An Improved Medical Image Fusion for Efficient Diagnostic Imaging Using Deep Learning and Gradient-Based Quality Assessment

Walid El-Shafai
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia | Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menoufia 32952, Egypt
welshafai@psu.edu.sa

Ahmad Taher Azar
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
aazar@psu.edu.sa (corresponding author)

Received: 18 February 2025 | Revised: 4 May 2025 and 25 May 2025 | Accepted: 30 May 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.10622*

ABSTRACT

Advancements in Deep Learning (DL) have significantly contributed to progress in image processing and pattern recognition, particularly in the medical imaging domain. This study proposes an improved DL-based fusion method for multi-modal medical images, referred to as Deep Learning Medical Image Fusion (DMF). The proposed technique integrates Convolutional Neural Networks (CNNs) with three conventional image fusion methods to generate enhanced fused outputs across various medical imaging modalities. By leveraging gradient-based quality assessments, including visibility, consistency, and exposure mapping, the DMF method effectively synthesizes multi-exposure images, resulting in high-resolution outputs with improved structural clarity and visual consistency. Unlike traditional fusion approaches, the proposed model capitalizes on gradient variations among multiple exposures to guide fusion decisions, leading to improved diagnostic relevance. The performance of the DMF technique is validated through comprehensive experiments, involving Magnetic Resonance Imaging-Computed Tomography (MRI-CT) and MRI-Positron Emission Tomography (MRI-PET) image pairs, which demonstrated its superior effectiveness and practicality in healthcare imaging and diagnostic applications.

Keywords-medical image fusion; deep learning; multi-modal imaging; convolutional neural networks; gradient-based quality assessment; image processing; diagnostic imaging; healthcare applications

I. INTRODUCTION

Image fusion technology enables the combination of multiple source images to produce a composite output with enhanced visual and informational quality, often surpassing that of any individual input. This process plays a crucial role in improving the interpretability and reliability of medical images. In clinical practice, multi-modal medical image fusion has become an essential tool for accurate diagnosis and treatment planning [1, 2]. Medical images are generally categorized into two primary types: anatomical images, which capture structural information, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), and functional images, which reveal physiological activity, such as Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT). These complementary modalities provide distinct yet interdependent insights, allowing clinicians to visualize the spatial localization of organs or lesions from multiple perspectives. Among the most widely used anatomical imaging modalities are MRI and CT, both of which offer high spatial resolution and detailed visualization of internal anatomical structures. In contrast, functional imaging typically includes SPECT and PET, which, despite their lower spatial resolution, provide critical information about physiological processes such as tissue metabolism and blood flow [3]. The fusion of anatomical and functional images enables clinicians to address complex diagnostic challenges by combining structural and functional perspectives into a single, comprehensive representation [4].

Beyond traditional medical imaging modalities, biomedical imaging also benefits significantly from fusion techniques. Two commonly used biological imaging approaches are Green Fluorescent Protein (GFP) imaging and Phase Contrast (PC) microscopy. GFP allows for the visualization of protein distribution within cells, while PC imaging enhances the structural details of transparent specimens, including subcellular organelles such as mitochondria and nuclei. The fusion of PC and GFP images serves as a powerful tool for investigating protein function, enabling researchers to correlate structural and molecular information within complex biological systems [5].

Image fusion techniques are commonly categorized into three levels: pixel-level, feature-level, and decision-level fusion strategies [6]. Among these, pixel-level fusion directly combines input images at the pixel level, offering superior data preservation and producing highly detailed fused outputs. Feature-level fusion involves the integration of extracted features, such as edges, textures, and structural patterns, from the source images into a unified representation. While this approach is generally less sensitive to precise image registration, it may result in the loss of fine-grained details. At the highest level of abstraction, decision-level fusion combines the outputs of individual image analyses or classifiers, often at the cost of significant loss of spatial and structural information [7]. Overall, pixel-level fusion remains the most straightforward and computationally efficient approach, making it a widely adopted technique in medical imaging applications.

Pixel-level image fusion techniques are generally divided into two categories: spatial domain methods and transform domain methods [8, 9]. Spatial domain fusion operates directly on the pixel intensities of source images [10, 11], whereas transform domain fusion begins by decomposing the source images into low- and high-frequency components using multi-scale transformation techniques. Fusion rules are then applied to combine these components, followed by an inverse transformation to reconstruct the final fused image. Within the transform domain, various techniques have been proposed [12-15], including methods based on wavelet transforms [13], Non-Subsampled Contourlet Transforms (NSCT) [16], and Non-Subsampled Shearlet Transforms (NSST) [17]. These approaches have shown significant success in enhancing fusion performance by preserving structural and spectral information.

Despite their effectiveness, conventional transform domain fusion methods often require complex frequency decomposition and reconstruction steps, which can sometimes lead to artifacts, structural distortion, or loss of critical information from the source images. To address these limitations, DL-based fusion techniques have gained increasing attention in the medical imaging community [18, 19, 21-23]. However, DL-based diagnostic systems typically require large volumes of annotated training data, often real-world clinical images, which can be difficult to acquire due to privacy constraints, data scarcity, and the high cost of expert annotation in medical contexts [20]. Authors in [21] proposed a multi-focus image fusion method based on residual neural networks, marking an advancement in DL-driven fusion. Building on this progress, DL has been increasingly applied to multi-modal medical image fusion. In [22], a CNN-based method was introduced, employing a Siamese architecture to generate weight maps for fusing contrast pyramid representations of the source images. These DL-powered techniques unify the processes of fusion rule formulation and decision map generation, effectively streamlining the overall fusion pipeline [23]. However, the success of DL-based medical image fusion depends heavily on access to large, high-quality training datasets. Ideally, these datasets should be composed of real-world clinical images; yet, acquiring such data remains a major challenge due to privacy concerns, limited availability, and annotation costs [24].

The primary contributions of this paper are:

- The study proposes an improved DMF technique that integrates three conventionally fused medical images as inputs to a CNN model. The method supports simultaneous computation of Visibility Assessment (VA), consistency assessment, and exposure mask generation, enabling the effective merging of outputs from multiple traditional fusion techniques to enhance diagnostic accuracy.
- The fusion of MRI-CT and MRI-PET image pairs is conducted using combinations of eight well-established traditional methods: Anisotropic Diffusion Fusion (ADF) [25], Cross Bilateral Filter (CBF) [26], Fourth-Order Partial Differential Equations (FPDE) [27], Medical Feature Extraction and Visual Information Preservation (MFEVIP) [28], Guided Filter-based Context Enhancement (GFCE) [29], Weighted Least Squares (WLS) [30], Two-Scale Image Fusion (TIF) [31], and Latent Low-Rank Representation (LatLRR) [32]. These techniques generate the three input images for CNN-based feature extraction, which in turn facilitates adaptive exposure mask computation and gradient-guided fusion.
- To address issues, such as contrast loss and image noise, the proposed method employs a luminance-aware fusion weight adjustment strategy, which enhances the preservation of key diagnostic features in the resulting fused images.
- A comprehensive quantitative and qualitative evaluation is conducted using publicly available datasets comprising 40 MRI-CT image pairs and 20 MRI-PET image pairs. The comparative analysis with state-of-the-art fusion methods demonstrates the superior performance of the proposed DMF technique across multiple fusion quality metrics.

II. PROPOSED DEEP LEARNING-BASED MEDICAL IMAGE ALGORITHM

The fusion of multiple medical images can produce a more informative and diagnostically relevant representation than any individual input. This is particularly beneficial given the complementary nature of different medical imaging modalities, each of which captures unique anatomical or functional characteristics. The image fusion process employed in this study is formulated in (1), where the goal is to synthesize a single, high-quality image from multiple source inputs:

$$F(h,v) = \sum_{j=1}^{N} W^{j}(h,v) I^{fj}(h,v) \quad (1)$$

where N denotes the number of input images (exposures), $I^{fj}(h,v)$ represents the pixel intensity at location (h,v) in the j^{th} input image, and $W^{j}(h,v)$ is the corresponding fusion weight. The fused output is denoted by $F(h, v)$.

This formulation, based on exposure composition, provides a streamlined and computationally efficient framework for multi-modal image fusion, offering advantages over many conventional methods. However, the effectiveness of this fusion approach is highly dependent on the computation of the weight parameters. $W^{j}(h,v)$, which must be accurately derived to preserve diagnostic information. To address this issue, a gradient-based quality assessment strategy was introduced that enables the precise estimation of fusion weights, ultimately improving the visibility, consistency, and exposure control across a variety of medical imaging modalities.

As illustrated in Figure 1, the proposed framework incorporates three established fusion methodologies to combine two input medical images. These fused outputs are then used as inputs to a CNN for feature extraction. A gradient-based quality assessment strategy is subsequently applied to compute the weight maps corresponding to each fused image. This assessment involves the calculation of three key components: visibility, consistency, and the exposure mask. These components collectively guide the generation of pixel-wise weights, which are used to construct the final fused image. The fusion process is completed using a multi-resolution spline interpolation technique, allowing for the seamless blending of all input exposures in accordance with their corresponding weight maps [22].

In deep CNN architectures, feature extraction is typically performed using either regression networks or classification networks [33]. In the context of medical image fusion, where accurate pixel-wise weight estimation is critical, the extraction of dense features becomes essential. Regression-based networks, such as those used for image denoising or super-resolution, are particularly effective for this task. On the other hand, classification networks, like VGG-Net [34, 35], can also capture dense features through their Convolutional Layers (CLs). However, these features are often influenced by the deeper layers of the network, which introduce sparsity and emphasize high-level semantic representations. The proposed method leverages both low- and mid-level dense features to achieve precise and adaptive weight generation for the fusion process.

Feature maps are extracted from each of the three traditionally fused input images using a CNN, as expressed in:

$$FM_{j}(h,v) = \mathrm{CNN}(I^{fj})(h,v) \quad (2)$$

where CNN(.) refers to a specific layer within a pre-trained deep network. For each pixel location (h, v), the output is a feature vector whose dimensionality corresponds to the number of filters in the selected CL.

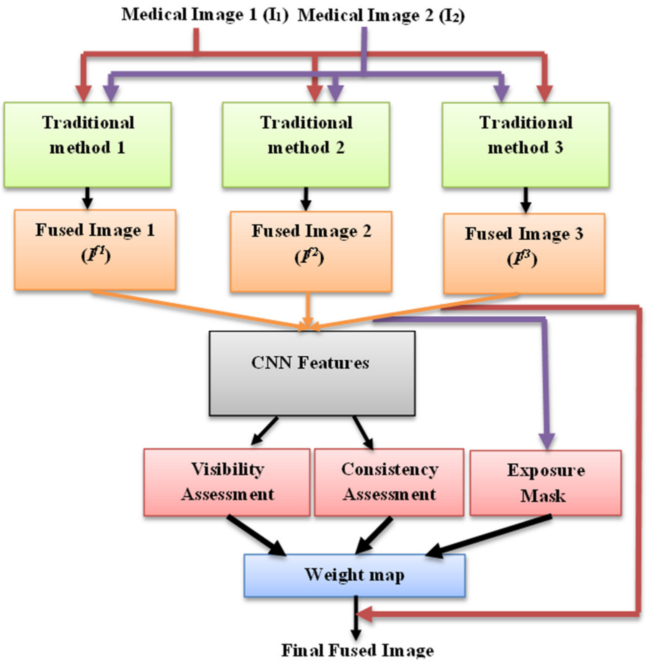


Fig. 1. Architecture of the proposed DL-based medical image fusion framework.

To compute the weighting maps, gradient information is leveraged, which is particularly effective for capturing local image structure and detail. The gradient magnitudes are obtained using the first-order derivatives of a 2D Gaussian kernel, a standard technique for edge and contrast detection. For each fused input image $I^{fj}(h,v)$, the gradient magnitude and gradient orientation at pixel (h, v) are denoted by $M^{j}(h,v)$ and $\emptyset^{j}(h,v$, respectively. These gradient-based features form the foundation for calculating visibility, consistency, and exposure assessments in the subsequent fusion process.

In multi-exposure image fusion, overexposure or underexposure in individual inputs can lead to the loss of important structural details that are otherwise visible in other exposures. The primary objective of the fusion process in such cases is to preserve all critical features across the input images and render them clearly in a single composite output.

The gradient magnitude of a pixel, which reflects local image structure, serves as a reliable indicator of exposure quality. Well-exposed pixels typically exhibit higher gradient magnitudes, while those affected by overexposure or underexposure tend to have reduced gradient values. This observation forms the basis for defining a VA metric, as formulated in:

$$VA^{j}(h,v) = \frac{M^{j}(h,v)}{\sum_{j=1}^{N} M^{j}(h,v)+\varepsilon} \quad (3)$$

where $M^{j}(h,v)$ denotes the gradient magnitude at pixel (h, v) in the j^{th} input image, and ε is a small constant (typically in the range of 10-25) introduced to prevent division by zero. This gradient-based visibility metric effectively guides the weight assignment process, ensuring that well-exposed pixels contribute more significantly to the fused result. In the context of medical image fusion, this strategy supports coherent image composition and the preservation of diagnostically relevant information.

To improve robustness against noise and enhance stability in weight computation, a window-based strategy was employed to evaluate local variations in gradient direction. For each pixel location (*h*, *v*) in the j^{th} input image, the gradient direction difference with respect to the k^{th} image is computed using:

$$D^{jk}(h,v) = \frac{\sum_{i=-l}^{l}\left|\emptyset^{j}(h+i,v+i)-\emptyset^{k}(h+i,v+i)\right|}{(2l+1)^{2}} \tag{4}$$

where $\emptyset^{j}(h+i,v+i)$ and $\emptyset^{k}(h+i,v+i)$ represent the gradient orientations at the pixel $(h+i,v+i)$ in the j^{th} and k^{th} images, respectively. The computation is performed over a local window of size $(2l+1)\times(2l+1)$, where l is typically set to 9. The dissimilarity measure is symmetric, i.e., $D^{jk}(h,v) = D^{kj}(h,v)$, and for identical images ($j = k$), the value is defined as zero, i.e., $D^{jj}(h,v) = 0$.

In many medical imaging sequences, it is often observed that disease-affected regions consistently appear across different exposures. This consistency arises because clinicians tend to focus on pathological areas during image acquisition, ensuring their prominence throughout the image set. To quantify this temporal stability, a consistency score $SC^{j}(h,v)$ is computed by aggregating gradient direction variations across all exposures, as defined in:

$$SC^{j}(h,v) = \sum_{k=1}^{N}\exp\left(\frac{-D^{jk}(h,v)^{2}}{2\sigma_{s}^{2}}\right) \tag{5}$$

where σ_s represents the standard deviation of the Gaussian kernel, which is empirically set to 0.2. A higher consistency score indicates minimal variation in gradient direction at pixel location (*h*, *v*), suggesting that the corresponding structure is stable and consistently present across exposures. This aligns with the assumption that medical image sequences tend to capture static anatomical elements, particularly those relevant to diagnosis [22].

However, changes in gradient direction may also result from exposure artifacts, such as overexposure or underexposure, potentially leading to misinterpretation. In such cases, relying solely on gradient-based dissimilarity $D^{jk}(h,v)$ may cause the fusion algorithm to incorrectly treat dynamic or diagnostic regions as noise or instability. To address this, an exposure quality parameter ($\delta^{j}(h,v)$) was introduced, which identifies well-exposed regions based on pixel intensity. This leads to the definition of the exposure-weighted consistency assessment, $CA^{j}(h,v)$, as given in:

$$CA^{j}(h,v) = \frac{SC^{j}(h,v)\,.\,\delta^{j}(h,v)}{\sum_{j=1}^{N}SC^{j}(h,v)\,.\delta^{j}(h,v)+\varepsilon} \tag{6}$$

The binary exposure indicator $\delta^{j}(h,v)$ is defined as:

$$\delta^{j}(h,v) = \begin{cases} 1 & 1-\tau < I^{fj}(h,v) < \tau \\ 0 & otherwise \end{cases} \tag{7}$$

where τ denotes the upper threshold for identifying well-exposed pixels, and is commonly set to 0.9. This formulation ensures that erroneous consistency scores from overexposed or underexposed regions are effectively suppressed.

Finally, the fusion weight $W^{j}(h,v)$ is computed by combining the VA metric ($VA^{j}(h,v)$) and the adjusted consistency assessment, $CA^{j}(h,v)$, as defined in:

$$W^{j}(h,v) = \frac{VA^{j}(h,v)\,.CA^{j}(h,v)}{\sum_{j=1}^{N}VA^{j}(h,v)\,.CA^{j}(h,v)+\varepsilon} \tag{8}$$

This final weighting mechanism ensures that the fused image benefits from both exposure-based visibility and structural consistency, enhancing the retention of diagnostically significant regions while mitigating the impact of exposure-related artifacts.

III. EXPERIMENTAL EVALUATION AND ANALYSIS

To evaluate the effectiveness of the proposed DMF technique, experiments were conducted using two publicly available multimodal medical imaging datasets. The first dataset comprises MRI and CT image pairs from the RIDER Neuro MRI-CT collection, hosted by The Cancer Imaging Archive (TCIA) [36]. The second dataset includes MRI and PET images obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [37]. All image pairs were pre-registered using mutual information-based rigid alignment and resized to a standardized resolution of 256×256 pixels. Intensity normalization was applied to ensure consistency across modalities. These datasets were selected based on their clinical relevance, accessibility, and suitability for research reproducibility.

The image fusion pipeline was implemented in Python 3.8 using the TensorFlow and Keras DL frameworks. All experiments were executed on a workstation equipped with an Intel Core i7-12700K CPU @ 3.60 GHz, 32 GB RAM, and an NVIDIA RTX 3080 GPU (10 GB VRAM), running Windows 11 with CUDA and cuDNN support for GPU acceleration. The framework was optimized for parallel processing to ensure efficient computation during training and inference.

An evaluation was conducted using diverse image combinations involving CT, MRI, and PET modalities. Representative examples of the image pairs used are shown in Figures 2 and 3. The fusion performance was quantitatively assessed utilizing a suite of widely recognized image quality metrics, including: Average Gradient (*AG*) [3], Edge Intensity (*EI*) [6], Entropy (*EN*) [8], gradient-based fusion metric ($Q^{(AB/F)}$) [15], Chen-Varshney metric (Q_{CV}) [17], Standard Deviation (*SD*) [24], and Spatial Frequency (*SF*) [33]. These metrics collectively offer a robust and multi-dimensional evaluation framework, assessing contrast, sharpness, spatial detail, and fusion fidelity.

In the context of multi-modal fusion, MRI and CT images are commonly used due to their complementary anatomical information. MRI provides high-resolution visualization of soft tissue structures, while CT imaging excels in capturing bone and dense tissue. Additionally, PET images offer valuable metabolic and functional insights, making them an important component in tri-modal fusion scenarios. The experiments utilized various combinations of these modalities to ensure a comprehensive evaluation from both anatomical and functional imaging perspectives.

All medical images were sourced from curated and validated clinical repositories, and were carefully registered before fusion. The inclusion of both traditional and DL-based fusion inputs enables rigorous analysis of the proposed method's generalization across modalities. Figures 2 and 3 illustrate samples of the input image datasets used in this study.

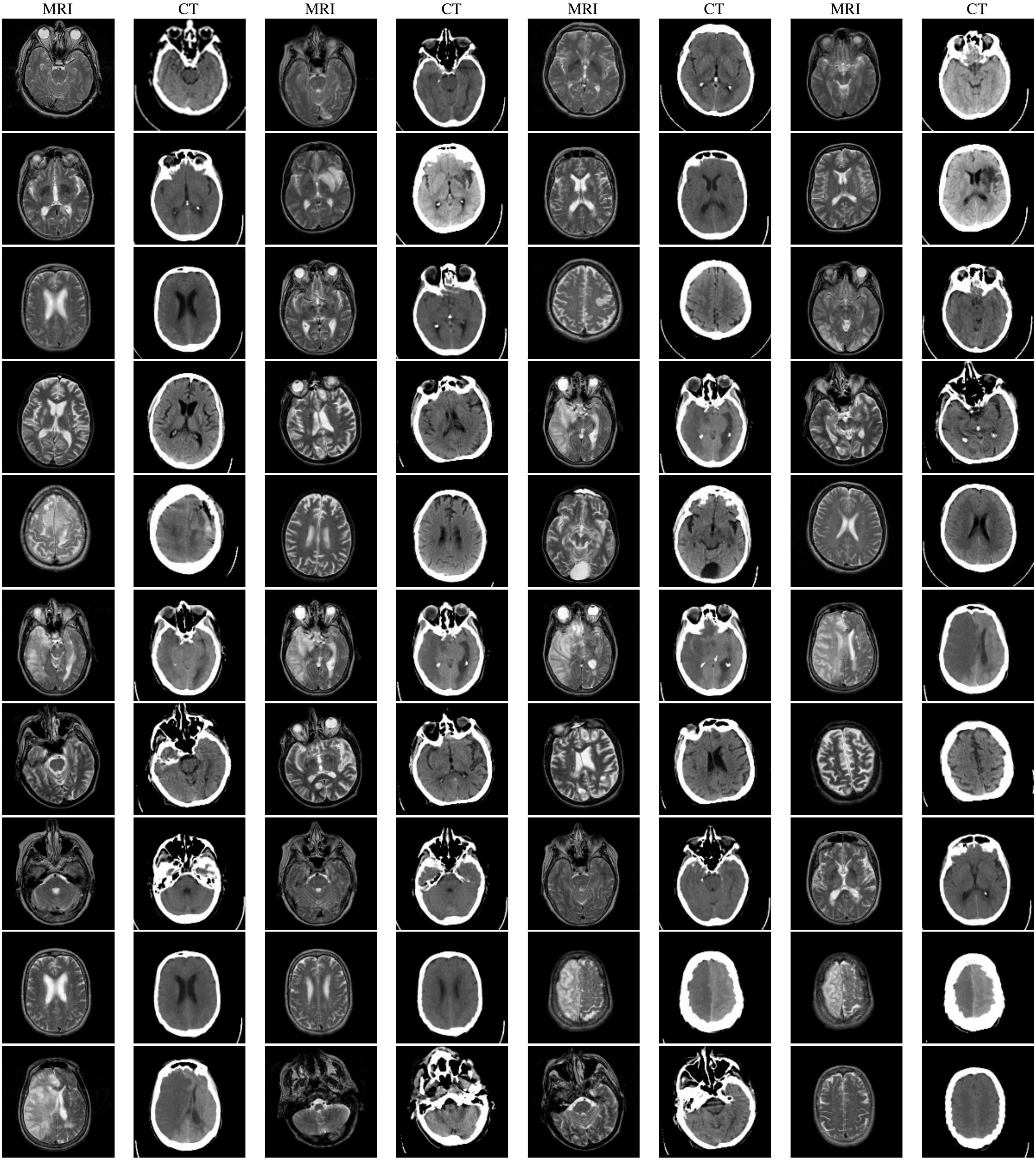


Fig. 2. Sample input images from the dataset: 40 pairs of MRI and CT medical images used for fusion evaluation.

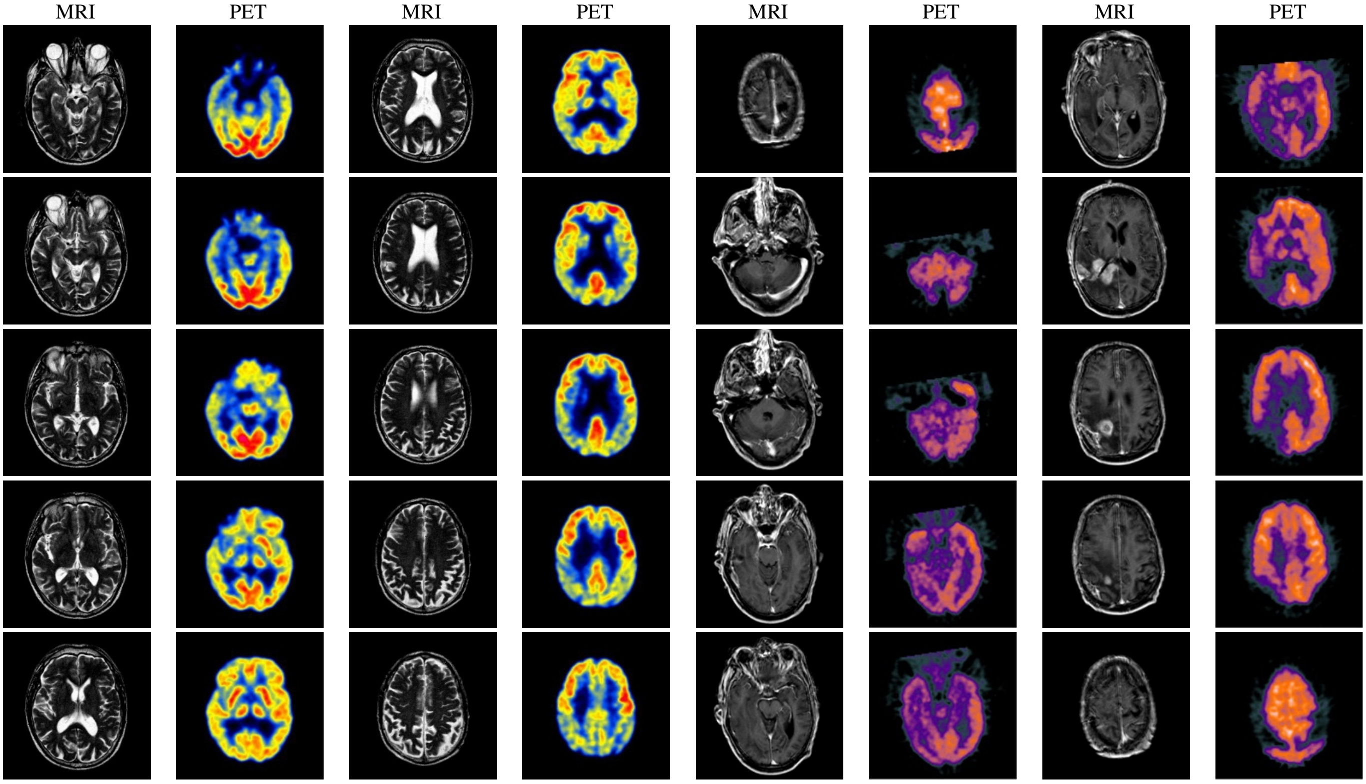


Fig. 3. Sample input images from the dataset: 20 pairs of MRI and PET medical images used for fusion evaluation.

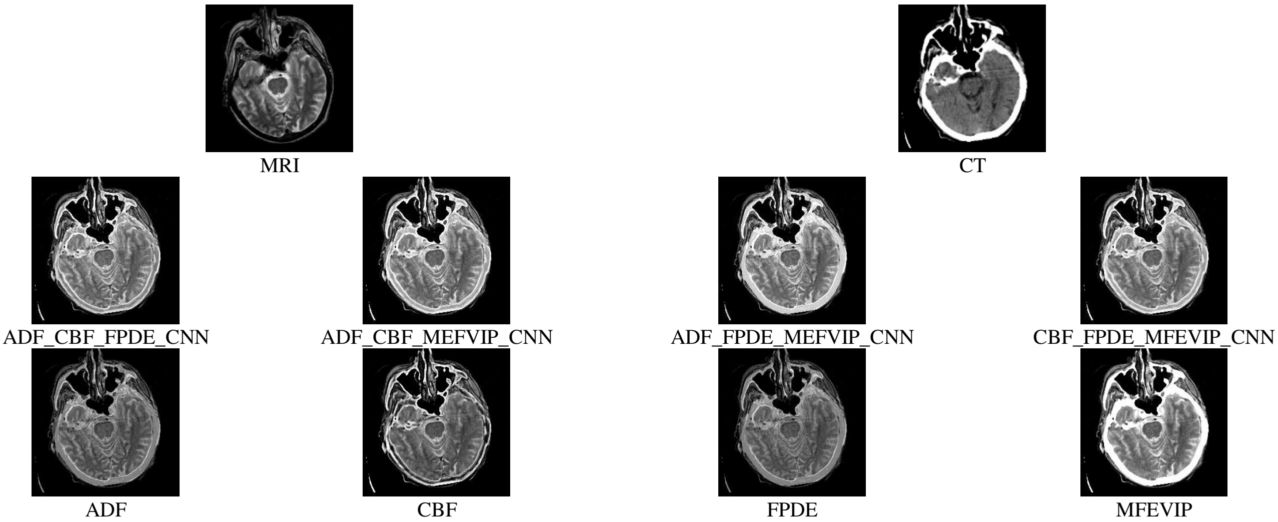
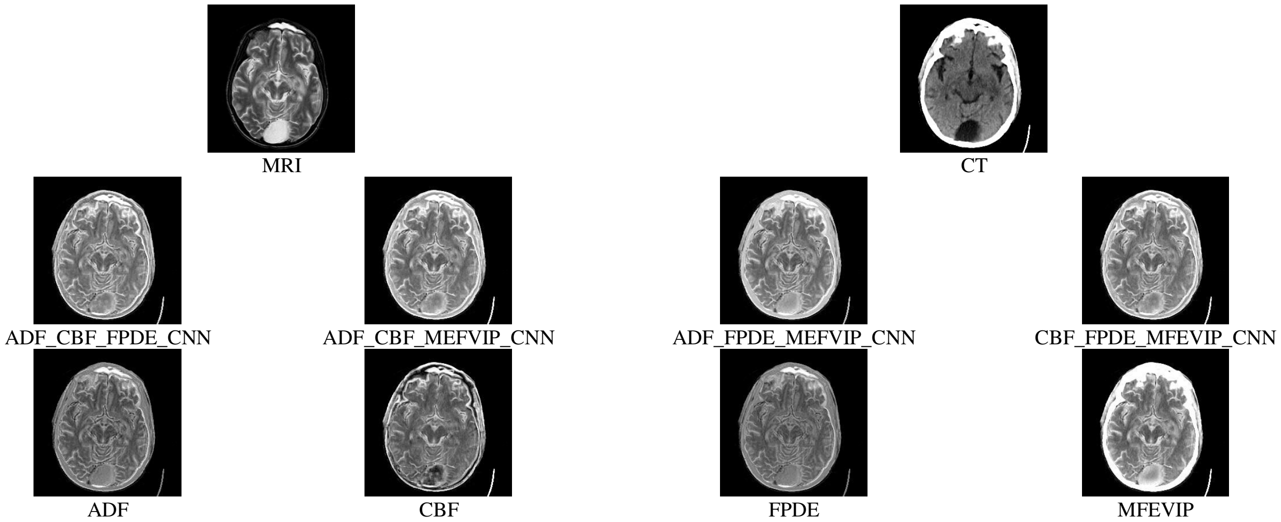

MRI
CT
ADF_CBF_FPDE_CNN
ADF_CBF_MEFVIP_CNN
ADF_FPDE_MEFVIP_CNN
CBF_FPDE_MFEVIP_CNN
ADF
CBF
FPDE
MFEVIP

Fig. 4. Example fusion results using different algorithms for one MRI-CT image pair (from the 40-pair dataset).

PET images, which contribute valuable data to the fusion process, are among the common image types employed in CT imaging technologies. To facilitate experiments, an assortment of traditional MRI, CT, and PET images was amalgamated/combined with DL images. Before experimenting, all images were meticulously registered and sourced from well-established medical image databases. To provide a visual representation, Figures 2 and 3 showcase select subsets of image databases employed in this paper.

To highlight the novelty and effectiveness of the proposed DMF technique, a comparative analysis is presented in Figures 4–7 and Tables I and II. The DMF method was evaluated against several recent traditional and deep learning-enhanced fusion techniques, including ADF, CBF, FPDE, MFEVIP, TIF, WLS, GFCE, LatLRR, and their hybrid CNN-based combinations (e.g., ADF_CBF_FPDE_CNN, TIF_WLS_LatLRR_CNN).



MRI CT

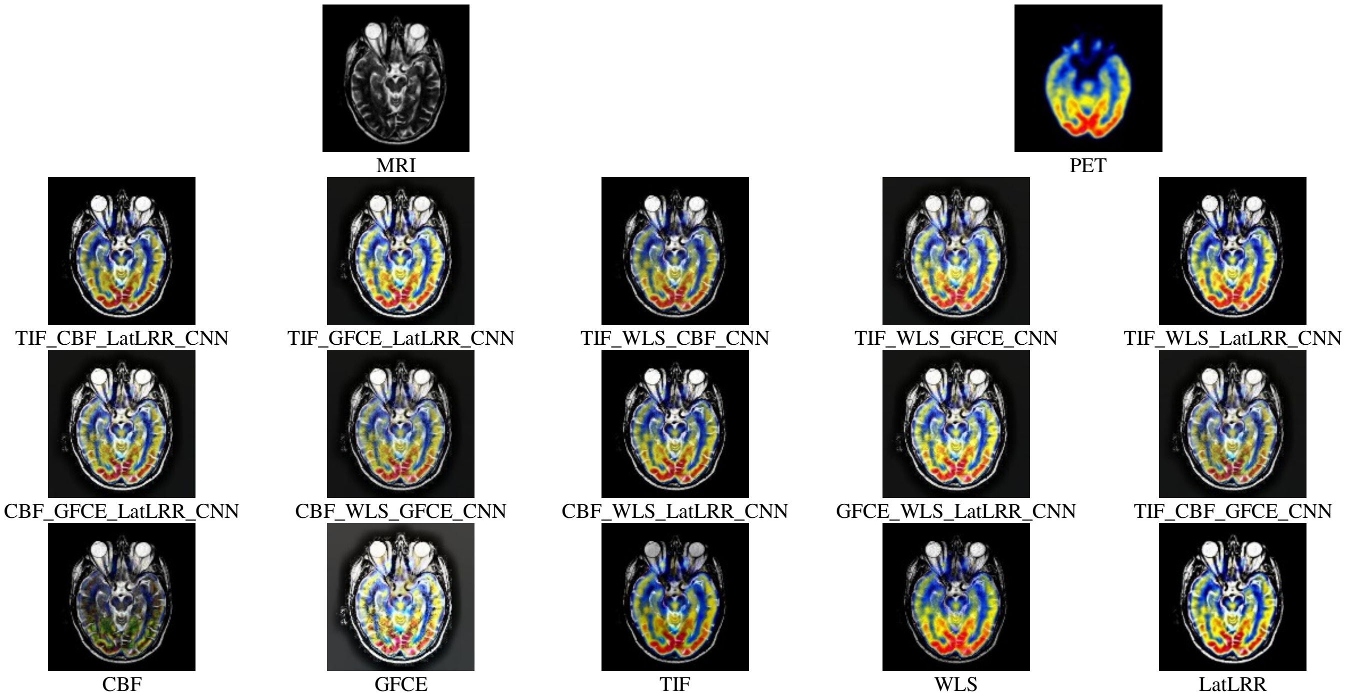
ADF_CBF_FPDE_CNN ADF_CBF_MEFVIP_CNN ADF_FPDE_MEFVIP_CNN CBF_FPDE_MFEVIP_CNN

ADF CBF FPDE MFEVIP

Fig. 5. Fusion results for the 40 MRI-CT image pairs using different rule-based fusion strategies within the evaluated algorithms.

TABLE I. QUANTITATIVE PERFORMANCE METRICS AVERAGED OVER 40 MRI-CT IMAGE PAIRS FOR EACH EXAMINED FUSION METHOD

Algorithm	***EI***	***AG***	$Q^{(AB/F)}$	***SF***	***EN***	***SD***	Q_{CV}
CBF_FPDE_MFEVIP_CNN	78.12792	7.731971	0.552564	30.24456	5.017528	81.89898	1173.891
ADF_FPDE_MEFVIP_CNN	71.69639	7.139887	0.520176	26.61368	5.011709	83.37993	1165.465
ADF_CBF_MEFVIP_CNN	79.20299	7.847331	0.556246	30.59601	5.013677	81.63195	1202.918
ADF_CBF_FPDE_CNN	80.78637	8.026103	0.553184	29.97839	4.986793	78.03174	1368.798
MFEVIP	72.35106	7.126903	0.512184	29.68373	4.748961	91.41878	1130.51
FPDE	55.00678	5.567731	0.394177	19.89671	4.746449	59.92041	1865.207
CBF	86.21341	8.493344	0.53812	33.4717	4.879071	64.98	2368.431
ADF	59.72915	6.091608	0.431931	22.16841	4.720815	60.23859	1993.145



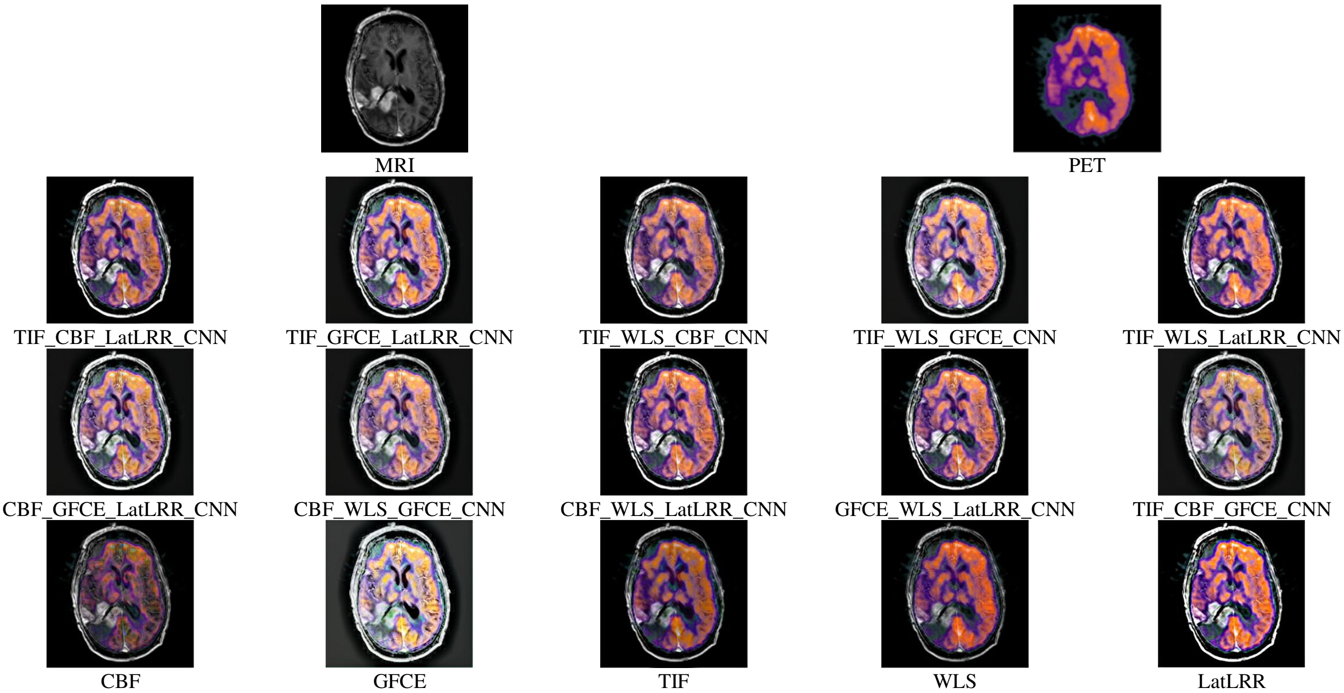
MRI PET

TIF_CBF_LatLRR_CNN TIF_GFCE_LatLRR_CNN TIF_WLS_CBF_CNN TIF_WLS_GFCE_CNN TIF_WLS_LatLRR_CNN

CBF_GFCE_LatLRR_CNN CBF_WLS_GFCE_CNN CBF_WLS_LatLRR_CNN GFCE_WLS_LatLRR_CNN TIF_CBF_GFCE_CNN

CBF GFCE TIF WLS LatLRR

Fig. 6. First example of fusion results using different algorithms on one MRI-PET image pair (from the 20-pair dataset).



MRI PET
TIF_CBF_LatLRR_CNN TIF_GFCE_LatLRR_CNN TIF_WLS_CBF_CNN TIF_WLS_GFCE_CNN TIF_WLS_LatLRR_CNN
CBF_GFCE_LatLRR_CNN CBF_WLS_GFCE_CNN CBF_WLS_LatLRR_CNN GFCE_WLS_LatLRR_CNN TIF_CBF_GFCE_CNN
CBF GFCE TIF WLS LatLRR

Fig. 7. Second example of fusion results using different algorithms on a different MRI-PET image pair from the 20-pair dataset.

TABLE II. QUANTITATIVE PERFORMANCE METRICS AVERAGED OVER 20 MRI-PET IMAGE PAIRS FOR EACH EXAMINED FUSION METHOD

Algorithm	***EI***	***AG***	Q_{CV}	***EN***	$Q^{(AB/F)}$	***SF***	***SD***
TIF_WLS_LatLRR_CNN	93.68164	9.244903	706.6002	4.888168	0.552819	32.03373	73.95617
TIF_WLS_GFCE_CNN	92.56898	9.20217	663.2106	6.593209	0.545897	29.21043	71.93798
TIF_WLS_CBF_CNN	81.99519	8.097168	494.0103	4.909907	0.60165	27.70182	67.51222
TIF_GFCE_LatLRR_CNN	105.0489	10.47289	738.7903	6.574837	0.525709	34.46772	74.28399
TIF_CBF_LatLRR_CNN	95.4365	9.379916	566.0772	4.920685	0.58074	32.4334	70.22616
GFCE_WLS_LatLRR_CNN	105.3751	10.57933	714.756	6.584664	0.525654	34.98914	75.16688
CBF_WLS_LatLRR_CNN	95.82341	9.491286	541.1313	4.941278	0.584818	32.71096	71.00709
CBF_WLS_GFCE_CNN	96.06986	9.595689	522.5791	6.597455	0.56842	30.66984	70.1085
CBF_GFCE_LatLRR_CNN	107.9071	10.79011	632.2274	6.582859	0.542446	35.59	72.43845
TIF_CBF_GFCE_CNN	95.25208	9.447546	536.7366	6.594488	0.569512	30.12274	68.90651
LatLRR	117.9522	11.70785	1023.075	4.729255	0.448301	43.0999	79.2323
WLS	66.92955	6.73272	683.5474	4.687889	0.544969	23.22147	58.80359
TIF	65.19256	6.355294	606.5297	4.578237	0.530517	22.34428	55.37364
GFCE	116.1719	12.00256	1047.961	7.188665	0.392422	40.96589	74.99643
CBF	76.22885	7.538242	261.2931	4.723125	0.635745	25.41379	52.4742

Performance was assessed using widely accepted image fusion metrics: *AG*, *EN*, $Q^{(AB/F)}$, Q_{CV}, *SD*, and *SF*. These metrics reflect edge clarity, information richness, global contrast, and spatial detail preservation. *AG* and *SD* are well-recognized indicators of local contrast, which is essential in evaluating the visibility and fine structural preservation of fused medical images.

The results show that the proposed DMF approach consistently outperforms competing methods across both MRI-CT and MRI-PET fusion tasks. This is evident in the higher *AG*, *EN*, and $Q^{(AB/F)}$ values, which confirm the method's ability to maintain detailed features and contrast transitions. While a slight reduction in Q_{CV} was observed in some CNN-based outputs, this is likely due to the smoothing effects introduced during CNN weight map generation, which enhance structural clarity but may marginally reduce contrast in decision-relevant regions. Nevertheless, the overall metric balance confirms the robustness and generalization capability of the DMF framework.

To ensure diverse architectural evaluations, the proposed approach combines multiple fusion outputs using CNNs, resulting in ten distinctive variants such as TIF_WLS_LatLRR_CNN, CBF_GFCE_LatLRR_CNN, and GFCE_WLS_LatLRR_CNN. These variants demonstrate the benefit of integrating low-quality with high-quality outputs, achieving enhanced realism, sharper edges, and improved feature definition in the final fused images.

The visual inspections demonstrated in Figures 4-7 reveal that the proposed method produces images with well-preserved anatomical structures, strong edge definition, and enhanced contrast. The fusion results successfully integrate key diagnostic features from both source modalities, eliminating virtual shadows and compensating for missing or ambiguous regions present in individual images. This results in images that offer a more complete and reliable basis for clinical interpretation.

Furthermore, the effectiveness of the proposed method is influenced by the availability of structurally comparable components within the source images. These findings reinforce the importance of utilizing complementary features across modalities to avoid redundancy and ensure comprehensive information capture. The method effectively adapts during the learning and testing stages, producing consistent outcomes when trained on high-quality, well-aligned datasets.

In summary, the proposed DMF technique demonstrates a high degree of adaptability and precision in integrating multimodal medical data. Its ability to capture and retain critical diagnostic information not only supports enhanced clinical decision-making but also holds promise for future integration into automated diagnostic systems and real-time healthcare applications.

IV. CONCLUSION AND FUTURE WORK

This study proposed an improved Deep Learning Medical Image Fusion (DMF) technique that integrates three traditionally fused medical images as Convolutional Neural Network (CNN) inputs, enhanced by gradient-based visibility, consistency, and exposure evaluations. The approach effectively preserves structural details and improves contrast, resulting in fused images with superior diagnostic quality. The evaluations conducted on publicly available Magnetic Resonance Imaging-Computed Tomography (MRI-CT) and MRI-Positron Emission Tomography (MRI-PET) datasets demonstrated that the proposed method outperforms several state-of-the-art traditional and hybrid techniques, as evidenced by the high scores across multiple quantitative metrics, including Average Gradient (AG), Edge Intensity (EI), Entropy (EN), gradient-based fusion metric ($Q^{(AB/F)}$), Chen-Varshney metric (Q_{CV}), Standard Deviation (SD), and Spatial Frequency (SF). While the DMF method demonstrates strong performance across diverse fusion tasks, its efficacy may be influenced by the quality and contrast level of the input images.

The novelty of the proposed DMF approach lies in its hybrid fusion strategy, which combines three traditionally fused images as CNN inputs and incorporates gradient-based visibility, consistency, and exposure assessments to generate adaptive fusion weights. Unlike conventional fusion methods, DMF dynamically adjusts weight computation based on pixel-level structural and exposure variations, enhancing contrast retention and diagnostic clarity. Comparative analysis against established methods—both traditional and CNN-based—demonstrates that DMF achieves superior fusion quality across multiple metrics, validating its effectiveness and generalization across modalities. In summary, the proposed DMF technique offers a robust, scalable, and efficient solution for multi-modal medical image fusion, with promising directions for further enhancement in contrast sensitivity, preprocessing, and deployment in real-world healthcare environments.

Although the experimental datasets naturally include variations in contrast, future work will focus on conducting targeted evaluations using curated low-contrast medical images. Additionally, integrating explicit local contrast metrics will provide a more detailed assessment of contrast enhancement capabilities. Further research will also explore the integration of preprocessing techniques, such as noise reduction and artifact correction, to improve input quality before fusion. Moreover, extending the DMF framework for real-time or near-real-time processing on edge devices or cloud platforms will enhance its applicability in time-sensitive clinical settings.

DATA AVAILABILITY

The data supporting the findings of this study can be made available upon request from the corresponding author.

ACKNOWLEDGMENTS

This study is supported by a research grant funded by the Research, Development, and Innovation Authority (RDIA), Kingdom of Saudi Arabia, with grant number 13382-psu-2023-PSNU-R-3-1-EI. The authors would like to acknowledge the support of Prince Sultan University, Riyadh, Saudi Arabia, for paying the article processing charges of this publication. This research is also supported by the Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia.

REFERENCES

[1] S. U. Khan, I. Ullah, N. Ullah, S. Shah, M. E. Affendi, and B. Lee, "A Novel CT Image De-Noising and Fusion Based Deep Learning Network to Screen for Disease (COVID-19)," *Scientific Reports*, vol. 13, no. 1, Apr. 2023, Art. no. 6601, https://doi.org/10.1038/s41598-023-33614-0.

[2] S. M. Fati, E. M. Senan, and N. ElHakim, "Deep and Hybrid Learning Technique for Early Detection of Tuberculosis Based on X-ray Images Using Feature Fusion," *Applied Sciences*, vol. 12, no. 14, July 2022, Art. no. 7092, https://doi.org/10.3390/app12147092.

[3] R. K. Jogi and M. U. Rani, "A Review of Multimodal Medical Imaging Fusion Methods," *GAMANAM: Global Advances in Multidisciplinary Applications in Next-Gen And Modern Technologies*, vol. 1, no. 2, pp. 90–101, Apr. 2025.

[4] N. Wickramasinghe, B. R. Thompson, and J. Xiao, "The Opportunities and Challenges of Digital Anatomy for Medical Sciences: Narrative Review," *JMIR Medical Education*, vol. 8, no. 2, May 2022, Art. no. e34687, https://doi.org/10.2196/34687.

[5] W. Tang, Y. Liu, J. Cheng, C. Li, and X. Chen, "Green Fluorescent Protein and Phase Contrast Image Fusion Via Detail Preserving Cross Network," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 584–597, May 2021, https://doi.org/10.1109/TCI.2021.3083965.

[6] S. Karim, G. Tong, J. Li, A. Qadir, U. Farooq, and Y. Yu, "Current Advances and Future Perspectives of Image Fusion: A Comprehensive Review," *Information Fusion*, vol. 90, pp. 185–217, Feb. 2023, https://doi.org/10.1016/j.inffus.2022.09.019.

[7] C. Ghandour, W. El-Shafai, and S. El-Rabaie, "Medical Image Fusion Based on Weighted Least Square Optimization and Deep Learning Algorithm," in *2021 9th International Japan-Africa Conference on Electronics, Communications, and Computations*, Alexandria, Egypt, Dec. 2021, pp. 159–163, https://doi.org/10.1109/JAC-ECC54461.2021.9691453.

[8] S. Singh *et al.*, "A Review of Image Fusion: Methods, Applications and Performance Metrics," *Digital Signal Processing*, vol. 137, June 2023, Art. no. 104020, https://doi.org/10.1016/j.dsp.2023.104020.

[9] Z. Wang, Y. Ma, and Y. Zhang, "Review of Pixel-Level Remote Sensing Image Fusion Based on Deep Learning," *Information Fusion*, vol. 90, pp. 36–58, Feb. 2023, https://doi.org/10.1016/j.inffus.2022.09.008.

[10] N. A. El-Hag *et al.*, "Utilization of Image Interpolation and Fusion in Brain Tumor Segmentation," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 37, no. 8, Aug. 2021, Art. no. e3449, https://doi.org/10.1002/cnm.3449.

[11] X. Liu, L. Jiao, L. Li, X. Tang, and Y. Guo, "Deep Multi-Level Fusion Network for Multi-Source Image Pixel-Wise Classification," *Knowledge-Based Systems*, vol. 221, June 2021, Art. no. 106921, https://doi.org/10.1016/j.knosys.2021.106921.

[12] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical Image Fusion With Parameter-Adaptive Pulse Coupled Neural Network in Nonsubsampled Shearlet Transform Domain," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 49–64, Jan. 2019, https://doi.org/10.1109/TIM.2018.2838778.

[13] P. Ganasala and A. D. Prasad, "Medical Image Fusion Based on Laws of Texture Energy Measures in Stationary Wavelet Transform Domain," *International Journal of Imaging Systems and Technology*, vol. 30, no. 3, pp. 544–557, Sept. 2020, https://doi.org/10.1002/ima.22393.

[14] H. Inbarani, A. T. Azar, and G. Jothi, "Leukemia Image Segmentation Using a Hybrid Histogram-Based Soft Covering Rough K-Means Clustering Algorithm," *Electronics*, vol. 9, no. 1, Jan. 2020, Art. no. 188, https://doi.org/10.3390/electronics9010188.

[15] A. D. Prasad and P. Ganasala, "MRI and SPECT Image Fusion Based on Adaptive Fusion Rule in Transform Domain," in *Intelligent Computing Techniques in Biomedical Imaging*. Amsterdam, Netherlands: Elsevier, 2025, pp. 281–293.

[16] M. Diwakar *et al.*, "Multimodality Medical Image Fusion Using Clustered Dictionary Learning in Non-Subsampled Shearlet Transform," *Diagnostics*, vol. 13, no. 8, Apr. 2023, Art. no. 1395, https://doi.org/10.3390/diagnostics13081395.

[17] D. K. Sharma, J. Sachdeva, C. K. Ahuja, and A. Singh, "A Novel Hybrid Multimodal Medical Image Fusion Scheme Based on Non-subsampled Shearlet Transform," *Circuits, Systems, and Signal Processing*, vol. 43, no. 6, pp. 3627–3648, June 2024, https://doi.org/10.1007/s00034-024-02614-4.

[18] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image Fusion Meets Deep Learning: A Survey and Perspective," *Information Fusion*, vol. 76, pp. 323–336, Dec. 2021, https://doi.org/10.1016/j.inffus.2021.06.008.

[19] S. M. Fati, E. M. Senan, and A. T. Azar, "Hybrid and Deep Learning Approach for Early Diagnosis of Lower Gastrointestinal Diseases," *Sensors*, vol. 22, no. 11, May 2022, Art. no. 4079, https://doi.org/10.3390/s22114079.

[20] A. Abdulrahman Albraikan, N. Nemri, M. Abdullah Alkhonaini, A. Mustafa Hilal, I. Yaseen, and A. Motwakel, "Automated Deep Learning Based Melanoma Detection and Classification Using Biomedical Dermoscopic Images," *Computers, Materials & Continua*, vol. 74, no. 2, pp. 2443–2459, 2023, https://doi.org/10.32604/cmc.2023.026379.

[21] Y. Liu, L. Wang, H. Li, and X. Chen, "Multi-Focus Image Fusion with Deep Residual Learning and Focus Property Detection," *Information Fusion*, vol. 86–87, pp. 1–16, Oct. 2022, https://doi.org/10.1016/j.inffus.2022.06.001.

[22] K. Wang, M. Zheng, H. Wei, G. Qi, and Y. Li, "Multi-Modality Medical Image Fusion Using Convolutional Neural Network and Contrast Pyramid," *Sensors*, vol. 20, no. 8, Apr. 2020, Art. no. 2169, https://doi.org/10.3390/s20082169.

[23] W. Kong, C. Li, and Y. Lei, "Multimodal Medical Image Fusion using Convolutional Neural Network and Extreme Learning Machine," *Frontiers in Neurorobotics*, vol. 16, Nov. 2022, Art. no. 1050981, https://doi.org/10.3389/fnbot.2022.1050981.

[24] T. Du *et al.*, "Overcoming the Challenges of Multi-Modal Medical Image Sharing: A Novel Data Distillation Strategy via Contrastive Learning," *Neurocomputing*, vol. 617, Feb. 2025, Art. no. 129043, https://doi.org/10.1016/j.neucom.2024.129043.

[25] K. Koteswara Rao and K. Veera Swamy, "Multimodal Medical Image Fusion Using the MBM-PCNN Model," *IETE Journal of Research*, vol. 71, no. 6, pp. 2161–2175, June 2025, https://doi.org/10.1080/03772063.2025.2477746.

[26] B. K. Shreyamsha Kumar, "Image Fusion Based on Pixel Significance using Cross Bilateral Filter," *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1193–1204, July 2015, https://doi.org/10.1007/s11760-013-0556-9.

[27] W. Xu, H. Ding, L. Deng, C. Liu, X. Lu, and L. Zhu, "Infrared and Visible Image Fusion Based on Fourth Order Partial Differential Equation and Cross Bilateral Filter," in *Eleventh International Symposium on Precision Mechanical Measurements*, Guangzhou, China, Sept. 2024, Art. no. 17, https://doi.org/10.1117/12.3032399.

[28] 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, https://doi.org/10.1109/ACCESS.2024.3402965.

[29] L. Ren, Z. Pan, J. Cao, J. Liao, and Y. Wang, "Infrared and Visible Image Fusion Based on Weighted Variance Guided Filter and Image Contrast Enhancement," *Infrared Physics & Technology*, vol. 114, May 2021, Art. no. 103662, https://doi.org/10.1016/j.infrared.2021.103662.

[30] S. Budhiraja, S. Agrawal, and N. Sharma, "Infrared and Visible Image Fusion based on Sparse Representation and Weighted Least Square Optimization," *IETE Journal of Research*, vol. 71, no. 5, pp. 1504–1516, May 2025, https://doi.org/10.1080/03772063.2025.2469642.

[31] X. Han *et al.*, "An Adaptive Two-Scale Image Fusion of Visible and Infrared Images," *IEEE Access*, vol. 7, pp. 56341–56352, 2019, https://doi.org/10.1109/ACCESS.2019.2913289.

[32] Y. Yang *et al.*, "LatLRR-CNN: An Infrared and Visible Image Fusion Method Combining Latent Low-Rank Representation and CNN," *Multimedia Tools and Applications*, vol. 82, no. 23, pp. 36303–36323, Sept. 2023, https://doi.org/10.1007/s11042-023-14967-0.

[33] M. M. Nanavati and M. Shah, "Implementation and Comparative Study of Pyramid-based Image Fusion Techniques for Lumbar Spine Images," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11139–11145, Aug. 2023, https://doi.org/10.48084/etasr.5960.

[34] X. Zeng, Z. Huang, L. Xu, and Y. Xie, "CP-GAN: Meet the High Requirements of Diagnose Report to Medical Image By Content Preservation," *IET Image Processing*, vol. 16, no. 1, pp. 29–38, Jan. 2022, https://doi.org/10.1049/ipr2.12145.

[35] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, "Automatic Diagnosis of Covid-19 Related Pneumonia from CXR and CT-Scan Images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, https://doi.org/10.48084/etasr.4613.

[36] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013, https://doi.org/10.1007/s10278-013-9622-7.

[37] C. R. Jack *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, Apr. 2008, https://doi.org/10.1002/jmri.21049.