Preserves texture + structure • Generalizes without fine-tuning
FDGNet [78]	• NMI: +68.68% • ($Q_{AB/F}$): +6.73% • (Q_P): +12.52% • (VIF^P): +18.33%	• Harvard dataset (n=166) • 6 modality categories • 256×256	• Public, de-identified • Pre-registered images	• Unsupervised end-to-end • Feature-difference guided weighting • Hybrid loss function	• Preserves luminance + texture + function • Prevents intensity degradation
G-CNNs + FNN [102]	• MI: +10-13% • SF: +10-20% • SD: +11-14% • ($Q_{AB/F}$): +22-43%	• Harvard + Shanxi datasets • CT-MR brain images • 256×256	• Public, de-identified • Pre-registered	• 16 G-CNNs + 5-layer FNN • GTX 1080Ti • Pre-train + fine-tune	• Preserves lesion boundaries • Handles uncertainty via fuzzy logic • Supports clinical diagnosis
DRIFA-Net [42]	• D1: 98.2 / 97.9% • D2: 95.6 / 95.5% • D3: 98.4 / 98.4% • D4: 99.7 / 99.5% • Dice (BraTS): WT 93.6%, TC 90.5%, ET 85.6%	• HAM10000, SIPaKMeD, MRI, Lung CT, BraTS2020 • 5 datasets	• Public datasets • Standard preprocessing • Data augmentation	• Ensemble MC Dropout • 200 epochs, batch=32 • RTX 4060 Ti	• Multi-cancer generalization • Improves reliability via uncertainty • Clinically robust decision support

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

c. features obtained from CNNs or Transformer architectures. Each modality is thus represented as a feature set capturing its unique characteristics.
d. The feature fusion operator integrates these features into a unified representation. This can involve concatenation followed by dimensionality reduction, weighted fusion based on feature importance, attention mechanisms, graph-based learning, or optimization-based selection (e.g., GA, GWO).
e. The result is a fused feature representation, which encodes complementary structural, textural, and semantic information from all modalities.
f. In the reconstruction/refinement step, the fused features are optionally mapped back to the image domain or refined using multi-scale enhancement, edge preservation, and artifact suppression techniques. The reconstructed image is then generated in fused image generation, followed by post-processing, (contrast enhancement, denoising, and sharpening)
g. The result, shown in the final fused image, is evaluated in quality assessment using both image quality metrics (e.g., SSIM, entropy) and task-specific metrics (e.g., Dice score, classification accuracy), demonstrating improved discriminative capability and robustness.

Decision-level fusion operates on outputs from independently processed modalities, employing rules or classifiers to produce a final diagnosis or interpretation. This taxonomy enables a systematic review of intelligent techniques across all stages of the fusion pipeline [83]. Decision-level fusion represents the highest level of image or data fusion, where the fusion process operates not on raw pixel intensities or extracted features, but on the discrete outputs or decisions made independently by multiple classifiers or sensors. Instead of combining data at the pixel or feature level, this approach first allows each modality or source to reach an individual preliminary decision (e.g., tumor present vs. absent, or a specific class label). These independent decisions are then combined using logical or probabilistic rules to produce a final, more informed consensus decision, a strategy widely adopted in recent medical imaging studies to integrate heterogeneous modalities [105], [106], [107]. This strategy minimizes the amount of transmitted or processed data, as only compact decision outputs, rather than entire images or feature vectors, need to be communicated and fused [108], [109].

Unlike pixel-level or feature-level fusion, which require strict registration or compatibility between feature spaces, decision-level fusion allows heterogeneous modalities to be processed independently using specialized algorithms tailored to each data type. This makes it particularly suitable for

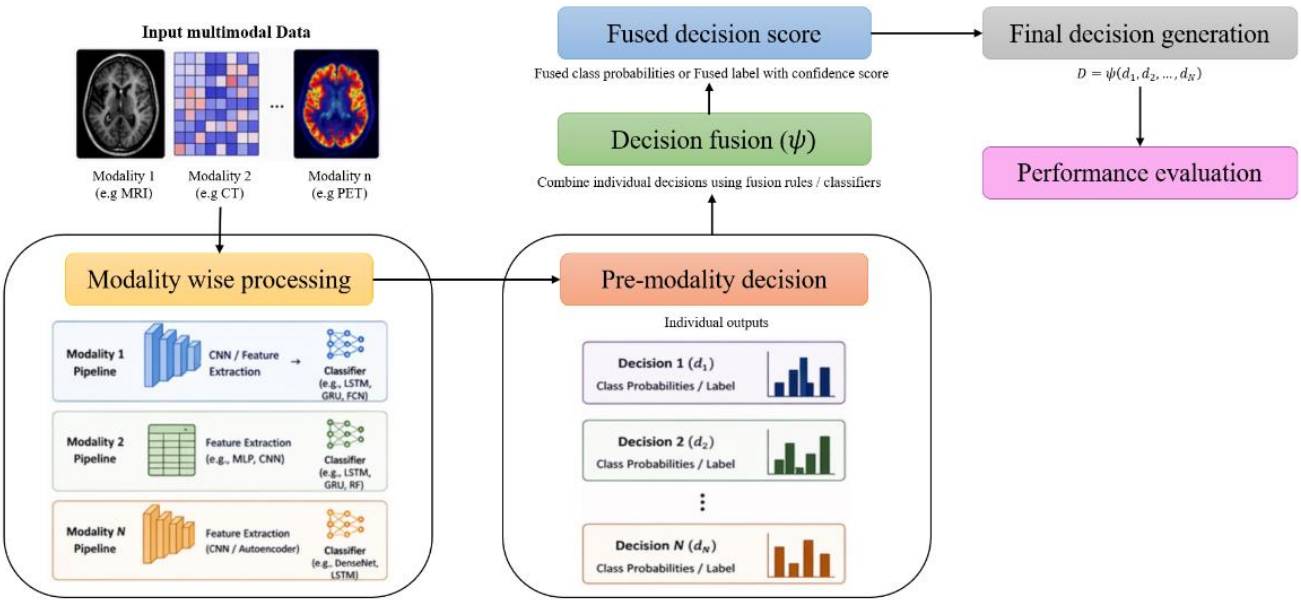

Fig. 3. Decision-Level fusion framework.

C. Decision-level fusion techniques

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

applications such as multimodal biometric verification [109] or medical diagnosis, where different sensors (e.g., MRI, CT, PET scanners) produce fundamentally different data structures that cannot be easily concatenated or compared directly [107], [109], [110]. The general principle can be expressed as a two-stage process: first, each modality classifier independently maps its input to a discrete decision $d_t \in$ {Accept,Reject} or a class label; second, a fusion rule $\phi_{decision}$ combines these individual decisions into a final global decision D, as shown in (13) [106], [108], [109]:

$$D = \phi_{decision}(d_1(I_1(x)), d_2(I_2(x)), \ldots, d_T(I_T(x))) \quad (13)$$

where d_t represents the decision output of the t-th modality classifier, and $\phi_{decision}$ is the decision-level fusion rule (e.g., majority voting, AND/OR logic, or Dempster-Shafer combination) [105], [109]. Several approaches have been proposed to achieve optimal decision-level fusion. One of the simplest and most widely used approaches is majority voting, where the final decision is determined by the majority of individual modality decisions. In medical imaging, similar voting mechanisms are widely adopted to aggregate multi-model predictions, such as in multi-omics breast cancer survival prediction [111] and glioma classification [106]. For binary classification the final decision can be expressed as in (14) [110]:

$$D = \begin{cases} \text{Accept} & \text{if } \sum_{t=1}^{T} \mathbb{1}(d_t = \text{Accept}) > T/2 \\ \text{Reject} & \text{otherwise} \end{cases} \quad (14)$$

where $\mathbb{1}(\cdot)$ is the indicator function, and T is the total number of modalities [110].

Another common formulation employs serial and parallel combination rules, also known as "AND" and "OR" logic. The AND rule requires all modalities to agree on acceptance to reach a final positive decision, which reduces the False Acceptance Rate (FAR) but may increase the False Rejection Rate (FRR). Conversely, the OR rule accepts if any modality accepts, improving FRR at the potential cost of FAR. These rules, which can also be adapted for medical sensitivity and specificity trade-offs [107], are expressed in (15) and (16) [110]:

$$D_{\text{AND}} = d_1 \wedge d_2 \wedge \ldots \wedge d_T \quad (15)$$

$$D_{\text{OR}} = d_1 \vee d_2 \vee \ldots \vee d_T \quad (16)$$

For more sophisticated decision fusion, the Dempster-Shafer Theory (DST) of evidence provides a rigorous mathematical framework for combining uncertain and potentially conflicting decisions. In this approach, each modality's decision is represented as a mass function m_t over the frame of discernment $\Theta = \{\theta_1, \theta_2\}$ (e.g., genuine and impostor). The combined mass function is obtained using Dempster's rule of combination, as shown in (17) [105], [109]:

Table 4. Decision-Level Fusion Techniques.

Technique	Evaluation Metrics	Datasets	Imaging Protocols and Privacy	Computational Complexity	Real-World Impact
Hybrid DL Decision-Level Fusion Model [111]	• Accuracy: 98.0% • Precision: 99.0% • Sensitivity: 99.2% • MCC: 93.6% • AUC: 98.2%	• METABRIC (n=1,980) • 1,489 long-term / 491 short-term survivors	• Multi-omics: clinical + gene expression + CNA • Public dataset • Preprocessing: min-max normalization, mRMR	• CNN (1 conv) + LSTM/GRU (4 layers) • Params: ~157K-189K • Adam (LR=10^{-3}), 10-fold CV	• Survival prediction (>5 yrs) • Supports precision oncology • Enhances treatment stratification
CNN (Imaging) + Deep ELM (EMR) via Fuzzy Decision Fusion. [107]	• 97.97% Accuracy; 2.5% Miss Rate.	• Multimodal Imaging + EMR (incl. Wisconsin dataset).	• MRI, CT, PET, Histopathology. Processed via Cloud framework.	• Hierarchical DL, multi-layer training, cloud infrastructure.	• Rapid, non-destructive CAD for early detection and staging
Multimodal Decision Fusion for Glioma Classification [106]	• Accuracy: 0.878 • AUC: 0.902 • Sensitivity: 0.772 • Specificity: 0.930 • Kappa: 0.773	• CPM-RadPath 2020 (n=221) • 3 classes: GBM / Astro / Oligo	• Preprocessing: bias correction, co-registration	• DenseNet-121 • Weighted decision fusion • 100 epochs, batch=8 • GPU: A100	• Glioma subtype classification • Supports preoperative planning

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C), \forall A \subseteq \Theta, A \neq \emptyset \quad (17)$$

where $\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ is the degree of conflict between the two pieces of evidence, and the final decision is made by maximizing the resulting plausibility or pignistic probability [109].

In the context of Dempster-Shafer decision fusion, the final decision can also be derived from the combined plausibility of each hypothesis. The plausibility $Pl(\theta_k)$ represents the degree to which the evidence does not rule out θ_k, and the final decision selects the hypothesis with maximum plausibility, as expressed in (18) [105], [109]:

$$D = \arg\max_{\theta_k \in \Theta} Pl(\theta_k) = \arg\max_{\theta_k \in \Theta} \left(\sum_{B \cap \{\theta_k\} \neq \emptyset} m(B) \right) \quad (18)$$

where m is the combined mass function from all modalities [105].

Information Content: Pixel-Level > Feature-Level > Decision-Level (19)

At the decision level, a framework for breast cancer survival prediction fuses multi-omics data using CNN-based feature extraction and LSTM/GRU classifiers, combining their outputs to enhance prediction robustness and accuracy across biological datasets [111]. In another approach for intelligent breast cancer diagnosis, a decision-based fusion empowered with fuzzy logic integrates deep learning outputs from multimodal medical imaging and electronic medical records, achieving an overall accuracy of 97.97% in predicting breast cancer stages [107]. In a dual-center study on rectal cancer, a multimodal deep transfer learning framework employs decision-level fusion to integrate optimal deep learning models built on CT and MRI modalities, combined with clinical risk factors via logistic regression and soft voting to enhance tumor

Table 4. Decision-Level Fusion Techniques (continued).

Technique	Evaluation Metrics	Datasets	Imaging Protocols and Privacy	Computational Complexity	Real-World Impact
DL + Dempster-Shafer Fusion Framework [105]	• Dice: up to 0.875 (BraTS) •Uncertainty: ECE↓, Brier↓, NLL↓ •Reliability coefficients (β)	•PET-CT lymphoma (n=173) • BraTS 2021 (n=1,251)	• PET-CT: registered, normalized (256×256×128) •MRI: FLAIR/T1/T1Gd/T2 • IRB approval (clinical data)	• nnUNet / nnFormer backbones • EM-based fusion: O(I(H+K)) • Training ~6.4h (A100)	• Tumor segmentation with uncertainty • Improves clinical trust • Explains modality contribution
Ensemble CNN for Multimodal Classification [112]	• Accuracy: 98.1% • F1: 97.5% • Kappa: 90.8% • Jaccard: 93.8%	• ISL dataset (MRI + PET) • Size not specified	• MRI + PET fusion • Image alignment (GIMP) •Morphological preprocessing	• Ensemble: VGG19, ResNet50, DenseNet121, SqueezeNet • Weighted voting • 5-fold CV	•Alzheimer's classification •Tumor segmentation • Robust via model diversity
2.5D Vgg11 + Clinical data via Soft Voting [113]	• AUC (0.898 internal, 0.868 external), Accuracy, Sensitivity, Specificity, F1-score, DCA.	• Dual-center cohort (355 patients total).	• Contrast CT & T1CE MRI (1.5T/3.0T). IRB approved, anonymized, consent waived.	• Pre-trained 2.5D CNNs + lightweight logistic fusion	• Non-invasive Tumor Budding grading to guide post-endoscopy treatment.

A key characteristic of decision-level fusion is the significant reduction in information compared to lower fusion levels. While pixel-level fusion preserves rich spatial details and feature-level fusion retains attribute-level information, decision-level fusion condenses all information into a small set of discrete outputs. This is illustrated conceptually in Eq. (19) [109], [110], [112]:

budding grading prediction [113]. Another framework combines deep learning with Dempster-Shafer theory to fuse evidence from multiple imaging modalities, discounting uncertainties to improve reliability and interpretability in medical segmentation [105]. For glioma subtype classification, a decision-level method employs DenseNet-based modality-specific models with a linear weighted fusion strategy, enhanced by

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

tumor-focused preprocessing and probabilistic integration across MRI sequences [106]. In tumor analysis, a supervised ensemble of CNNs, VGG19, ResNet50, SqueezeNet, and DenseNet121, is fine-tuned on multimodal inputs and combined via weighted voting, significantly improving segmentation and classification accuracy over individual models [112].

The primary advantages of decision-level fusion include computational efficiency, robustness to modality failure, and the ability to combine heterogeneous data types without requiring common feature spaces [105], [109]. Additionally, decision-level fusion can provide transparency and explainability, as the contribution of each modality to the final decision can be explicitly analyzed, for example, through learned reliability coefficients [105]. However, challenges include the potential loss of valuable discriminative information that is present in raw data or features but lost during the thresholding process into discrete decisions, suboptimal performance compared to feature-level fusion in some scenarios, and the need for careful calibration of individual modality thresholds to avoid bias [105], [108], [109]. The reviewed decision-level fusion approaches consistently demonstrate superior performance by effectively integrating complementary modality-specific predictions (Table 4). The proposed CNN-LSTM/GRU hard voting ensemble achieved outstanding results on the METABRIC dataset, with 98.0% accuracy, 99.2% sensitivity, and 98.2% AUC, clearly outperforming individual classifiers and conventional baselines in binary breast cancer survival prediction [111]. Furthermore, the intelligent breast cancer prediction system utilizing fuzzy logic decision fusion achieved a high diagnostic accuracy of 97.97% by effectively combining imaging and clinical feature-based models [107]. Similarly, the multimodal deep transfer learning model for rectal cancer demonstrated that decision-level fusion of CT and MRI models with clinical data significantly outperformed single-modal approaches, achieving an AUC of 0.898 in internal validation and 0.868 in external validation cohorts [113]. In the context of uncertainty-aware fusion, the MMEF-UNet framework significantly reduced uncertainty metrics (ECE, Brier score, and NLL) while achieving Dice scores of 0.811 for lymphoma and 0.875 on BraTS2021. The incorporation of learnable reliability coefficients enabled adaptive modality weighting, enhancing segmentation reliability and clinical trust compared to conventional pixel-level fusion methods [105]. For glioma classification, a linear weighted decision fusion of multiple DenseNet-121 models achieved 87.8% accuracy, 0.902 AUC, and a Cohen's Kappa of 0.773 on the CPM-RadPath dataset. This approach effectively leveraged the complementarity of multi-sequence MRI, outperforming both radiomics-based and early fusion techniques [106]. Ensemble-based decision fusion further demonstrated strong performance, with a weighted voting scheme combining VGG19, ResNet50, SqueezeNet, and DenseNet121 achieving 98.1% accuracy and 97.5% F1-score on the ISL dataset. The integration of MRI and PET modalities consistently outperformed single-modality baselines in both tumor segmentation and classification tasks [112].

Fig. 3 depicts the decision-level fusion framework, which combines outputs from independently processed modalities. The pipeline begins with:

a. Input multimodal data, which may include imaging data (MRI, CT, PET), as well as non-imaging sources such as clinical or omics data.
b. Each modality is processed separately in modality-wise processing pipelines, where features are extracted (e.g., via CNNs or autoencoders) and passed to classifiers.
c. The results in per-modality decisions, where each branch produces outputs such as class probabilities, predicted labels, or isk scores.
d. The individual decisions are then combined in decision fusion, which employs various strategies, including rule-based methods (e.g., majority voting, weighted voting, Bayesian fusion), ensemble learning (e.g., stacking, averaging), or uncertainty-aware approaches (e.g., Dempster–Shafer theory, reliability weighting).
e. The fusion process produces fused decision scores, representing combined probabilities or belief measures. These are then used in the final decision-making step, where the ultimate prediction (e.g., classification) is determined. Optional post-processing steps refine the output through thresholding, calibration (e.g., Platt scaling), conflict resolution, and uncertainty estimation.
f. The final output provides a robust and interpretable prediction, including confidence or reliability measures. Performance evaluation assesses the framework using classification, probabilistic, segmentation (Dice, IoU), and calibration metrics.

IV. Datasets and evaluation metrics

Based on Table 2, Table 3, and Table 4 from the reviewed MMIF literature [83], the datasets employed across fusion levels predominantly consist of publicly available, de-identified medical imaging repositories that facilitate reproducible benchmarking while adhering to privacy regulations such as HIPAA and GDPR [3]. At the pixel level, studies frequently utilize the Whole Brain Atlas (AANLIB) with standardized 256×256 pre-registered image pairs spanning MRI-CT, MRI-PET, and MRI-SPECT modalities [88], [89], [87], alongside the Harvard Medical School dataset and the

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

Table 5. Summary of State-of-the-Art MMIF Methods by Fusion Level.

Model / Study	Architecture / Method	Scalability	Complexity	Advantages	Limits / Application (Disease)
Pixel-Level Fusion					
OWAF-PSO [88]	DWT decomposition + Optimum Weighted Average Fusion weights optimized by PSO.	Medium	Medium	Robust against noise; better than standard DWT/FFT.	Requires pre-registered images; moderate speed.
KDE-GAN [90]	Knowledge Distillation + Explainable GAN + Dual Discriminators + Perceptual Loss	High	High	Requires less training data, early stopping avoids overfitting	Complexity of dual discriminators; requires explainability tuning
U-Patch GAN [44]	U-Net Generator + Dual PatchGAN Discriminator + Spectral Normalization + VGG-16 Feature Loss	High	High	End-to-end; captures fine detail and semantics; robust to modality variation	Requires large datasets and training time; less interpretable than rule-based methods; Brain imaging
Variable-Order MO-DPSO Fusion [87]	Variable-order fractional-order + Multi-objective Darwinian PSO + Gradient compass	High	High	Adaptive convergence, enhanced spatial details, real-time capable	Complex implementation, requires parameter tuning; CT-MRI fusion for diagnosis and treatment planning

Cancer Imaging Archive (TCIA) GLIS-RT cohort comprising 230 patients with over 13,000 annotated slices [43]. Evaluation at this level emphasizes structural fidelity and information preservation through metrics such as Structural Similarity Index (SSIM: 0.915-0.946) [88], Peak Signal-to-Noise Ratio (PSNR: 32.1-37.8 dB) [92], Mutual Information (MI: 3.50-4.97) [91], edge preservation (Q^AB/F: 0.77-2.74) [88], [91], and entropy-based measures (4.9-10.93 bits/pixel) [89], [90] with clinical segmentation tasks additionally reporting Dice coefficients (0.96±0.02) and Hausdorff Distance (1.22 mm) for tumor delineation validation [43]. At the feature level, datasets expand to include multi-institutional benchmarks such as BraTS2019 (n=335 glioma cases) [38], HAM10000 for dermatological imaging [42], and SIPaKMeD for cervical cancer screening [42], often incorporating data augmentation and cross-modality generalization tests (e.g., CT-MRI, PET-MRI, SPECT-MRI) [36], [78].

Metrics here balance image quality with semantic robustness: Spatial Frequency (SF: 27.08-60.20) and Average Gradient (AG: 6.31-13.39) quantify texture preservation [37], [99] while task-driven validation employs Dice scores (0.88-0.93) and 95th-percentile Hausdorff Distance (HD95 ≈ 3.2 mm) for segmentation accuracy [38]. Notably, classification-oriented feature fusion reports accuracy ranges of 95.6-99.7% across five diverse medical datasets, underscoring the clinical

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

Table 5. Summary of State-of-the-Art MMIF Methods by Fusion Level (continued).

Model / Study	Architecture / Method	Scalability	Complexity	Advantages	Limits / Application (Disease)
MMIF-MDWTAOA [89]	Bilateral Filtering (denoising) + Modified DWT (MDWT) + Fusion rule parameters optimized by Arithmetic Optimization Algorithm (AOA).	Medium	Medium	High structural similarity; effective noise removal.	Higher computational cost than basic DWT.
DTV-LSC-NSCT [91]	CNN Denoising + Non-Subsampled Contourlet Transform (NSCT) + Directional Total Variation (Low-freq) and SML (High-freq) fusion.	Medium	High	Highest quality; shift-invariant; superior edge/texture preservation.	High complexity; not suitable for real-time or ultrasound.
MedFusionGAN [43]	GAN-based architecture with one generator, PatchGAN discriminator; trained with L1, perceptual, SSIM, and gradient losses	High	High	End-to-end unsupervised training; preserves both bone and soft tissue contrast; no artifact generation	Computationally intensive; sensitive to training dynamics and hyperparameter settings; Brain tumors, radiotherapy planning
YUV based SVD-VGG [92]	YUV color space + SVD luminance decomposition + lightweight VGG19 feature gating + hybrid denoising (Bilateral/NLM/Guided filters)	Medium	Medium	Color-preserving (PET chrominance intact); noise-aware preprocessing; computationally efficient; modular design; no ground-truth dependency for fusion	Evaluated on brain MRI-PET with synthetic noise; scalar high-frequency gate may soften very fine textures; MRI-PET fusion for tumor delineation and metabolic-anatomical integration
Anisotropic-CBF Fusion Framework [93]	Anisotropic diffusion + morphological edge processing + pixel significance weighting + cross bilateral filtering (CBF)	Medium	Medium	Preserves fine edges, reduces blur/noise, effective pixel-level weighting strategy	Limited adaptability to high-level semantic features, non-learning-based; General imaging (brain, bone, etc.)

relevance of uncertainty-aware metrics like Expected Calibration Error (ECE) and Brier score [42].

For decision-level fusion, studies integrate heterogeneous data types beyond imaging alone,

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

Table 5. Summary of State-of-the-Art MMIF Methods by Fusion Level (continued).

Model / Study	Architecture / Method	Scalability	Complexity	Advantages	Limits / Application (Disease)
PPMF-Net [31]	Hierarchical multi-path CNN with DEEM, NIFE (cross-modal interaction), TEGM (transformer global modeling), and unsupervised multi-objective loss	High	High	Preserves local structure and global context; cross-modal interaction; robust to low SNR	High computational cost; may require GPU inference; Neuroimaging, oncology (PET-MRI fusion tasks)
LWT-SVD [25]	VGG-19 feature extraction + Dynamic inlier selection + Thin Plate Spline Interpolation (TPSI) + LWT decomposition + SVD-based fusion	High	High	Edge-preserving, fast, robust to multimodal variations, clinically validated	TPSI cost, parameter-sensitive; Brain tumor/lesion diagnosis (MRI/CT/PET)
Feature-Level Fusion					
TIEF [38]	SDCL-Block + Entropy-based Feature Selection + Transformer Fusion + Attention + Modality-aware Loss	High	High	Robust cross-modal fusion, spatial attention, strong generalization, segmentation integration	Computational complexity due to transformer and attention layers; training requires multi-sequence data; Glioma and meningioma brain tumors
AMMNet [37]	Multi-scale CNN + DenseNet + ECA-S attention + MobileNetV3 decomposition network	High	High	Enhanced texture retention, deep feature reuse, lightweight MobileNetV3, attention-guided fusion	Requires large annotated datasets; training cost is non-trivial; General diagnosis
G-CNNs + Fuzzy NN Fusion [102]	Gabor filter bank + Multi-CNN ensemble (G-CNNs) + Fuzzy Neural Network (FNN) fusion mechanism	Medium	High	Enhanced edge/texture fusion, uncertainty handling via FNN, modality-specific Gabor enhancement	Complex training pipeline; sensitive to Gabor parameter selection; Brain and lesion imaging
FDGNet [78]	End-to-end unsupervised CNN; feature-weighted guided learning + hybrid loss (weighted fidelity + feature difference)	High	High	Preserves complementary features, adaptive weighting, luminance retention	High training cost; needs substantial labeled or unlabeled multimodal data for generalization

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

Table 5. Summary of State-of-the-Art MMIF Methods by Fusion Level (continued).

Model / Study	Architecture / Method	Scalability	Complexity	Advantages	Limits / Application (Disease)
M4FNet [36]	Dual-Branch Dense Hybrid Dilated Convolution Blocks (DHDCB) + Wavelet + Attention-Aware Fusion + Deconvolution + Custom Structural Similarity Loss	High	High	Captures long-range and multi-scale features, strong semantic preservation, task-adaptive	Requires large labeled data and high computational resources; General organ-level medical fusion
FDCuT + Type-2 Fuzzy Entropy + A-EFO [103]	Fast Discrete Curvelet Transform + Averaging (LF) + Optimized Type-2 fuzzy entropy (HF) + Adaptive Electric Fish Optimization	Medium	Medium	Handles degradation, robust fusion of frequency sub-bands, multi-objective quality optimization	Requires parameter tuning and sufficient frequency component alignment; General multimodal diagnosis
DRIFA-Net [42]	Dual attention (multi-branch + multimodal) + deep neural network + Monte Carlo dropout for uncertainty estimation	High	High	Generalizable across modalities and diseases, dual attention improves representation, uncertainty modeling	Attention module increases model complexity; requires multi-modal data alignment
MATR [100]	Multiscale DL architecture with adaptive convolution and Transformer modules; Structural + Region Mutual Information loss	High	High	Captures both local and global features; multiscale design; strong generalization	High computational cost; requires large training data; CT-MRI, PET-MRI fusion for diagnosis and navigation
Region-Based Fusion with LSC + GA [99]	Superpixel segmentation (LSC) + Feature extraction (LGF, SML) + Genetic Algorithm optimization	Medium	Medium	Preserves structural detail and regional semantics; adaptive optimization of weights	Increased computational cost due to segmentation and GA; depends on quality of superpixels

(n=1,980) with multi-omics profiles (gene expression, copy-number alterations, clinical variables) [111], and the CPM-RadPath 2020 glioma collection featuring multi-center, multi-field-strength MRI (1T-3T) with

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

Table 5. Summary of State-of-the-Art MMIF Methods by Fusion Level (continued).

Model / Study	Architecture / Method	Scalability	Complexity	Advantages	Limits / Application (Disease)
Decision-Level Fusion					
Hybrid DL Decision-Level Fusion Model [111]	CNN for feature extraction + LSTM and GRU classifiers + Decision fusion strategy (voting)	High	Medium-High	Combines temporal and spatial patterns, robust to data heterogeneity, improves interpretability	Limited to METABRIC dataset; fusion logic is rule-based; Breast cancer survival prediction
CNN (Imaging) + Deep ELM (EMR) via Fuzzy Decision Fusion. [107]	CNN (Imaging) + Deep ELM (EMR features) combined via Decision-based Fuzzy Fusion.	High	High	Achieved high accuracy; rapid, non-destructive CAD system for staging.	Requires cloud infrastructure; system must retrain if initial learning criteria are not met.
DL + Dempster-Shafer Fusion Framework [105]	Deep learning for segmentation + Dempster-Shafer theory + contextual discounting + evidence aggregation	Medium	Medium	Uncertainty modeling, interpretable decision fusion, improved segmentation accuracy	Fusion relies on quality of discounting and evidence modeling; Lymphoma detection, brain tumor segmentation
Multimodal Decision Fusion for Glioma Classification [106]	DenseNet (per modality) + tumor segmentation + decision-level linear weighted fusion	Medium	Medium	High accuracy, interpretable decision fusion, modular and extensible to other modalities	Requires pre-trained segmentation and separate modality-specific training; Glioma subtype classification
Ensemble CNN for Multimodal Classification [112]	MRI + PET fusion at preprocessing + CNNs (VGG19, ResNet50, SqueezeNet, DenseNet121) + Weighted Voting Ensemble	High	Medium-High	Combines multi-modal input with ensemble learning, high accuracy, robust segmentation	Requires large multimodal datasets; less spatial feature modeling during fusion; Tumor detection, lesion segmentation
2.5D Vgg11 + Clinical data via Soft Voting [113]	2.5D Vgg11 (CT & MRI T1CE) + Clinical factors integrated via Soft Voting / Logistic Regression.	High	Medium	Non-invasive tumor budding grading; High AUC; Grad-CAM provides clinical interpretability	Retrospective design; 2.5D inputs miss some 3D spatial info; only utilized T1CE MRI sequence

standardized bias correction and co-registration [106]. Evaluation prioritizes diagnostic performance: accuracy (87.8-98.1%), AUC-ROC (0.902-0.982), F1-score (97.5%), and Cohen's Kappa (0.773-0.908) [111], [106], [112], while uncertainty-aware frameworks incorporate reliability coefficients (β) and evidence discounting via Dempster-Shafer theory to quantify modality contribution and decision confidence [105]. These metrics reflect an evolving paradigm wherein quantitative benchmarks are increasingly complemented by clinical utility measures, such as inter-observer variability reduction and workflow integration, to ensure MMIF systems translate effectively from algorithmic innovation to precision medicine practice [6], [9], [13]

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

V. Discussion

The rapid evolution of MMIF has significantly advanced diagnostic imaging, enabling clinicians to leverage complementary information from diverse modalities, such as MRI, CT, PET, and SPECT. As illustrated in Section III, MMIF techniques can be systematically categorized into three hierarchical levels: pixel-level, feature-level, and decision-level fusion. Each level offers distinct advantages and trade-offs in terms of information fidelity, computational complexity, interpretability, and clinical applicability (Table 5). However, a critical meta-analysis of the reviewed literature reveals substantial heterogeneity in reported performance, with conflicting results frequently arising from dataset composition, evaluation metric selection, and deployment context.

A. Pixel-Level Fusion: High Fidelity at a Computational Cost

Pixel-level fusion methods aim to preserve maximum spatial detail by directly combining raw intensity values across modalities. Recent deep learning-based approaches like U-Patch GAN [44], MedFusionGAN [43], and PPMF-Net [31] have demonstrated superior performance in preserving both structural and functional details, particularly in neuroimaging and oncology. These models leverage adversarial training, perceptual losses, and transformer-based attention mechanisms to generate high-resolution, artifact-free fused images that retain critical diagnostic features, such as bone contrast in CT and soft-tissue delineation in MRI.

However, performance claims in the literature are often dataset-dependent and occasionally contradictory. For instance, while GAN-based methods consistently outperform classical transforms in SSIM and MI on curated public datasets [44], [43], several studies report comparable or superior edge retention using hybrid optimization-driven frameworks (e.g., PSO, AOA) when trained on small or noisy clinical cohorts [87], [89]. This discrepancy highlights that adversarial architectures may overfit to synthetic distributions or idealized alignment, whereas meta-heuristic optimization remains more robust to limited data but struggles with semantic adaptation.

Non-learning methods like the Anisotropic-CBF framework [93] offer competitive edge preservation and noise reduction with moderate complexity, making them suitable for resource-constrained environments. However, they lack adaptability to semantic content and cannot learn from data, highlighting a key limitation compared to deep architectures. Optimization techniques such as Darwinian PSO and AOA [87], [89] address this gap by dynamically tuning fusion weights without gradient descent, achieving convergence in under 1 second per slice. Yet, compared with standard gradient-based or fixed-rule approaches, these meta-heuristics require careful parameter initialization, exhibit stochastic variability across runs, and lack theoretical convergence guarantees, thereby limiting their reproducibility in regulated clinical pipelines.

B. Feature-Level Fusion: Balancing Semantics and Structure

Feature-level fusion bridges the gap between low-level pixel data and high-level decisions by extracting and integrating modality-specific features such as edges, textures, and regions. This level has seen a surge in hybrid architectures combining convolutional networks, transformers, and attention mechanisms. Models like MATR [100] and TIEF [38] exemplify the trend toward multiscale and cross-modal attention, enabling robust fusion of local textures and global semantics. The integration of entropy-guided selection and transformer fusion in TIEF proves particularly effective for glioma analysis, where spatially-aware feature integration improves segmentation accuracy (high Dice, low HD95). Similarly, DRIFA-Net [42] introduces dual attention and uncertainty estimation via Monte Carlo dropout, enhancing generalization across diverse diseases, including brain tumors and skin cancer.

Despite their high performance, transformer-based architectures suffer from pronounced data dependence and a risk of overfitting. Vision Transformers and hybrid attention models typically require thousands of well-aligned, multi-institutional samples to stabilize self-attention weights, a condition rarely met in medical imaging due to strict privacy regulations and fragmented data silos [3]. When evaluated on external cohorts or different scanner manufacturers, performance degradation exceeding 20-35% is frequently reported [8], [15]. This underscores that architectural sophistication does not guarantee clinical robustness; rather, sample efficiency, domain invariance, and regularization strategies (e.g., dropout, weight decay, contrastive pretraining) are critical mitigators. Methods like G-CNNs + Fuzzy NN [102] attempt to handle uncertainty through fuzzy logic, offering a more interpretable fusion mechanism, albeit at the cost of a complex training pipeline.

A notable trend in feature-level fusion is the use of decomposition strategies (e.g., wavelet, curvelet) combined with intelligent weighting schemes. For example, FDCuT + A-EFO [103] uses multi-objective optimization to fuse frequency sub-bands, achieving robust structural fidelity. This suggests that hybridizing classical signal processing with meta-heuristic optimization remains a viable path, especially when deep learning data requirements are prohibitive.

C. Decision-Level Fusion: Interpretability and Clinical Integration

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

At the highest level, decision-level fusion combines independently processed outputs, such as classification or segmentation results, using ensemble or rule-based strategies. These methods are particularly valuable in clinical decision support systems, where interpretability and reliability are paramount. The Hybrid DL Decision-Level Fusion Model [111] combines CNN-extracted features with LSTM/GRU classifiers, achieving up to 98% accuracy in breast cancer survival prediction. Similarly, Ensemble CNN frameworks [112] leverage weighted voting across multiple deep networks (VGG19, ResNet50, etc.), yielding high recall and F1 scores for tumor classification.

A key innovation is the integration of uncertainty-aware fusion using theories like Dempster-Shafer [105], which allows for evidence discounting and improved reliability in ambiguous cases, critical in lymphoma and brain tumor diagnosis.

While decision-level methods offer modular design and high interpretability, they depend heavily on the quality of individual modality-specific models and require careful alignment and preprocessing. Moreover, their performance is often dataset-specific, limiting generalization without extensive retraining. The scarcity of large-scale, multimodal, and clinically annotated datasets, coupled with strict privacy regulations such as HIPAA and GDPR, severely restricts benchmark diversity [3]. Consequently, many models are validated on small, de-identified public repositories (e.g., AANLIB, BraTS), introducing selection bias and inflating reported performance. Real-world clinical workflows frequently encounter missing modalities, inconsistent slice thickness, and variable reconstruction kernels, which decision-level ensembles must explicitly accommodate through imputation, fallback routing, or confidence-weighted voting.

While decision-level fusion is widely valued for its interpretability and modularity, it inherently bypasses early-stage spatial and feature-level alignment, introducing distinct integration challenges that warrant careful consideration. First, although pixel-wise registration is not strictly required, consistent region-of-interest (ROI) extraction or central cropping remains critical to ensure that independent classifiers analyse anatomically comparable structures; spatial mismatches can otherwise produce divergent predictions that are difficult to reconcile during ensemble. Second, decision-level frameworks must contend with semantic and label inconsistencies across modalities (e.g., slice-level imaging annotations versus patient-level clinical records), which can degrade voting or weighting performance if not explicitly mitigated through strategies such as label masking or maximum-likelihood selection. Third, the heterogeneity of modality-specific feature distributions and varying clinical reliability necessitate adaptive integration mechanisms; fixed weighting or hard voting may underutilize complementary evidence or overtrust noisy sources. Recent studies address these limitations by incorporating learnable reliability coefficients, uncertainty-aware discounting, or adaptive likelihood ensembling, thereby preserving the transparency of decision-level fusion while enhancing its capacity to harmonize conflicting or asynchronous multimodal evidence.

D. Modality-Specific Information Preservation

A fundamental yet frequently overlooked challenge in MMIF is the preservation of modality-specific diagnostic signatures during fusion, a critical requirement for clinical utility that transcends mere visual quality metrics [8], [9]. Each imaging modality provides irreplaceable diagnostic information: PET reveals metabolic hotspots through radiotracer uptake that may be anatomically invisible, MRI delineates soft-tissue boundaries with exquisite contrast, CT visualizes bone architecture with millimeter precision, and SPECT captures perfusion dynamics essential for functional assessment. When fusion algorithms indiscriminately blend these modalities, diagnostically critical signatures risk suppression or distortion, such as PET's subtle metabolic gradients being overwhelmed by MRI's structural dominance, or SPECT's blood flow patterns becoming obscured by CT's high-contrast bone visualization [3].

This preservation challenge manifests in documented clinical cases with tangible diagnostic consequences. Successful preservation is demonstrated by frameworks like MedFusionGAN [43] and YUV-SVD-VGG [92], which explicitly decouple structural and functional channels. MedFusionGAN's adversarial loss enforces CT bone retention while preserving MRI soft-tissue contrast, enabling precise radiotherapy target delineation (Dice: 0.96±0.02). Similarly, the YUV-SVD approach maintains PET chrominance fidelity while fusing luminance gradients, achieving sub-second inference without metabolic information loss. Conversely, failed cases frequently arise from aggressive global weighting or irregularized attention. Mirzaei et al. [7] documented instances where PET-MRI fusion for early Alzheimer's diagnosis suppressed subtle amyloid-beta deposition patterns, yielding false-negative interpretations despite high SSIM/PSNR scores. Other studies report that transformer-based cross-attention modules, when insufficiently constrained, tend to prioritize high-gradient anatomical boundaries (MRI/CT) over low-contrast functional signals (PET/SPECT), effectively

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

"washing out" metabolically active but structurally homogeneous lesions [8], [38].

Recent advances address this through three complementary strategies. First, boundary-measured PCNN approaches modulated by multi-scale morphological gradient operators, fidelity [3], adaptively weight fusion decisions based on local structural characteristics, preserving edges from complementary modalities without suppression. Second, dual-branch architectures with modality-specific feature extractors maintain essential characteristics through dedicated pathways before strategic integration [6], ensuring that PET's metabolic signatures and MRI's anatomical details retain diagnostic integrity throughout fusion. Third, information-theoretic fusion rules leveraging mutual information and entropy-based measures dynamically preserve modality-specific signatures by quantifying the information contribution at the pixel and regional levels [10].

Critically, evaluation must evolve beyond conventional metrics like SSIM or PSNR that correlate poorly with clinical utility. Instead, modality preservation requires task-specific validation where radiologists assess whether fused images retain sufficient modality-specific information for diagnosis, a paradigm shift advocated across multiple recent studies [8], [9]. As neurological applications increasingly demand fusion of temporally asynchronous modalities (e.g., EEG-fMRI with hemodynamic lag), preservation strategies must incorporate temporal alignment mechanisms that maintain both spatial correspondence and physiological timing relationships, a frontier challenge requiring specialized attention in future MMIF research [7].

E. Interpretability Crisis and Explainable Fusion

Radiologists cannot ethically base diagnostic decisions on fused images whose generation process remains opaque, particularly when fusion artifacts might mimic pathology or obscure subtle lesions. This interpretability crisis highlights the critical role of XAI, which aims to make AI-driven processes transparent and understandable, improving trust and accountability in clinical decision-making, and it manifests through three interconnected dimensions that demand specialized solutions beyond generic XAI techniques.

First, clinicians require transparent explanations of which modality contributed what information to specific regions of the fused output, a necessity for diagnostic accountability that standard saliency maps fail to provide. For instance, when a fused PET-MRI image reveals a suspicious region, radiologists must know whether this finding derives primarily from PET's metabolic activity, MRI's structural abnormality, or an artifact of fusion, information essential for determining appropriate follow-up [6]. Second, fusion decisions must be explainable at multiple abstraction levels: pixel-level explanations for edge preservation decisions, feature-level explanations for texture integration strategies, and semantic-level explanations for how complementary information resolves diagnostic ambiguity [10]. Third, explanations must align with clinical cognition rather than mathematical abstractions; radiologists think in anatomical structures and imaging signs, not activation maps or gradient flows [7].

Emerging solutions address these requirements through three promising directions. Layer-wise relevance propagation (LRP) adapted specifically for fusion architectures decomposes fusion decisions into modality-specific contributions, generating heatmaps that highlight PET-derived metabolic signatures versus MRI-derived structural features within the same fused image [6]. Transformer-based architectures with cross-attention visualization reveal inter-modality relationships by mapping how specific MRI regions influenced PET interpretation during fusion, a capability particularly valuable for neurological applications where structural lesions may alter functional interpretations [7]. Uncertainty quantification through Bayesian deep learning generates pixel-wise confidence estimates that flag regions where fusion decisions lack diagnostic certainty, enabling radiologists to exercise appropriate caution when interpreting ambiguous fused regions [8].

Despite methodological progress, interpretability in MMIF lacks standardized evaluation frameworks. Most XAI applications are validated using technical proxies (e.g., faithfulness, sensitivity, perturbation stability) rather than clinical endpoints. To bridge this gap, future MMIF systems must adopt structured validation protocols aligned with emerging AI reporting guidelines (e.g., CLAIM, TRIPOD-AI, DECIDE-AI). These should mandate multi-reader, multi-case (MRMC) studies measuring diagnostic confidence, inter-observer agreement (Cohen's κ), and decision latency before and after XAI integration. Only through prospective clinical trials demonstrating reduced diagnostic errors, streamlined workflows, and improved patient stratification can interpretability transition from a technical feature to a clinical necessity [4], [13].

Critically, as Haribubu et al. [9] emphasize, interpretability must be evaluated through clinician-in-the-loop studies measuring whether explanations actually improve diagnostic confidence and accuracy, not merely through technical metrics of explanation fidelity. The ultimate validation of interpretable fusion lies not in algorithmic elegance but in demonstrable improvements in diagnostic accuracy, reduced inter-observer variability, and enhanced radiologist trust,

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

metrics that must become standard in MMIF evaluation frameworks to bridge the chasm between technical innovation and clinical utility.

F. Post-Fusion Artifact Reduction and Clinical Evaluation

Despite advances in fusion architectures, residual artifacts, including pseudo-Gibbs oscillations, contrast inversion, edge blurring, and modality-specific signal bleeding, frequently compromise the diagnostic usability of fused outputs. The review reveals a critical gap: post-fusion refinement techniques are rarely evaluated using clinically meaningful benchmarks. Most studies rely on synthetic artifact injection or full-reference metrics (PSNR, SSIM, MI) computed against pseudo-ground-truth or individual source images, which fail to capture radiologically relevant distortions such as false lesion enhancement, suppressed microcalcifications, or altered perfusion gradients [9], [10].

Effective post-fusion refinement requires task-driven evaluation pipelines. Recent frameworks integrate denoising (e.g., DnCNN, RGF), edge restoration (anisotropic diffusion, guided filtering), and contrast normalization as explicit refinement stages [28], [93]. However, comparative analysis remains fragmented. A clinically robust evaluation should incorporate: (1) detectability indices (e.g., channelized Hotelling observer, AUC-ROC on task-specific lesion detection), (2) radiologist consensus scoring across standardized Likert scales for artifact severity, diagnostic confidence, and clinical actionability, and (3) downstream task correlation, measuring how artifact reduction impacts segmentation Dice scores or classification sensitivity. Without these, claims of "artifact-free fusion" remain algorithmic abstractions rather than clinically verified outcomes. Future MMIF pipelines must embed refinement as a mandatory, quantitatively audited stage, with benchmark datasets explicitly annotated for clinically relevant artifact types rather than purely mathematical distortions.

VI. Conclusion

MMIF represents a new, innovative approach to diagnostic imaging by addressing the significant limitation that no single modality can fully capture the underlying pathophysiology of disease. In this review, we have characterized state-of-the-art MMIF techniques across three levels of hierarchical fusion, which showcases how intelligent methods, especially CNNs, GANs, and Vision Transformers, have changed the way fusion is performed compared to what was previously possible with handcrafted methods. However, as this critical synthesis shows, just because an architecture is sophisticated does not automatically mean it will produce clinically viable applications. Pixel-level methods provide excellent spatial fidelity but also create major computational and data-related bottlenecks; feature-level architectures provide an appropriate balance between structure and semantics but remain sensitive to protocol shifts and sample inefficient; decision-level approaches offer strong interpretability and modularity but require consistent ROI alignment and suffer from dataset bias.

The use of deep learning in clinical practice faces four fundamental issues that prevent the general uptake of AI technology: (1) clinicians do not yet trust the “black box” nature of AI models; (2) diagnostic features are lost during data fusion, limiting their use for diagnostic purposes; (3) differences in vendor scanners and acquisition protocols limit generalization to new datasets; and (4) artifacts caused by a lack of post-fusion refinement reduce the overall value of AI-derived images for disease diagnosis.

To create MMIF systems and adapt academic benchmarks into precision medicine tools, research needs to shift toward a clinically driven validation model and develop methods for cross-site robustness. Below are four targeted pathways:

a. Use standardized, cross-institutional datasets from multiple vendors, field strengths, and reconstruction protocols to evaluate models via leave-one-site-out cross-validation, along with reporting protocol-shift degradation metrics.
b. Replace purely mathematical optimization with imaging diagnostic accuracy (MRMC) studies that measure inter-observer variability, integrate workflow procedures, and assess how patient outcomes correlate with imaging-based decisions. Adoption of both DECIDE-AI and CLAIM should be made mandatory for clinical utilization.
c. Develop federated and/or split-learning architectures that enable cross-hospital training without sharing raw data, while addressing constraints posed by HIPAA/GDPR and improving demographic and scanner diversity in training datasets.
d. Implement a standardized, multi-phase post-fusion evaluation pipeline that utilizes radiologist consensus scores, task-based detectability indices, and modality contribution maps (describing the contribution of multimodal imaging). Quality audits of fusion outputs should occur in clinical environments, not only in mathematically optimized settings.

Ultimately, the success of MMIF will be measured not by SSIM or PSNR scores alone, but by demonstrable improvements in diagnostic confidence, reduced inter-observer variability, and seamless integration into clinical workflows. Only through sustained collaboration between AI developers, radiologists, medical physicists,

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

and regulatory bodies can fused images transition from algorithmic curiosities to trusted partners in precision medicine.

Acknowledgment

No acknowledgment.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

No new data were created or analyzed in this study. Data sharing is not applicable to this article as it is based on previously published literature.

Author Contribution

Majda Maatallah and Abdelmadjid Benmachiche contributed to the conceptualization and design of the study.

Khadija Rais conducted the literature review, performed the analysis, and drafted the manuscript.

Selma Touam contributed to the supervision and critical revision of the manuscript. All authors read and approved the final manuscript.

Declarations

Ethical Approval

Not applicable. This study does not involve human participants or animal subjects.

Consent for Publication Participants.

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

References

[1] N. Goswami, A. Dogra, S. Bakshi, and B. Goyal, "Multimodal Medical Image Fusion: Techniques, Databases, Evaluation Metrics, and Clinical Applications -A Comprehensive Review", doi: 10.2174/0118744400417835251022042920.

[2] M. Zubair, M. Hussain, M. A. Albashrawi, M. Bendechache, and M. Owais, "A comprehensive review of techniques, algorithms, advancements, challenges, and clinical applications of multi-modal medical image fusion for improved diagnosis," *Comput. Methods Programs Biomed.*, vol. 272, p. 109014, Dec. 2025, doi: 10.1016/j.cmpb.2025.109014.

[3] W. Tan, P. Tiwari, H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," *Neural Comput. Appl.*, vol. 37, no. 28, pp. 22995–23015, 2025, doi: 10.1007/s00521-020-05173-2.

[4] F. Zhao, C. Zhang, and B. Geng, "Deep Multimodal Data Fusion," *ACM Comput. Surv.*, vol. 56, no. 9, pp. 1–36, Oct. 2024, doi: 10.1145/3649447.

[5] Y. Li *et al.*, "A review of deep learning-based information fusion techniques for multimodal medical image classification," *Comput. Biol. Med.*, vol. 177, p. 108635, Jul. 2024, doi: 10.1016/j.compbiomed.2024.108635.

[6] V. A. Barola, P. Singh, and M. Diwakar, "A Recent Survey on Multi-modal Medical Image Fusion," *Biomed. Inform. Smart Healthc.*, vol. 1, no. 3, pp. 89–97, 2025, doi: 10.62762/BISH.2025.414869.

[7] G. Mirzaei, A. Gupta, and H. Adeli, "Data fusion of medical imaging in neurological disorders," *Rev. Neurosci.*, vol. 37, no. 1, pp. 43–60, 2026, doi: 10.1515/revneuro-2025-0062.

[8] M. A. Saleh, A. A. Ali, K. Ahmed, and A. M. Sarhan, "A brief analysis of multimodal medical image fusion techniques," *Electronics*, vol. 12, no. 1, p. 97, 2022, doi: 10.3390/electronics12010097.

[9] M. Haribabu, V. Guruviah, and P. Yogarajah, "Recent advancements in multimodal medical image fusion techniques for better diagnosis: an overview," *Curr. Med. Imaging Rev.*, vol. 19, no. 7, pp. 673–694, 2023, doi: 10.2174/1573405618666220606161137.

[10] M. Diwakar, P. Singh, V. Ravi, and A. Maurya, "A non-conventional review on multi-modality-based medical image fusion," *Diagnostics*, vol. 13, no. 5, p. 820, 2023, doi: 10.3390/diagnostics13050820.

[11] J. Sui, D. Zhi, and V. D. Calhoun, "Data-driven multimodal fusion: approaches and applications in psychiatric research," *Psychoradiology*, vol. 3, p. kkad026, 2023, doi: 10.1093/psyrad/kkad026.

[12] S. Ullah Khan, M. Ahmad Khan, M. Azhar, F. Khan, Y. Lee, and M. Javed, "Multimodal medical image fusion towards future research: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 8, p. 101733, Sep. 2023, doi: 10.1016/j.jksuci.2023.101733.

[13] S. Steyaert *et al.*, "Multimodal data fusion for cancer biomarker discovery with deep learning," *Nat. Mach. Intell.*, vol. 5, no. 4, pp. 351–362, Apr. 2023, doi: 10.1038/s42256-023-00633-5.

[14] S. Kalamkar and G. M. A., "Multimodal image fusion: A systematic review," *Decis. Anal. J.*, vol. 9, p. 100327, Dec. 2023, doi: 10.1016/j.dajour.2023.100327.

[15] S. Bhosekar, P. Singh, D. Garg, V. Ravi, and M. Diwakar, "A Review of Deep Learning-based

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

Multi-modal Medical Image Fusion", doi: 10.2174/0118750362370697250630063814.

[16] T. M. Hayat and S. Madhavi D., "A Comprehensive Analysis of Medical Image Fusion Techniques: A Detailed Review:," in *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics*, Virtual, India: SCITEPRESS - Science and Technology Publications, 2023, pp. 147–152. doi: 10.5220/0012603200003739.

[17] M. A. Azam *et al.*, "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," *Comput. Biol. Med.*, vol. 144, p. 105253, May 2022, doi: 10.1016/j.compbiomed.2022.105253.

[18] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, "A Review of Multimodal Medical Image Fusion Techniques," *Comput. Math. Methods Med.*, vol. 2020, p. 8279342, Apr. 2020, doi: 10.1155/2020/8279342.

[19] S. Liu, M. Wang, L. Yin, X. Sun, Y.-D. Zhang, and J. Zhao, "Two-Scale Multimodal Medical Image Fusion Based on Structure Preservation," *Front. Comput. Neurosci.*, vol. 15, Jan. 2022, doi: 10.3389/fncom.2021.803724.

[20] M. M. Almasri and A. M. Alajlan, "Artificial Intelligence-Based Multimodal Medical Image Fusion Using Hybrid S2 Optimal CNN," *Electronics*, vol. 11, no. 14, p. 2124, Jan. 2022, doi: 10.3390/electronics11142124.

[21] W. Kong, C. Li, and Y. Lei, "Multimodal medical image fusion using convolutional neural network and extreme learning machine," *Front. Neurorobotics*, vol. 16, p. 1050981, 2022, doi: 10.3389/fnbot.2022.1050981.

[22] K. Vanitha, D. Satyanarayana, and M. N. G. Prasad, "Multi-modal Medical Image Fusion Algorithm Based on Spatial Frequency Motivated PA-PCNN in the NSST Domain," *Curr. Med. Imaging*, vol. 17, no. 5, pp. 634–643, May 2021, doi: 10.2174/1573405616666201118123220.

[23] S. Goyal, V. Singh, A. Rani, and N. Yadav, "Multimodal image fusion and denoising in NSCT domain using CNN and FOTGV," *Biomed. Signal Process. Control*, vol. 71, p. 103214, 2022, doi: 10.1016/j.bspc.2021.103214.

[24] K. Vanitha, D. Satyanarayana, and M. N. Giri Prasad, "Medical Image Fusion Based on Energy Attribute and PA-PCNN in NSST Domain," in *Soft Computing and Signal Processing*, V. S. Reddy, V. K. Prasad, J. Wang, and K. T. V. Reddy, Eds., Singapore: Springer Nature, 2022, pp. 457–467. doi: 10.1007/978-981-16-7088-6_42.

[25] P. Arora, R. Mehta, and P. K. Soni, "Multimodal medical image analysis using deep learning registration and LWT-SVD fusion," *Discov. Comput.*, vol. 29, no. 1, p. 4, 2026, doi: 10.1007/s10791-025-09857-y.

[26] Q. Zuo, J. Zhang, and Y. Yang, "DMC-Fusion: Deep Multi-Cascade Fusion With Classifier-Based Feature Synthesis for Medical Multi-Modal Images," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3438–3449, Sep. 2021, doi: 10.1109/JBHI.2021.3083752.

[27] A. S. Nisha and T. S. Siva Rani, "Novel hybrid CNN with Bi-LSTM multi-focus image fusion method based on modified tetrolet transform in MRI and CT images," *J. Intell. Fuzzy Syst.*, vol. 45, no. 4, pp. 6767–6783, Oct. 2023, doi: 10.3233/JIFS-224439.

[28] D. K. Chaudhary, P. Singh, and A. Shankar, "RGF-DnCNN-GMM: Multi-Modal Medical Image Fusion Using Rolling Guidance Filtering, CNN Denoising, and Gradient-Based Adaptive Fusion," in *2025 5th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, Nov. 2025, pp. 1–5. doi: 10.1109/IOT-SIU65919.2025.11402756.

[29] V. S. Parvathy, S. Pothiraj, and J. Sampson, "Hyperparameter Optimization of Deep Neural Network in Multimodality Fused Medical Image Classification for Medical and Industrial IoT," in *Smart Sensors for Industrial Internet of Things: Challenges, Solutions and Applications*, D. Gupta, V. Hugo C. de Albuquerque, A. Khanna, and P. L. Mehta, Eds., Cham: Springer International Publishing, 2021, pp. 127–146. doi: 10.1007/978-3-030-52624-5_9.

[30] P.-H. Dinh, "A novel approach based on Three-scale image decomposition and Marine predators algorithm for multi-modal medical image fusion," *Biomed. Signal Process. Control*, vol. 67, p. 102536, May 2021, doi: 10.1016/j.bspc.2021.102536.

[31] P. Peng and Y. Luo, "Multimodal Medical Image Fusion Using a Progressive Parallel Strategy Based on Deep Learning," *Electronics*, vol. 14, no. 11, p. 2266, Jan. 2025, doi: https://doi.org/10.3390/electronics14112266.

[32] S. Yu, M. He, R. Nie, C. Wang, and X. Wang, *An unsupervised hybrid model based on CNN and ViT for multimodal medical image fusion*.

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

2021, p. 240. doi: 10.1109/CECIT53797.2021.00048.

[33] W. Li, Y. Zhang, G. Wang, Y. Huang, and R. Li, "DFENet: A dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion," *Biomed. Signal Process. Control*, vol. 80, p. 104402, 2023, doi: 10.1016/j.bspc.2022.104402.

[34] X. Xie *et al.*, "Mrscfusion: Joint residual swin transformer and multiscale cnn for unsupervised multimodal medical image fusion," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–17, 2023, doi: 10.1109/TIM.2023.3317470.

[35] B. Zou *et al.*, "FocalNetFuse: enhancing multimodal image fusion quality with focal modulation networks: FocalNetFuse: enhancing multimodal image fusion quality…B. Zou et al.," *Vis. Comput.*, vol. 42, Jan. 2026, doi: 10.1007/s00371-025-04206-y.

[36] Z. Ding, H. Li, Y. Guo, D. Zhou, Y. Liu, and S. Xie, "M4FNet: Multimodal medical image fusion network via multi-receptive-field and multi-scale feature integration," *Comput. Biol. Med.*, vol. 159, p. 106923, Jun. 2023, doi: 10.1016/j.compbiomed.2023.106923.

[37] J. Di, W. Guo, J. Liu, L. Ren, and J. Lian, "AMMNet: A multimodal medical image fusion method based on an attention mechanism and MobileNetV3," *Biomed. Signal Process. Control*, vol. 96, p. 106561, Oct. 2024, doi: 10.1016/j.bspc.2024.106561.

[38] Y. Zhou, X. Yang, S. Liu, and J. Yin, "Multimodal Medical Image Fusion Network Based on Target Information Enhancement," *IEEE Access*, vol. 12, pp. 70851–70869, 2024, doi: 10.1109/ACCESS.2024.3402965.

[39] X. Sun, H. Liu, G. Chen, Y. Sheng, and C. Zhang, "Multimodal Medical Image Fusion via Manifold Structure Modeling and Information Geometry Enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, 2026, doi: 10.1109/TCSVT.2026.3661187.

[40] W. Li, P. Jia, D. He, S. Liu, G. Wang, and Y. Huang, "SAFusion: Scenario-Adaptive Network for Multimodal Medical Image Fusion," *IEEE J. Biomed. Health Inform.*, 2026, doi: 10.1109/JBHI.2026.3651957.

[41] A. A. Kamara, S. He, and A. J. Fofanah, "FAMAFuse: Functional-Anatomical Multiscale Attention for Multimodal Image Fusion," *IEEE Trans. Circuits Syst. Video Technol.*, 2025, doi: 10.1109/TCSVT.2025.3626562.

[42] J. Dhar *et al.*, "Multimodal Fusion Learning with Dual Attention for Medical Imaging," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Feb. 2025, pp. 4362–4371. doi: 10.1109/WACV61041.2025.00428.

[43] M. Safari, A. Fatemi, and L. Archambault, "MedFusionGAN: multimodal medical image fusion using an unsupervised deep generative adversarial network," *BMC Med. Imaging*, vol. 23, no. 1, p. 203, 2023, doi: 10.1186/s12880-023-01160-w.

[44] C. Fan, H. Lin, and Y. Qiu, "U-Patch GAN: A Medical Image Fusion Method Based on GAN," *J. Digit. Imaging*, vol. 36, no. 1, pp. 339–355, Feb. 2023, doi: 10.1007/s10278-022-00696-7.

[45] N. Anita, M. R. Devi, R. A. M. Rose, and J. S. J. Lijha, "MIMO-TGAN: Multi-Modality Medical Image Fusion via Triple Generator Network for Brain Abnormality Detection," *Int. J. Comput. Intell. Syst.*, Mar. 2026, doi: 10.1007/s44196-026-01189-z.

[46] C. Sui *et al.*, "IG-GAN: Interactive Guided Generative Adversarial Networks for Multimodal Image Fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–19, 2024, doi: 10.1109/TGRS.2024.3433619.

[47] K. Guo, X. Hu, and X. Li, "MMFGAN: A novel multimodal brain medical image fusion based on the improvement of generative adversarial network," *Multimed. Tools Appl.*, vol. 81, no. 4, pp. 5889–5927, Feb. 2022, doi: 10.1007/s11042-021-11822-y.

[48] Y. Zhou, K. He, D. Xu, J. Gong, and W. Mei, "DIFusion: Multimodal medical image fusion based on detail preservation and invertible neural networks," *Biomed. Signal Process. Control*, vol. 115, p. 109415, 2026, doi: 10.1016/j.bspc.2025.109415.

[49] G. C. Kumar, K. M, J. S, and N. S, "Structured constraints based Deep guided Generative adversarial network(GAN) for deformable multimodal medical image fusion(MMIF) and enhancement," in *2025 2nd International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS)*, Jul. 2025, pp. 1–5. doi: 10.1109/ICCAMS65118.2025.11234098.

[50] J. Wang, M. Liu, W. Shen, R. Ding, Y. Wang, and E. Meijering, "EPDiff: Erasure Perception Diffusion Model for Unsupervised Anomaly Detection in Preoperative Multimodal Images," *IEEE Trans. Med. Imaging*, vol. 45, no. 1, pp. 379–390, Jan. 2026, doi: 10.1109/TMI.2025.3597545.

[51] L. Han, "AD-Diff: enhancing Alzheimer's disease prediction accuracy through multimodal

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

fusion," *Front. Comput. Neurosci.*, vol. 19, Mar. 2025, doi: 10.3389/fncom.2025.1484540.

[52] A. Snani, M. Khadir, A. Pranolo, and M. Abdalla, "GAN-Enhanced multimodal fusion and ensemble learning for imbalanced chest X-Ray classification," *Int. J. Adv. Intell. Inform.*, vol. 11, p. 514, Aug. 2025, doi: 10.26555/ijain.v11i3.2092.

[53] T. Zhou, L. Liu, H. Lu, M. Zhang, and Z. Zhang, "MAC-GAN: Medical advice condition GAN for multimodal lung tumor image fusion," *Biomed. Signal Process. Control*, vol. 119, p. 109981, Jun. 2026, doi: 10.1016/j.bspc.2026.109981.

[54] Y. Kang *et al.*, "Deep learning-based multimodal fusion of MRI and whole slide image for predicting neoadjuvant therapy response in locally advanced head and neck squamous cell carcinoma," *BMC Med. Imaging*, 2026, doi: 10.1186/s12880-026-02173-x.

[55] Y. Akbari, F. Abdullakutty, S. Al-Maadeed, R. Al Saady, A. Bouridane, and R. Hamoudi, "A novel virtual patient approach for cross-patient multimodal fusion in enhanced breast cancer detection," *Comput. Med. Imaging Graph.*, vol. 127, p. 102687, 2026, doi: 10.1016/j.compmedimag.2025.102687.

[56] D. Duenias, B. Nichyporuk, T. Arbel, T. R. Raviv, and others, "Hyperfusion: A hypernetwork approach to multimodal integration of tabular and medical imaging data for predictive modeling," *Med. Image Anal.*, vol. 102, p. 103503, 2025, doi: 10.1016/j.media.2025.103503.

[57] C. Yu, J. Ye, Y. Liu, X. Zhang, and Z. Zhang, "AMF-MedIT: An efficient align-modulation-fusion framework for medical image–tabular data," *Biomed. Signal Process. Control*, vol. 118, p. 109772, 2026, doi: 10.1016/j.bspc.2026.109772.

[58] B. Zeng *et al.*, "C2HFusion: Clinical context-driven hierarchical fusion of multimodal data for personalized and quantitative prognostic assessment in pancreatic cancer," *Med. Image Anal.*, p. 103937, 2026, doi: 10.1016/j.media.2026.103937.

[59] Q. Ye, M. Luo, J. Zhou, C. Cheng, L. Peng, and J. Wu, "NMD-FusionNet: a multimodal fusion-based medical imaging-assisted diagnostic model for liver cancer," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 37, no. 6, p. 147, 2025, doi: 10.1007/s44443-025-00162-8.

[60] H. Xiang, H. Zhang, Y. Cheng, X. Quan, and W. Huang, "SMFusion: Semantic-Preserving Fusion of Multimodal Medical Images for Enhanced Clinical Diagnosis," *IEEE J. Biomed. Health Inform.*, 2025, doi: 10.1109/JBHI.2025.3649749.

[61] G. Dai *et al.*, "Prompt-Level Contrastive Learning for Context-Aware Multi-modal Image Representation in Medical Diagnosis," *Pattern Recognit.*, p. 113027, 2026, doi: 10.1016/j.patcog.2025.113027.

[62] R. S. Rao, A. Mishra, and S. Swain, "CAMF-SkinNet: Cross-Attention Multimodal Fusion of Visual, Textual, and Dermatology-Specific Embeddings for Skin Disease Classification," in *International Conference on Distributed Computing and Intelligent Technology*, Springer, 2026, pp. 425–435. doi: 10.1007/978-3-032-16632-6_27.

[63] W. Guo, L. Wang, J. Zeng, Q. Han, K. Jin, and X. Wang, "SMAFusion: Multimodal medical image fusion based on spatial registration and local-global multi-scale feature adaptive fusion," *Neurocomputing*, p. 131039, 2025, doi: 10.1016/j.neucom.2025.131039.

[64] M. A. Azam, K. B. Khan, M. Ahmad, and M. Mazzara, "Multimodal Medical Image Registration and Fusion for Quality Enhancement," *Comput. Mater. Contin.*, vol. 68, pp. 821–840, Feb. 2021, doi: 10.32604/cmc.2021.016131.

[65] Y. Wu, J. Chen, L. Hu, H. Xu, H. Liang, and J. Wu, "OmniFuse: A general modality fusion framework for multi-modality learning on low-quality medical data," *Inf. Fusion*, vol. 117, p. 102890, 2025, doi: 10.1016/j.inffus.2024.102890.

[66] D. M. Pathak *et al.*, "Optimal feature selection for medical image fusion using deep learning with transformer," *Biomed. Signal Process. Control*, vol. 111, p. 108377, 2026, doi: 10.1016/j.bspc.2025.108377.

[67] S. Sangeetha *et al.*, "An enhanced multimodal fusion deep learning neural network for lung cancer classification," *Syst. Soft Comput.*, vol. 6, p. 200068, 2024, doi: 10.1016/j.sasc.2023.200068.

[68] J. Cheng, F. Liu, and S. Wei, "Multimodal fusion network with multi-scale structure and metabolic focus for enhancing Alzheimer's disease prediction," *Appl. Intell.*, vol. 56, no. 2, p. 66, 2026, doi: 10.1007/s10489-026-07105-4.

[69] J. Xu, S. Zhuang, Y. He, H. Wang, Z. Zhuang, and H. Zeng, "Multimodal Sparse Fusion Transformer Network with Spatio-Temporal Decoupling for Breast Tumor Classification," *Med. Image Anal.*, p. 103966, 2026, doi: 10.1016/j.media.2026.103966.

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

[70] J. Huang *et al.*, "UltraMamba: Mamba-based Multimodal Ultrasound Image Adaptive Fusion for Breast Lesion Segmentation," *IEEE Trans. Med. Imaging*, 2026, doi: 10.1109/TMI.2026.3653779.

[71] L. Li *et al.*, "SymUnet-DynCFC: Multimodal MRI Fusion for Robust Cartilage Segmentation and Clinically Confirmed Moderate-to-Severe KOA Diagnosis," *Inf. Fusion*, p. 104145, 2026, doi: 10.1016/j.inffus.2026.104145.

[72] L. Dong *et al.*, "AI-based prediction of best-corrected visual acuity in patients with multiple retinal diseases using multimodal medical imaging," *Br. J. Ophthalmol.*, vol. 110, no. 2, pp. 158–165, 2026, doi: 10.1136/bjo-2025-327189s.

[73] J. Chen and J. Chen, "Multimodal image feature fusion for improving medical ultrasound image segmentation," *Biomed. Signal Process. Control*, vol. 89, p. 105705, 2024, doi: 10.1016/j.bspc.2023.105705.

[74] Q. Lu, L. Zheng, J. Su, W. Ma, H. Ma, and Y. Zhang, "MCAB-GFEResNet: A multimodal fusion model for pre-treatment prediction of neoadjuvant chemoradiotherapy response in rectal cancer," *Biomed. Signal Process. Control*, vol. 117, p. 109672, 2026, doi: 10.1016/j.bspc.2026.109672.

[75] P. Ravikumaran, K. Vimala Devi, and K. Valarmathi, "An Improved Kidney Tumor Prediction Using Deep Convolutional Neural Network-Restricted Boltzmann Machine Technique in Medical Image Segmentation," *J. Med. Imaging Health Inform.*, vol. 11, no. 12, pp. 3191–3198, Dec. 2021, doi: 10.1166/jmihi.2021.3917.

[76] H. Lu, M. Yu, X. Wei, X. Xu, and J. Xu, "Unbiased Multimodal Fusion for Medical Image Segmentation Based on Dual-Stream Adapter," *Knowl.-Based Syst.*, p. 114653, 2025, doi: 10.1016/j.knosys.2025.114653.

[77] M. Wang *et al.*, "Task-generalized adaptive cross-domain learning for multimodal image fusion," *IEEE Trans. Multimed.*, 2026, doi: 10.1109/TMM.2026.3660142.

[78] G. Zhang, R. Nie, J. Cao, L. Chen, and Y. Zhu, "FDGNet: A pair feature difference guided network for multimodal medical image fusion," *Biomed. Signal Process. Control*, vol. 81, p. 104545, Mar. 2023, doi: 10.1016/j.bspc.2022.104545.

[79] A. Bouamrane, M. Derdour, A. Bennour, A. Benmachiche, and M. Gasmi, "Machine Learning for Medical Image Analysis," in *AI for Medical Image Analysis: Reconciling Innovation and Ethical Considerations*, N. Ben Aoun, S. Ahmad, and M. Hammad, Eds., Cham: Springer Nature Switzerland, 2026, pp. 97–125. doi: 10.1007/978-3-032-02963-8_4.

[80] T. B. Nguyen-Tat, T. Q. Hung, P. T. Nam, and V. M. Ngo, "Evaluating pre-processing and deep learning methods in medical imaging: Combined effectiveness across multiple modalities," *Alex. Eng. J.*, vol. 119, pp. 558–586, Apr. 2025, doi: 10.1016/j.aej.2025.01.090.

[81] L. He, "Non-rigid Multi-Modal Medical Image Registration Based on Improved Maximum Mutual Information PV Image Interpolation Method," *Front. Public Health*, vol. 10, Jun. 2022, doi: 10.3389/fpubh.2022.863307.

[82] A. Khorasani, N. Dadashi serej, M. Jalilian, A. Shayganfar, and M. B. Tavakoli, "Performance comparison of different medical image fusion algorithms for clinical glioma grade classification with advanced magnetic resonance imaging (MRI)," *Sci. Rep.*, vol. 13, no. 1, p. 17646, Oct. 2023, doi: 10.1038/s41598-023-43874-5.

[83] T. Tirupal, B. Mohan, and S. Kumar, "Multimodal Medical Image Fusion Techniques – A Review," *Curr. Signal Transduct. Ther.*, vol. 15, Feb. 2020, doi: 10.2174/1574362415666200226103116.

[84] J. Chen, L. Chen, and M. Shabaz, "Image Fusion Algorithm at Pixel Level Based on Edge Detection," *J. Healthc. Eng.*, vol. 2021, p. 5760660, Aug. 2021, doi: 10.1155/2021/5760660.

[85] K. P. Indira, R. Rani Hemamalini, and R. Indhumathi, "Pixel based Medical Image Fusion Techniques using Discrete Wavelet Transform and Stationary Wavelet Transform," *Indian J. Sci. Technol.*, vol. 8, no. 26, Oct. 2015, doi: 10.17485/ijst/2015/v8i26/56192.

[86] B. Miles, M. W. K. Law, I. Ben-Ayed, G. Garvin, A. Fenster, and S. Li, "Pixel level image fusion for medical imaging: an energy minimizing approach," presented at the SPIE Medical Imaging, B. Van Ginneken and C. L. Novak, Eds., San Diego, California, USA, Feb. 2012, p. 831511. doi: 10.1117/12.911613.

[87] C. E. Ogbuanya, A. Obayi, S. Larabi-Marie-Sainte, A. O. Saad, and L. Berriche, "A hybrid optimization approach for accelerated multimodal medical image fusion," *PLOS One*, vol. 20, no. 7, p. e0324973, Jul. 2025, doi: 10.1371/journal.pone.0324973.

[88] S. Shehanaz, E. Daniel, S. R. Guntur, and S. Satrasupalli, "Optimum weighted multimodal medical image fusion using particle swarm

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

optimization," *Optik*, vol. 231, p. 166413, Apr. 2021, doi: 10.1016/j.ijleo.2021.166413.

[89] A. A. Alzahrani, "Enhanced multimodal medical image fusion via modified DWT with arithmetic optimization algorithm," *Sci. Rep.*, vol. 14, no. 1, p. 19261, Aug. 2024, doi: 10.1038/s41598-024-69997-x.

[90] J. Mi, L. Wang, Y. Liu, and J. Zhang, "KDE-GAN: A multimodal medical image-fusion model based on knowledge distillation and explainable AI modules," *Comput. Biol. Med.*, vol. 151, p. 106273, Dec. 2022, doi: 10.1016/j.compbiomed.2022.106273.

[91] M. Z. Khan *et al.*, "Multimodality medical image fusion using directional total variation based linear spectral clustering in NSCT domain," *Sci. Rep.*, vol. 16, no. 1, p. 5367, Feb. 2026, doi: 10.1038/s41598-025-26916-y.

[92] K. S. S. V. V. Ramesh and S. S. Kumar, "YUV-based SVD-VGG hybrid fusion for multimodal MRI-PET image integration," *PLOS ONE*, vol. 21, no. 1, p. e0340781, Jan. 2026, doi: 10.1371/journal.pone.0340781.

[93] D. C. Lepcha *et al.*, "Multimodal Medical Image Fusion based on Pixel Significance using Anisotropic Diffusion and Cross Bilateral Filter," *Hum.-Centric Comput. Inf. Sci.*, vol. 12, no. 0, pp. 190–206, Mar. 2022, doi: 10.22967/HCIS.2022.12.015.

[94] L. Wei, R. Zhu, X. Li, L. Zhao, X. Hu, and X. Zhang, "Pixel-level structure awareness for enhancing multi-modal medical image fusion," *Biomed. Signal Process. Control*, vol. 97, p. 106694, Nov. 2024, doi: 10.1016/j.bspc.2024.106694.

[95] P. Kavita, D. R. Alli, and A. B. Rao, "Study of image fusion optimization techniques for medical applications," *Int. J. Cogn. Comput. Eng.*, vol. 3, pp. 136–143, Jun. 2022, doi: 10.1016/j.ijcce.2022.05.002.

[96] T. Zhou, Q. Cheng, H. Lu, Q. Li, X. Zhang, and S. Qiu, "Deep learning methods for medical image fusion: A review," *Comput. Biol. Med.*, vol. 160, p. 106959, Jun. 2023, doi: 10.1016/j.compbiomed.2023.106959.

[97] N. Liang, "Medical image fusion with deep neural networks," *Sci. Rep.*, vol. 14, no. 1, p. 7972, Apr. 2024, doi: 10.1038/s41598-024-58665-9.

[98] F. Luo, D. Wu, L. R. Pino, and W. Ding, "A novel multimodel medical image fusion framework with edge enhancement and cross-scale transformer," *Sci. Rep.*, vol. 15, no. 1, p. 11657, Apr. 2025, doi: 10.1038/s41598-025-93616-y.

[99] J. Duan, S. Mao, J. Jin, Z. Zhou, L. Chen, and C. L. P. Chen, "A Novel GA-Based Optimized Approach for Regional Multimodal Medical Image Fusion With Superpixel Segmentation," *IEEE Access*, vol. 9, pp. 96353–96366, 2021, doi: 10.1109/ACCESS.2021.3094972.

[100] W. Tang, F. He, Y. Liu, and Y. Duan, "MATR: Multimodal Medical Image Fusion via Multiscale Adaptive Transformer," *IEEE Trans. Image Process.*, vol. 31, pp. 5134–5149, 2022, doi: 10.1109/TIP.2022.3193288.

[101] R. Prathipa and R. Ramadevi, "Feature Level Medical Image Fusion with Deep Learning," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, Oct. 2023, pp. 1–8. doi: 10.1109/EASCT59475.2023.10393457.

[102] L. Wang, J. Zhang, Y. Liu, J. Mi, and J. Zhang, "Multimodal Medical Image Fusion Based on Gabor Representation Combination of Multi-CNN and Fuzzy Neural Network," *IEEE Access*, vol. 9, pp. 67634–67647, 2021, doi: 10.1109/ACCESS.2021.3075953.

[103] N. Nagaraja Kumar, T. Jayachandra Prasad, and K. Satya Prasad, "Multimodal Medical Image Fusion with Improved Multi-Objective Meta-Heuristic Algorithm with Fuzzy Entropy," *J. Inf. Knowl. Manag.*, vol. 22, no. 1, p. 2250063, Feb. 2023, doi: 10.1142/S0219649222500630.

[104] Z. Wang, H. Di, R. Zhang, and F. Liu, "DTCFormer: Deep tensor chain frequency guided transformer for multi-modal medical image classification," *Biomed. Signal Process. Control*, vol. 113, p. 109152, Mar. 2026, doi: 10.1016/j.bspc.2025.109152.

[105] L. Huang, S. Ruan, P. Decazes, and T. Denœux, "Deep evidential fusion with uncertainty quantification and reliability learning for multimodal medical image segmentation," *Inf. Fusion*, vol. 113, p. 102648, Jan. 2025, doi: 10.1016/j.inffus.2024.102648.

[106] S. Guo, L. Wang, Q. Chen, L. Wang, J. Zhang, and Y. Zhu, "Multimodal MRI Image Decision Fusion-Based Network for Glioma Classification," *Front. Oncol.*, vol. 12, Feb. 2022, doi: 10.3389/fonc.2022.819673.

[107] S. Siddiqui *et al.*, "Intelligent Breast Cancer Prediction Empowered with Fusion and Deep Learning," *Comput. Mater. Contin.*, vol. 67, no. 1, pp. 1033–1049, 2021, doi: 10.32604/cmc.2021.013952.

[108] S. Roheda, H. Krim, Z.-Q. Luo, and T. Wu, "Decision Level Fusion: An Event Driven Approach," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018,

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.

pp. 2598–2602. doi: 10.23919/EUSIPCO.2018.8553412.
[109] P. Szczuko, A. Harasimiuk, and A. Czyżewski, "Evaluation of Decision Fusion Methods for Multimodal Biometrics in the Banking Application," *Sensors*, vol. 22, no. 6, p. 2356, Mar. 2022, doi: 10.3390/s22062356.
[110] V. Sireesha and K. Sandhyarani, "Overview of Fusion Techniques in Multimodal Biometrics," *Int. J. Eng. Res.*, 2014.
[111] N. A. Othman, M. A. Abdel-Fattah, and A. T. Ali, "A Hybrid Deep Learning Framework with Decision-Level Fusion for Breast Cancer Survival Prediction," *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 50, Mar. 2023, doi: 10.3390/bdcc7010050.
[112] A. Karthik *et al.*, "Ensemble-based multimodal medical imaging fusion for tumor segmentation," *Biomed. Signal Process. Control*, vol. 96, p. 106550, Oct. 2024, doi: 10.1016/j.bspc.2024.106550.
[113] Z. Liu, J. Jia, F. Bai, Y. Ding, L. Han, and G. Bai, "Predicting rectal cancer tumor budding grading based on MRI and CT with multimodal deep transfer learning: A dual-center study," *Heliyon*, vol. 10, no. 7, Apr. 2024, doi: 10.1016/j.heliyon.2024.e28769.

Author Biography

Majda Maatallah is an Associate Professor in Computer Science at Chadli Bendjedid University in El Tarf, Algeria. She specializes in artificial intelligence, recommender systems, and technology-enhanced learning, with a strong focus on designing intelligent solutions for modern educational environments. Her research interests include intelligent systems, collaborative filtering techniques, and the analysis of learning behaviors in MOOCs. In addition to her research activities, she is actively involved in teaching at the master's level, where she supervises graduate students and delivers courses in artificial intelligence, reasoning, and related fields. Dr. Maatallah contributes to various academic and scientific activities, including participation in international conferences and collaborative research projects, supporting the continuous advancement of AI-driven educational systems and innovative learning technologies.

Abdelmadjid Benmachiche is a Professor of Computer Science at Chadli Bendjedid University in El Tarf, Algeria, and serves as the Director of the House of Artificial Intelligence. His research interests cover a wide range of domains, including artificial intelligence, cybersecurity, e-learning systems, robotics, medical imaging, and IoT. Over the years, he has made substantial contributions to deep learning, recommendation systems, autonomous navigation, and intelligent educational platforms. His work emphasizes developing innovative, real-world AI-driven solutions that address complex interdisciplinary challenges. Professor Benmachiche has authored and co-authored numerous scientific publications in reputable peer-reviewed journals and international conferences. In addition to his research output, he actively leads and participates in national and international research projects and collaborations. Through these efforts, he continues to advance knowledge, foster innovation, and support the integration of artificial intelligence technologies.

Khadija Rais is a researcher in Artificial Intelligence and Medical Imaging at the Laboratory of Mathematics, Informatics and Systems (LAMIS), Echahid Cheikh Larbi Tebessi University, Algeria. She holds a PhD in Artificial Intelligence and a master's degree in Multimedia and Systems. Her research focuses on deep learning for medical imaging, with expertise in generative models, data augmentation, and intelligent healthcare systems. She has also contributed to interdisciplinary areas, including cybersecurity and e-learning systems. She has authored and co-authored numerous publications in peer-reviewed journals and at national and international conferences, reflecting her active engagement in advancing applied AI research and developing innovative solutions.

Selma Touam is a professor in condensed matter physics at Chadli Bendjedid University in El Tarf, Algeria. She holds a PhD in Physics from Badji Mokhtar University, Annaba. Her teaching activities include courses in electromagnetism, materials physics, and related subjects. Her research focuses on materials science, particularly ab initio calculations based on density functional theory (DFT), with an emphasis on investigating the electronic, optical, and structural properties of semiconductors. She has contributed to several scientific publications in peer-reviewed journals and has participated in national and international conferences. In addition, she actively supervises student research and contributes to various academic and scientific activities.

Manuscript received 3 February 2026; Revised 10 April 2026; Accepted 5 May 2026; Available online 18 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1527

Note: the original version contained 24 self-citations and that these citations have since been removed.