

A Novel Approach for Enhanced Multimodal Medical Image Fusion

Nassima Mezhoud^{1*}, Gueltoum Bendiab^{1,2,3}, Kamel Zeltni^{1,2}, Rayene Chelghoum¹,
Imed Eddine Chama^{1,2}

¹Department of Electronics, University of Frères Mentouri, Constantine 25000, Algeria

²Lsiacio Laboratory, University of Frères Mentouri, Constantine 25000, Algeria.

³Lire Laboratory, University of Abdelhamid Mehri Constantine 2, Constantine 25000, Algeria.

Email: nassima.mezhoud@umc.edu.dz, bendiab.kelthoum@umc.edu.dz, zeltni.kamel@umc.edu.dz,
rayene.chelghoum@umc.edu.dz, chama.imed@gmail.com.

*Corresponding Author

Received: 25.04.2024

Revised: 18.08.2024

Accepted: 02.10.2024

ABSTRACT

The fusion of multimodal medical imaging is crucial in enhancing the clinical utility of medical images for diagnosis and medical issue assessment. Fused images significantly augment the quality of reference images while reducing randomness and redundancy. However, the efficacy of the fused image heavily relies on the chosen fusion techniques. Numerous algorithms have been proposed to enhance the clinical precision of image-based decisions; however, developing efficient fusion methods continues to pose a significant challenge for researchers. This paper proposes a novel multimodal framework for medical image fusion aimed at enhancing output image clarity and diagnosis accuracy. Our approach integrates classical fusion methods such as CV2, DWT, and PCA with multimodal source images. We implemented and evaluated this method in a simulated environment, demonstrating through quantitative analysis that our approach markedly enhances fusion quality compared to existing methods.

Keywords: Multimodal image fusion, medical imaging, CV2, DWT, PCA.

1. INTRODUCTION

Medical imaging has been widely applied for analysing and diagnosing diseases such as atherosclerosis, ageing and cancer. In this context, Multimodal Medical Image Fusion (MMIF) has become one of the most powerful tools that can help medical experts obtain richer and more accurate information by combining medical images from different modalities [1–3]. This technique aims to overcome the limitations of single-modal medical images : by combining various images of the same or different modalities with complementary information. The result is a single composite image that has comprehensive and detailed descriptions of the medical problems [2]. The top imaging modalities commonly used to produce the fused image include Microscopy, magnetic resonance imaging (MRI), Positron Emission Tomography (PET), radiology, printed signals (waves), Fluoroscopy, computed tomography (CT) scanners, Single-Photon Emission Computed Tomography (SPECT), ultrasound and medical radiation [2,4,5]. The produced image is more informative and accurate than any single source image since it contains all the necessary information, leading to better and more appropriate medical decision-making [2,5,6].

The overall multimodal image fusion process includes five basic steps that should be performed after selecting the different imaging modalities and the appropriate fusion algorithm [4,5]. These steps can be summarised as follows: reference image registration, original image decomposition by a decomposition algorithm, fusion method, image reconstruction, and evaluation methods [3,4,6]. The output fused image, obtained from the fusion method, must have all pertinent medical information that was extracted from the source image. Nevertheless it must not contain any additional information that is not present in the reference image. The primary concern regarding this process lies in the fact that the resulting fused medical image's is heavily reliant on the fusion technique's efficiency. To address this, researchers have suggested various fusion algorithms intended to enhance the accuracy of fused image details. Each algorithm has its drawbacks and benefits [1,2,5]. Nonetheless, a trustworthy fusion technique is still necessary for producing more accurate, comprehensive, and potentially more easily interpreted results for various types of medical images. [1,5,7]. In this article, we aim to address these particular issues by proposing a novel multimodal medical image fusion algorithm that combines the output fused images obtained from the convolutional fusion methods CV2 [8], discrete wavelet transform (DWT) and Principal

Component Analyses (PCA) with the multimodal source image in order to further improve the fusion quality of medical images as well as the accuracy of the obtained results. The proposed fusion method can be used in a wide range of medical diagnostic problems. The quantitative experimental results demonstrate that the proposed algorithm significantly improves the fusion quality and performs better than most of the existing fusion methods in most cases.

The rest of this paper is structured as follows. In section 2, we present some significant work in the multimodal medical image fusion domain. In section 3, we give a detailed description of the proposed fusion method. After that, we present in Section 4 the set of experiments that were performed to evaluate the proposed method along with an extensive analysis of the obtained results. Finally, in Section 5, the main conclusions of this work are presented along with some future directions to further enhance and extend their scope.

2. Multimodal Image Fusion Techniques

As previously discussed, multimodal medical image fusion is a research field aimed at developing advanced and reliable techniques that allow for the integration of image data captured from multiple imaging modalities [3,5]. These methods can be classified into three main levels: pixel level, decision level and feature level [4,5]. Decision-level fusion techniques merge information at a high level of abstraction and combine the results from numerous methods to achieve a common final decision that is better than the individual decisions of local fusion methods [2,9]. The fusion process is generally composed of three main steps. First, features are identified and extracted from each source image, then classified using local methods to form the corresponding results. Finally, decision rules are applied to fuse them, aiming to enhance the performance of the classification task. Bayesian techniques [10], dictionary learning, deep learning and machine learning are the most commonly used methods at the decision-fusion level [2,5,11]. In feature-level fusion, the sets of relevant features, such as edges, corner points, contours, texture and shapes [12], that were extracted from multiple biometric sources are combined into a single feature set by applying the appropriate feature normalisation, transformation, and reduction schemes [5,13]. Retrieved features in this category are more informative and better describe the content than pixels [2]. Further, this category resolves many issues related to contrast, noise sensitivity, and mis-registration with pixel-based algorithms [2,14]. The most important fusion methods that have been applied at this level are machine learning, deep learning [15], region-based, and similarity matching [2]. For instance, authors in [16] proposed a multimodal medical image fusion approach based on deep learning convolutional neural networks (CNN) for the fusion process. This approach has been tested on nine sets of input medical images from different modalities. The authors affirmed that their approach achieved better results compared to existing fusion methods.

Pixel-level techniques, which are commonly employed in medical image fusion, operate at the most fundamental level of image processing by directly merging the pixel data from multiple input images. They are generally used without applying any image enhancement technique to the original images [5,17]. These techniques are generally performed in either the transform or spatial domain [18,19]. The transform domain is based on frequency, where the source image is divided into several scales, and the transform coefficients are then merged together in accordance with predetermined fusion rules [2]. Then, the inverse transform of the fused coefficients is utilized to build the output fused image [2,18]. In this domain, wavelet-based algorithms, such as Discrete Wavelet Transformation (DWT), are extremely effective due to their rapid speed of computation, low energy consumption, and good fusion quality [18]. On the other hand, the spatial domain techniques are related to pixels, where pixel values from two or more images are brought together and manipulated to obtain the fused image [19]. This category includes many algorithms like Principal component analysis (PCA), Simple average, select maximum, select minimum, etc.

Multi-scale decomposition (MSD) based fusion is one of the most widely used techniques at the pixel level [20]. In this technique, the source image is first decomposed into multi-scale layers and then, images from different sensors are combined at different scales to produce the fused image [2,20]. Many other popular pixel-level image fusion algorithms have used the Component Substitution (CS) approach to produce the fused image [12,20]. At the pixel level, the fused image is very rich in information content, which improves decision accuracy. However, we must deal with the highest amount of data and the processing demands are larger as compared to feature level and decision level fusion techniques [2,12]. Table 1 presents a comparison between the three levels of image fusion (pixel-level, feature-level, and decision-level) based on some criteria [2,6,9].

Table 1. Comparison between the image fusion levels [2].

Criteria	Pixel	Feature	Decision
Information loss	Minimum	Medium	Maximum
Information content	Highest	Medium	Lowest
Method difficulty	Easiest	Medium	Hardest
Preprocessing	Maximum	Medium	Minimum
Classification per ¹	Worst	Medium	Best
Sensitivity to noise	Highest	Medium	Lowest

¹ Performance.

In the following sections, we will present the three pixel-level image fusion techniques that were used in this study; DWT, PCA and CV2.

2.1. Discrete Wavelet Transform

Discrete Wavelet Transformation (DWT) converts the image from the spatial domain to the frequency domain. In this method, the input source images are first transformed by DWT to their corresponding wavelet coefficient images at each scale level. Corresponding approximation coefficients and detail coefficients of the source images at each level are then fused, respectively, based on a certain fusion rule [5,19] (e.g., pixel-level averaging rule, weighted averaging rule and min-max rule). This rule can be a simple addition or averaging, or a PCA-based weighted averaging. The fused approximation and detail coefficients at each level are used in the final reconstruction of a single output fused image by an inverse DWT [2,21]. Fig. 1 illustrates the block diagram of a pixel-level image fusion process using wavelet transform and pixel-level averaging fusion rule [21].

This technique has found extensive application in the image fusion field, demonstrating notably high performances. For instance, N. Julaiha et al. [22], proposed a multimodal image fusion approach for detecting brain tumours based on (Computed Tomography) scans and MRI (Magnetic Resonance Imaging) images. The fusion was executed utilizing DWT, IDWT, and VGG19 algorithms. This method was evaluated on the public datasets Brain Tumor Segmentation (BraTS) Challenge [23] and Cancer Imaging Archive (TCIA) [24], achieving an impressive accuracy of approximately 90%.

2.2. Principal component analysis

Principal Component Analysis (PCA), also known as the Karhunen-Loève transform (or the Hotelling transform), is a statistical analysis for transforming high-dimensional data into a new coordinate system, where the axes are chosen to be the directions of maximum variance in the data. This linear transformation is mainly used for reducing the dimension

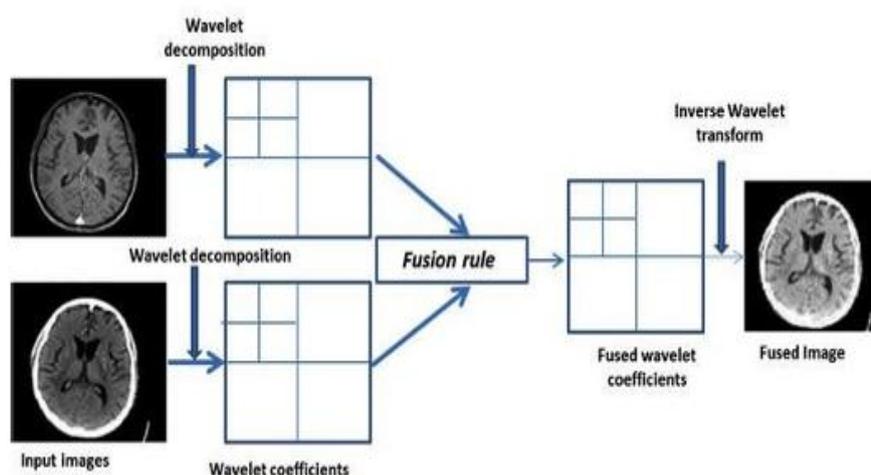


Figure 1. Scheme of Generic DWT fusion[21]

ality of data while retaining as much information as possible. It is widely used in various fields, including statistics, signal processing, computer vision, and machine learning. This approach can be applied to any N number of modalities [25,26], and easily implemented for applications that analyse a huge amount of

data, such as in data compression and pattern 133 matching [25,26]. In addition, PCA plays a crucial role in preserving the essential features 134 of the original image while simultaneously reducing noise. By transforming the image data 135 into a new set of orthogonal components, PCA captures the most significant variations 136 in the data—those that contribute the most to the overall structure of the image. This allows the method to retain key information such as important textures, edges, or patterns, 138 which are critical for interpretation and analysis. At the same time, PCA minimizes less relevant information, such as noise or small fluctuations that do not significantly impact 140 the image's primary content. This noise reduction, coupled with the preservation of key features, enhances the quality of the final output, making it a highly effective tool for tasks like medical image fusion, where clarity and accuracy are paramount [26].

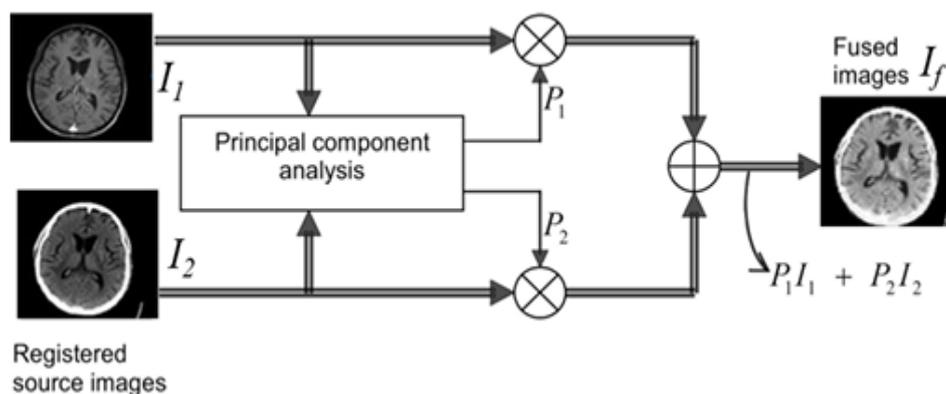


Figure 2. Scheme of Generic PCA fusion[27]

In image fusion, PCA can be used to combine two or more source images from different * modalities. As shown in Fig. 2, PCA generic fusion involves a mathematical procedure that transforms a number of correlated variables into a number of uncorrelated variables called principal components [7,27]. First, the source input images ($I_1(x, y)$, $I_2(x, y)$) are organised in two-column vectors and their empirical means are subtracted. The resulting vector has a dimension of $n \times 2$, where n is the length of each image vector. After that, the eigenvector and eigenvalues for this resulting vector are computed and the eigenvectors corresponding to the larger eigenvalue are achieved. The principal components P_1 and P_2 (i.e., $P_1 + P_2 = 1$) are computed from the obtained eigenvector. The fused image I_f is [27]:

$$I_f = P_1 * I_1(x, y) + P_2 * I_2(x, y) \quad (1)$$

2.3. Open Source Computer Vision Library CV2

OpenCV, also known as CV2, is a robust open-source library primarily utilized for tasks related to computer vision and image processing. This library offers a diverse range of functions designed for performing operations like image resizing, cropping, and filtering, which are essential steps in preparing images for fusion. Moreover, OpenCV equips users with a comprehensive set of tools and functions that enable the implementation of various image fusion algorithms. Some common and fundamental fusion techniques achievable with OpenCV encompass averaging, weighted summation, minimum or maximum operations, Laplacian pyramid blending, and guided filter-based fusion. Notably, CV2 has found extensive application in the fusion of medical images utilizing multiple modalities [28,29]. For instance, authors in [28] leveraged CV2 in conjunction with a machine learning classifier for the early screening of coronary heart disease, achieved through the multimodal fusion of ultrasonic images and electronic medical records (EMRs).

3. Proposed methodology

In this section, we elucidate our methodology which is working to fuse medical images sourced from diverse modalities, including X-ray, Medical Resonance Images (MRI), Computed Tomography scan (CT), Position Emission tomography PET, Single Photon Emission Computed Tomography (SPECT), ultrasound, and nuclear medicine imaging. The fusion of images obtained from different modalities could significantly enhance the quality and clarity of medical images, ultimately facilitating healthcare professionals in their decision-making endeavours. As illustrated in Fig. 3, Our methodology involves various stages, beginning with the loading and pre-processing of the source images. This is followed by the fusion process, where

we adjust the fusion parameters. Lastly, we perform post-processing tasks, such as noise reduction and enhancement, based on the resulting fused image that integrates information from all modalities. More details about these stages are given in the following sections.

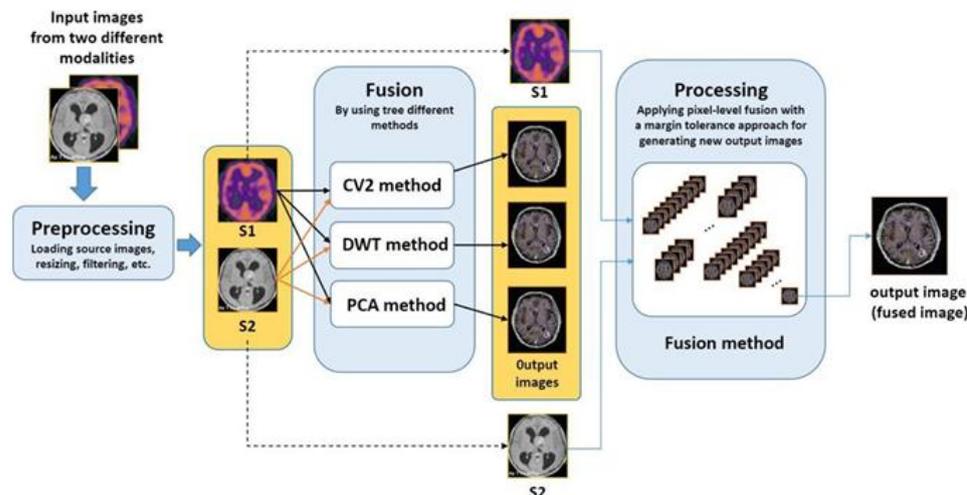


Figure 3. High-level architecture of the proposed approach

3.1. Source images loading and preprocessing

Initially, we load two multi-modal source images as input, referred to as S1 and S2. The source images must have a square size, specifically 500 x 500 pixels. Otherwise, performing some pretreatment is necessary, like normalisation and resizing, which may be applied to the images as needed. This preprocessing is crucial for the compatibility of images and to give both images the required shape for the fusion process. In medical imaging, the normalization of source images is imperative due to potential fluctuations in lighting conditions and colour spaces during image acquisition. By meticulously adjusting brightness and contrast and standardizing colour spaces for coloured images, we achieve a consistent level of intensity and colour information across all regions of the source images. This crucial process serves to alleviate potential disparities, guaranteeing that subsequent analyses or fusion techniques are applied consistently and with precision, thus upholding the integrity and accuracy of diagnostic procedures. Moreover, it is worth noting that medical images are often obtained from diverse sensors, resulting in variations in spatial resolutions. However, our image fusion methodology necessitates spatially matched input images. Consequently, it becomes imperative to resize the input images to uniform dimensions and align them within the same coordinate system. Nonetheless, it is pivotal to emphasize that while this preprocessing stage is strongly recommended, it is not obligatory. There are instances where it may be omitted, particularly when source images exhibit identical origins, precise alignment, uniform dimensions, and consistent coverage, rendering them inherently suitable for fusion without further adjustment.

3.2. Fusion process

Following the necessary preprocessing procedures, the selected fusion technique is applied to source images named S1 and S2. In this paper, we propose a novel approach to combine the resulting source images. This approach involves several sequential steps. Firstly, we utilise the image augmentation technique, which is commonly used in image processing and computer vision, to enhance the size and variety of available images by generating new versions of data. This technique is largely used in machine learning and deep learning applications such as object detection, segmentation, and classification, where it allows for the expansion of the training dataset and the generation of new training samples, particularly when the available data is limited. We then fuse the image sources S1 and S2 together to create new images using three different techniques: CV2, DWT, and PCA. This results in new three images called RCV2, RDWT, and RPCA, produced by the CV2, DWT, and PCA methods, respectively. Next, we use the three generated images from the previous step, along with the source images, (S1 and S2) to generate more images through an iterative process using intensity-based methods.

The low-level fusion methods are used to combine the intensity component, operating at the pixel level, to generate additional images by merging the pixel values from the input images. The fusion process is performed at a low level by pairing up the images and comparing the grey level of each pixel while adhering to a tolerance margin of τ . As previously stated, each image in our approach needs to have a

resolution of 500x500 pixels. Consequently, every image is encoded as a tuple consisting of 250000 elements, with each element representing a pixel value within the image. The fusion process compares two input tuples In_T from two input images and combines them into a newly merged tuple F_T to create the output fused image. In the first iteration, the fusion is performed, and we will take the three generated images RCV2, RDWT, and RPCA and the source images $S1$ and $S2$ as input. Then, the output tuple is determined by calculating the average value if the grey values of each pixel in the input tuples are within a specified tolerance margin. Otherwise, a null value is assigned to the output tuple, as demonstrated in equation 2. Besides, it is important to note that the fusion between the two source images only occurs in the first iteration.

$$F_tup = \begin{cases} (In_T1 + In_T2) / 2, & \text{if } In_T1 \in [In_T2 - \tau, In_T2 + \tau] \\ 0, & \text{else} \end{cases} \quad (2)$$

Upon the completion of this iteration, we will possess a total of nine novel images. The subsequent step involves the combination of the nine produced images to create more images that are both enhanced and contain valuable information, as outlined in equation 3.

$$F_tup = \begin{cases} (In_T1 + In_T2) / 2, & \text{if } In_T1 \in [In_T2 - \tau, In_T2 + \tau] \\ S_{1_T}, & \text{if } In_T1 \in [S_{1_T} - \tau, S_{1_T} + \tau] \\ S_{2_T}, & \text{if } In_T2 \in [S_{2_T} - \tau, S_{2_T} + \tau] \\ 0, & \text{else} \end{cases} \quad (3)$$

Where f_T is the tuple of the fused image, In_T is a tuple of an input image subject to -- the fusion, and S_T is a tuple of a source image. We can keep going through this process until we have the desired amount of images. At each iteration i , we will generate N_i new images based on the previous number of images N_{i-1} generated in the previous iteration $i-1$. Once the augmentation process is complete, we obtain a collection of N additional tuples. In the current step, we will reconstruct the final image from the N tuples generated previously. We engage in a reverse iterative process until we obtain a single, highly detailed image. We will continuously work towards minimizing the number of tuples by extracting relevant pieces of information from each one.

The first step in this process is to eliminate any null tuples, and once that is done, we proceed to separate the remaining non-zero tuples into two distinct packets. The next step involves comparing the tuples from the first packet with the tuples from the second packet, one by one, using equation 2, until we are left with a single tuple. The final tuple that is obtained will be used to reconstruct the resulting image in its final form.

4. Experimental analysis

For the implementation of our approach, we employed the Python 3 programming language alongside various open-source libraries on the Windows operating system. Specifically, we leveraged OpenCV 1 and PyWavelets 2 to implement the CV2 and DWT image fusion methods, respectively. Moreover, we developed a user-friendly interface that facilitates seamless interaction with the multitude of features and functionalities integrated into our framework. To evaluate the capabilities and limitations of our image fusion approach, we conducted computational experiments and validation. These experiments were performed on a CPU equipped with an i7 9750 H, 9th Generation Intel® Core™ i7 processor featuring 6 Cores and 12 Threads running at a speed of 2.60 GHz. Additionally, the system is equipped with 16 GB DDR4 RAM and a 1TB storage capacity hard drive.

4.1. Dataset

To test and validate our approach, we have created a comprehensive multimodal medical imaging dataset, comprising images from four distinct modalities of the brain: MRI, PET, SPECT, and CT scan. Each modality offers unique insights into brain anatomy and function. MRI (magnetic resonance imaging) provides detailed anatomical information on soft tissues, while CT (computed tomography) images offer clear visualization of bone structures. PET (positron emission tomography) images primarily reveal functional information such as metabolic activity. Additionally, SPECT (single-photon emission computed tomography) images provide valuable data regarding the distribution of radioactive tracers within the brain. The original medical images were sourced from the Harvard Medical Image Library website [30]. From this repository, we extracted a collection of 100 registered MRI-CT, MRI-PET, MRI-

SPECT and CT-SPECT image pairs showcasing various brain tumours (see Fig. 5). These images were obtained at high resolutions, with MRI images sized at 256x256 pixels and PET images sized at 128x128 pixels, ensuring a comprehensive and detailed dataset for our tests and validation purposes.

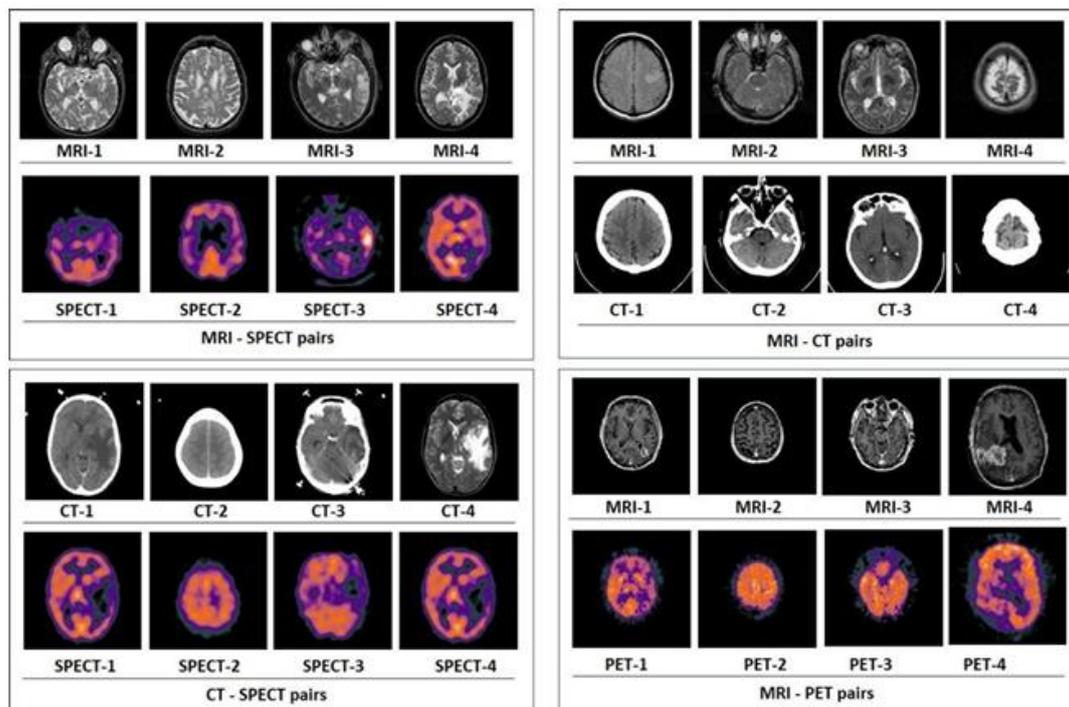


Figure 4. Examples of pairs of source images derived from MRI, CT, MRI and SPECT, and CT and SPECT modalities. The dataset encompasses 25 pairs each of MRI-SPECT, MRI-CT, MRI-PET, and CT-SPECT source images. For instance, Pair-1 includes MRI-1 and CT-1, Pair-2 includes MRI-2 and CT-2, and so forth.

4.2. Fusion results and analysis

In this section, we present the fusion results obtained using the proposed method and provide a detailed analysis of its performance. The results are evaluated based on key metrics, such as image clarity, standard deviation, and noise reduction.

4.2.1. Visual Assessment

The main purpose of visual assessment is to determine whether the proposed fusion method produces images that are more informative, clear, and useful for interpretation. Figure 5 shows the fused images obtained from methods PCA, DWT, CV2 and our approach. Upon closer examination of the fusion results, it becomes evident that each fusion combination yields noteworthy visual qualities. The fusion of MRI and PET images stands out for its exceptional visual quality. This fusion capitalizes on the strengths of both modalities: MRI provides detailed anatomical information with high-resolution clarity, while PET offers valuable functional insights, resulting in a fused image that seamlessly integrates both anatomical and functional data. Similarly, the fused images resulting from MRI and SPECT source images also achieved outstanding visual quality. Here, MRI contributes detailed anatomical information, capturing the intricate structures within the body, while SPECT complements this by providing functional data, resulting in a fused image that offers comprehensive insights into both anatomical structure and functional activity. Concerning the fusion of MRI and CT scan images, we observe yet another instance of exceptional visual quality. In this fusion, MRI excels in providing soft tissue contrast and precise delineation of structures, complemented by the CT scan's ability to offer valuable information about bone density. The resulting fused image thus combines the strengths of both modalities to present a comprehensive depiction of anatomical structures and tissue characteristics. Lastly, the fusion of a CT scan and SPECT images also produces fused images of remarkable visual quality. Here, the CT scan contributes precise anatomical information, offering detailed insights into anatomical structures, while SPECT provides functional data, highlighting areas of functional activity. Together, these modalities create a fused image that offers a comprehensive understanding of both anatomical structure and functional dynamics. Overall, the proposed fusion method yields fused images of exceptional visual quality, each offering unique insights into anatomical structures and functional characteristics. It effectively capitalizes on the complementary

strengths of each modality, resulting in fused images that provide comprehensive and detailed information for diagnostic and analytical purposes.

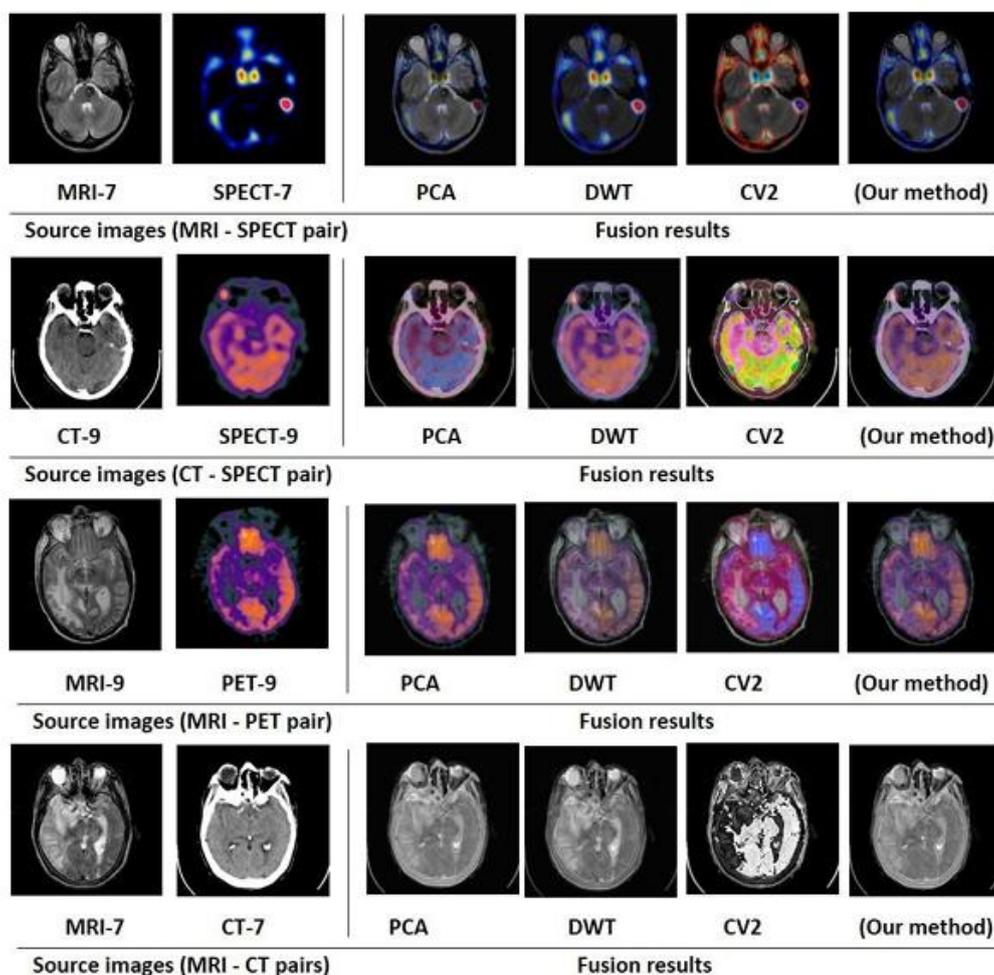


Figure 5. The fusion results of the fusion methods PCA, DWT, CV2 and our method.

4.3. Performance Evaluation and Discussion

In this section, we utilize objective metrics to evaluate the quality of the fused images systematically. These metrics provide quantifiable measurements that help assess various aspects of image quality. Specifically, we focus on the key metrics Standard Deviation (STD), Correlation Factor (Corr), Deviation (D), and Peak Signal-to-Noise Ratio (PSNR).

4.3.1. Correlation (Corr)

Correlation (Corr) helps to evaluate how well the fused image preserves the information from the original modalities by analyzing the Corr between pixel intensities. Higher Corr values indicate that the fusion method effectively integrates the significant features from the input modalities. Table 2 displays Corr values achieved across different fusion methods and depending on the combination of source image modalities, for the ten first samples for each type of source image modalities. Notably, our fusion method consistently achieved the highest Corr values across all samples, with values ranging from 0.71263 to 0.9200 for all samples (20 samples of each subset of the source image modalities). The highest Corr value of 0.9200 was achieved by pair #07 in the MRI-TET subset of samples (see Figure 7). This proves that the structural similarity and Corr between the source images are exceptionally well preserved in the fused images, ensuring that the key structural details from the original modalities are maintained in the final output. This high level of Corr reflects the robustness of our fusion method in preserving critical features and enhancing the overall image quality compared to the other fusion methods.

Table 2. Correlation metric using different fusion methods and source image modalities.

MRI-CT	CV2	PCA	DWT	Our Method	CT-SPECT	CV2	PCA	DWT	Our Method
#01	0.7699	0.8209	0.8548	0.8570	#01	0.7311	0.8050	0.8050	0.8746
#02	0.7522	0.8425	0.8050	0.8448	#02	0.8308	0.8648	0.8814	0.8948
#03	0.6891	0.7893	0.8141	0.8293	#03	0.8153	0.8496	0.8799	0.8864
#04	0.7825	0.8078	0.8249	0.8657	#04	0.7547	0.7590	0.7634	0.7943
#05	0.7834	0.8121	0.8227	0.8246	#05	0.7755	0.8402	0.8409	0.8411
#06	0.6170	0.6254	0.7160	0.8153	#06	0.7539	0.7785	0.7797	0.7895
#07	0.7893	0.8100	0.8828	0.8962	#07	0.7618	0.8475	0.8545	0.8853
#08	0.6013	0.6091	0.7079	0.8087	#08	0.7883	0.8612	0.8621	0.8854
#09	0.7404	0.7864	0.8196	0.8374	#09	0.7586	0.7856	0.7884	0.8104
#10	0.7359	0.7542	0.7652	0.8081	#10	0.7403	0.7994	0.8066	0.8257

MRI-TEP	CV2	PCA	DWT	Our Method	MRI-SPECT	CV2	PCA	DWT	Our Method
#01	0.7699	0.8209	0.8548	0.8570	#01	0.7311	0.8050	0.8050	0.8746
#02	0.7522	0.8425	0.8050	0.8448	#02	0.8308	0.8648	0.8814	0.8948
#03	0.6891	0.7893	0.8141	0.8293	#03	0.8153	0.8496	0.8799	0.8864
#04	0.7825	0.8078	0.8249	0.8657	#04	0.7547	0.7590	0.7634	0.7943
#05	0.7834	0.8121	0.8227	0.8246	#05	0.7755	0.8402	0.8409	0.8411
#06	0.6170	0.6254	0.7160	0.8153	#06	0.7539	0.7785	0.7797	0.7895
#07	0.7893	0.8100	0.8828	0.8962	#07	0.7618	0.8475	0.8545	0.8853
#08	0.6013	0.6091	0.7079	0.8087	#08	0.7883	0.8612	0.8621	0.8854
#09	0.7404	0.7864	0.8196	0.8374	#09	0.7586	0.7856	0.7884	0.8104
#10	0.7359	0.7542	0.7652	0.8081	#10	0.7403	0.7994	0.8066	0.8257

Figure 7 illustrates the highest and lowest correlation values achieved by our method for the four types of input source images, alongside the values obtained by other methods.

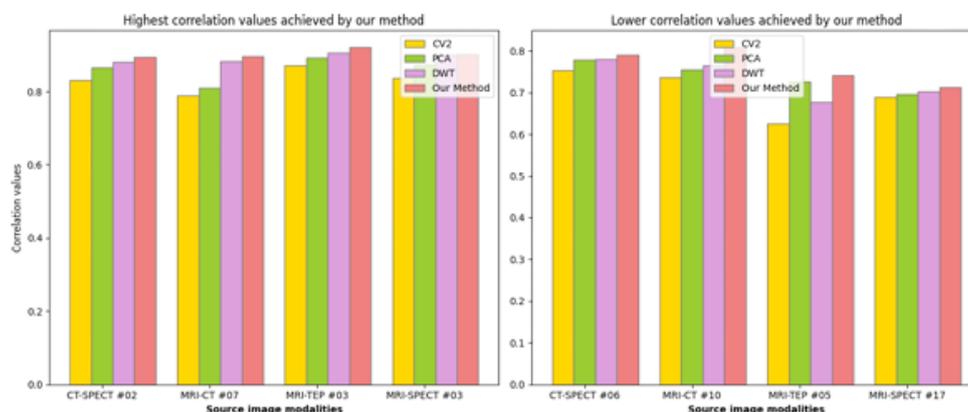


Figure 6. The higher and lower correlation values achieved by our method, for each subset of source image modalities.

The analysis also reveals that the CV2 method consistently produced the lowest Corr values across most samples among the four types of source image modalities. In contrast, the DWT method regularly achieved high correlation values, indicating its effectiveness in maintaining structural similarity between the source images in the fused output. This suggests that the DWT-based fusion technique is particularly stronger than CV2 at preserving important features from the original images. PCA also performed well, with its correlation values frequently comparable to, or slightly higher than, those of DWT. This indicates that the PCA-based fusion method is capable of maintaining Corr between the source images, while also preserving structural information in the fused outputs. While CV2 and DWT may be suitable options when a moderate level of correlation with some variation in the fused image is acceptable, PCA and our proposed method stand out as stronger contenders when the goal is to maximize structural similarity and retain high correlation between the source images. This makes PCA, and particularly our approach, more suitable in medical imaging for accurate diagnosis and analysis.

4.3.2. Deviation (D)

The deviation matrix is used to particularly quantify how much the output fused image deviates from the original source images. A lower deviation typically indicates that the fusion method effectively combines the significant features of the source images without introducing distortions or losing important medical information. Table 3 represents the deviation values achieved across the different fusion methods, for

different combinations of medical image modalities. As shown in the table, our fusion method consistently exhibits the lowest deviation values across all samples and modalities (20 samples for each subset). For example, in the MRI-CT modality, our method achieves the lowest deviation for sample #04 with a value of 43.5875, significantly lower than CV2 (193.0943) and PCA (188.9802). Similarly, in the CT-SPECT modality, our method also maintains a lower deviation value of 40.8229 for sample #12 compared to other methods. This trend is consistent across various image combinations, suggesting that our method outperforms others in minimizing distortion and better preserving essential features. The consistently higher deviation values from CV2 and PCA indicate that these methods introduce more variation, which may compromise the quality of the fused image. DWT generally performs better than CV2 and PCA but remains inferior to our method, making it particularly effective in medical imaging for accurate diagnosis and analysis (see Figure 3).

Table 3. Deviation metric using different fusion methods and source image modalities.

MRI-CT	CV2	PCA	DWT	Our Method	CT-SPECT	CV2	PCA	DWT	Our Method
#01	209,5712	181,6458	76,6021	52,5045	#01	200,2186	165,3716	67,8285	51,6155
#02	198,9905	179,8668	67,8584	45,3911	#02	205,0162	169,5380	70,6753	59,1587
#03	206,9651	196,7930	72,4547	47,8006	#03	199,8577	164,0082	70,6753	52,9374
#04	193,0943	188,9802	68,1565	43,5875	#04	189,0751	148,3376	64,8168	49,7846
#05	199,8696	193,0034	74,8876	56,6278	#05	208,7859	150,7480	77,2079	60,7501
#06	188,7130	180,5397	69,4918	55,8487	#06	205,2286	149,8373	63,2000	55,8086
#07	209,5712	180,5397	75,0583	60,7036	#07	210,4914	162,9102	72,6343	60,1785
#08	198,9905	189,8668	70,1673	51,2916	#08	206,6956	156,7953	66,2071	52,2906
#09	209,7930	206,9651	74,2819	55,1779	#09	181,4313	147,2543	60,4462	47,8041
#10	190,9802	183,0943	69,1709	50,4042	#10	204,5568	172,8912	83,4280	64,3356
MRI-TEP	CV2	PCA	DWT	Our Method	MRI-SPECT	CV2	PCA	DWT	Our Method
#01	200,2703	185,2731	69,6916	60,7666	#01	182,7714	111,2556	79,5243	44,3638
#02	207,6985	189,8909	77,7609	72,3961	#02	198,2597	121,5225	81,6027	48,8440
#03	195,9520	187,9820	66,1477	60,4645	#03	186,1328	105,0848	78,8030	43,4807
#04	207,1557	198,8534	74,1988	69,9611	#04	205,5901	142,8609	88,6166	52,0163
#05	197,7735	181,3007	69,9884	62,3239	#05	196,6324	104,2041	86,3707	45,5449
#06	208,0381	200,0104	75,8498	70,1619	#06	201,4567	131,4744	90,7243	54,3842
#07	212,5279	204,5583	76,6782	70,1386	#07	198,4749	122,5875	86,6394	50,7468
#08	205,6818	198,4773	62,8339	358,7861	#08	209,5640	143,4603	89,0808	64,8379
#09	208,3912	202,3264	70,4677	65,8965	#09	182,5725	108,3043	86,4830	51,6681
#10	191,0329	184,2477	65,0423	60,5232	#10	188,8593	132,9772	89,7858	69,1362

Figure 7 illustrates the highest and lowest deviation values achieved by our method for the four types of input source images, alongside the values obtained by other methods.

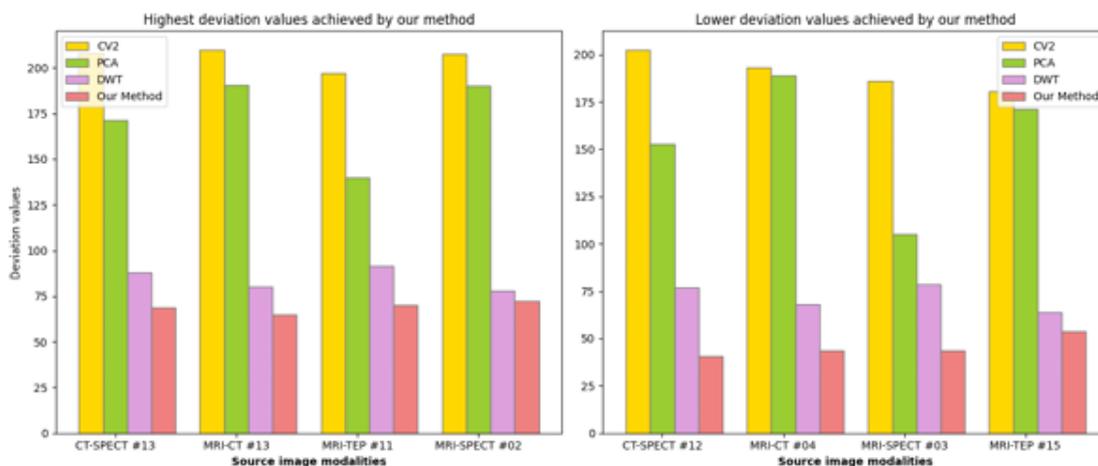


Figure 7. The higher and lower correlation values achieved by our method, for each subset of source image modalities.

4.3.3. Peak Signal-to-Noise Ratio (PSNR)

The PSNR metric assesses how well our fusion method preserves the important details from different imaging modalities while minimizing noise and artifacts. A higher PSNR value is desirable, as it implies better image quality, which is essential for maintaining diagnostic accuracy and clinical utility. Table 4

represents PSNR values achieved across the different fusion methods, for different combinations of medical image modalities (10 first samples). Based on the obtained results, we observe that particularly for CT-SPECT, our fusion method yields the highest PSNR values across almost all cases, outperforming CV2 by significant margins (e.g., 32.847 for CT-SPECT #01 in our method vs. 30.720 in CV2), reflecting more effective noise reduction and detail preservation. Similar trends are observed for MRI-CT fusion, where our method consistently achieves superior PSNR scores, such as 32.400 for MRI-CT #02, compared to 30.566 in CV2. In MRI-SPECT and MRI-TEP combinations, our method again excels, often providing the best balance between enhancing image detail and maintaining overall image integrity, with scores like 32.031 (MRI-SPECT #01) and 32.551 (MRI-TEP #19). Overall, our method consistently outperforms DWT, PCA, and CV2, highlighting its effectiveness in producing high-quality fused images.

Figure 8 illustrates the highest and lowest PSNR values achieved by our method for the four types of input source images, alongside the values obtained by other methods.

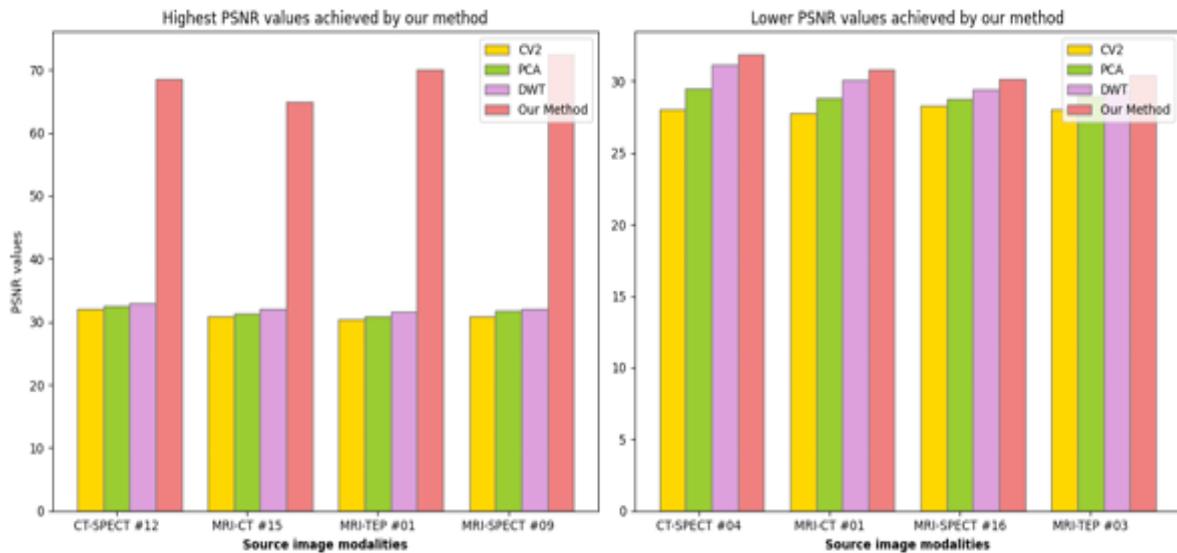


Figure 8. The higher and lower PSNR values achieved by our method, for each subset of source image modalities.

Table 4. PSNR metric using different fusion methods and source image modalities.

MRI-CT	CV2	PCA	DWT	Our Method	CT-SPECT	CV2	PCA	DWT	Our Method
#01	27.7997	28.8466	30.1232	30.7965	#01	30.7208	32.1688	32.3025	32.8470
#02	30.5664	31.8083	31.9695	32.4005	#02	32.1881	32.5251	32.7803	32.9104
#03	30.2257	31.3899	31.8204	32.0318	#03	30.1658	30.2757	31.1300	32.0993
#04	27.5548	29.8320	30.8127	31.7251	#04	28.0360	29.5232	31.1374	31.8643
#05	29.3871	30.7729	31.8157	32.0857	#05	31.4699	31.7510	32.1584	32.5708
#06	28.6410	29.9198	31.4126	31.7441	#06	31.6397	32.7492	32.9303	33.0069
#07	29.9443	30.9739	31.6464	32.0326	#07	30.9191	31.4904	32.5394	32.9404
#08	30.8928	31.1287	31.9052	32.1331	#08	32.3669	32.7346	32.9735	33.0255
#09	27.8687	29.0516	30.1746	31.3340	#09	31.3070	31.4321	32.0787	32.5102
#10	29.0837	31.7086	31.9001	32.0000	#10	31.8910	32.3890	32.6025	32.9201
MRI-TEP	CV2	PCA	DWT	Our Method	MRI-SPECT	CV2	PCA	DWT	Our Method
#01	29.2027	29.5496	29.8653	31.0541	#01	30.3391	30.9199	31.5513	32.0311
#02	28.6337	29.2850	29.5501	30.8635	#02	29.7802	30.0965	30.2307	30.7381
#03	28.0557	28.9968	29.3468	30.4012	#03	30.3720	31.5686	31.7647	31.9438
#04	30.2329	30.8840	31.6047	31.8189	#04	28.7099	29.6151	30.4096	31.5635
#05	29.4491	30.4174	31.1510	31.4740	#05	29.4256	30.1050	30.8159	31.5475
#06	29.5407	30.9442	31.3876	31.5292	#06	29.0129	29.7857	29.9056	30.1907
#07	30.7456	31.7082	31.8361	31.9810	#07	29.5002	30.8018	31.5165	32.0162
#08	28.8341	30.2578	30.8351	31.6056	#08	29.4024	30.7608	31.3460	31.9565
#09	30.9556	31.1318	31.9473	32.0365	#09	27.7206	28.4284	29.5534	30.9308
#10	31.5295	31.7725	31.9724	32.0794	#10	27.8189	28.8493	29.8006	31.3899

4.3.4. Standard Deviation (STD)

The standard deviation metric (STD) is particularly used to evaluate the level of detail and contrast retained in the fusion process. A lower standard deviation indicates that the values are closer to the mean, suggesting less variability in the data. In the context of medical image fusion, a lower standard deviation across modalities implies a more consistent fusion result, as there is less variation in intensity across the fused images. This is mainly desirable as it means a smoother or more coherent fusion, which is beneficial for visualization or diagnostic consistency. Table 5 shows the STD values for the first ten pair-source modality images, for each category. Based on the achieved results, our method shows consistently lower standard deviation values across all fusion combinations. As shown in Figure x, for the CT-SPECT pairs (sample #20), our method achieved a value of 60.99, significantly lower than other methods like DWT at 78.84, PCA at 82.46, and CV2 at 87.31. Similarly, in the MRI-CT pairs (sample #10), our method shows a reduced deviation of 70.56, outperforming DWT (82.84), PCA (85.40), and CV2 (87.13). In more challenging MRI-TEP and MRI-SPECT combinations, where conventional methods struggle to stabilize (e.g., MRI-TEP sample #11 shows CV2 at 100.01 and PCA at 79.93), our method achieves far lower variability with standard deviations of 43.76 and 52.87, respectively. These results indicate that our fusion method is likely the most stable and reliable fusion method among those listed. This stability suggests that it could provide smoother or less variable fusion images, which may be advantageous in clinical settings where consistency is crucial.

CV2 generally has the highest values across the combinations, indicating more variability and potentially less stable fusion results. On the other hand, PCA shows moderate values, lower than CV2 in many cases but not consistently the lowest. DWT has as slightly better (lower) values than PCA but does not outperform our method across all entries.

Figure 9 illustrates the highest and lowest standard deviation values achieved by our method for the four types of input source images, alongside the values obtained by other methods for the same pairs of source modality images.

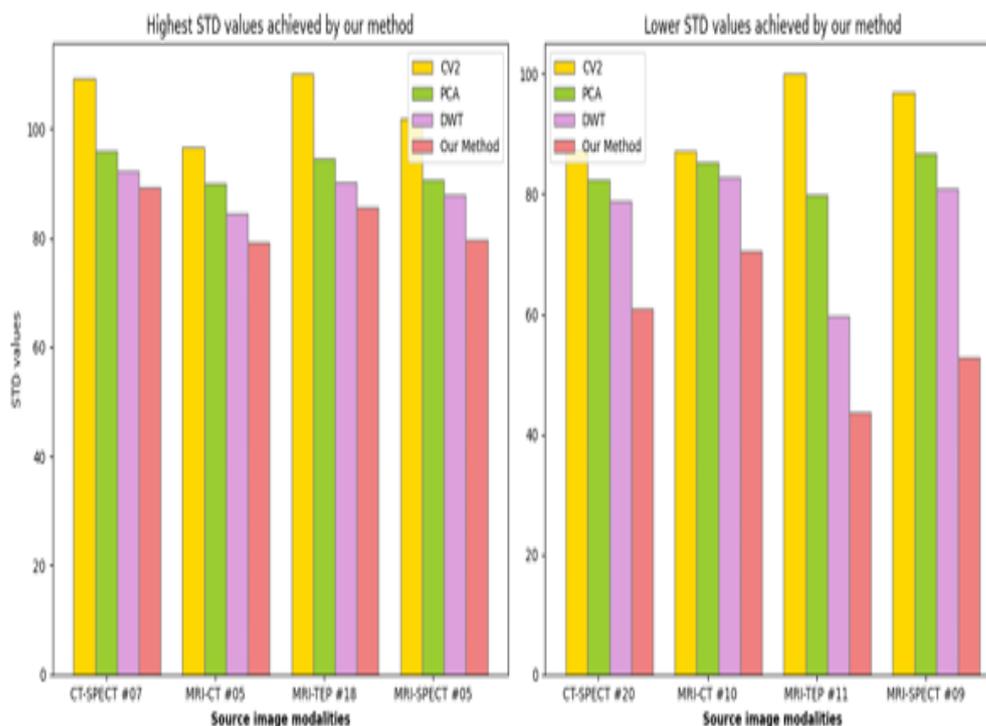


Figure 9. The higher and lower PSNR values achieved by our method, for each subset of source image modalities.

Table 5. STD metric using different fusion methods and source image modalities.

MRI-CT	CV2	PCA	DWT	Our Method	CT-SPECT	CV2	PCA	DWT	Our Method
#01	97.5661	90.0048	85.0875	75.9454	#01	107.0120	93.3806	89.6556	84.5132
#02	95.1438	87.8374	83.8116	72.9337	#02	106.7242	95.8297	93.1059	88.4104
#03	96.2400	88.9058	84.4049	73.9588	#03	100.0550	93.3380	89.9679	81.9510
#04	93.2270	85.0034	81.5898	71.4368	#04	104.3790	94.1686	90.3720	87.7564
#05	96.5719	89.9016	84.5757	79.3184	#05	98.0833	91.5979	88.6856	83.5563
#06	94.8593	88.9125	82.7194	77.1869	#06	108.2152	92.4274	90.1073	88.4319
#07	90.0872	85.9885	79.9686	72.7889	#07	109.1435	95.8864	92.2318	89.2579
#08	87.2692	81.8633	76.8396	70.9529	#08	101.5713	93.3754	89.6860	84.5595
#09	99.7963	91.2881	87.3695	76.2315	#09	109.7156	98.4901	93.8539	89.9717
#10	87.1393	85.4008	82.8477	70.5601	#10	103.0159	95.1029	90.5589	88.0264
MRI-TEP	CV2	PCA	DWT	Our Method	MRI-SPECT	CV2	PCA	DWT	Our Method
#01	98.5964	90.6509	85.8189	80.3215	#01	95.1502	85.1138	71.8383	61.900
#02	108.4998	94.0097	90.0332	85.3583	#02	100.2064	89.9811	73.6106	62.9982
#03	104.3709	91.9746	88.6255	82.7055	#03	104.5351	90.5752	86.0853	67.8602
#04	109.0698	94.6987	89.9229	84.8535	#04	12.5788	15.3766	15.2367	15.7588
#05	102.3556	92.8811	89.0453	80.1154	#05	101.9175	90.6336	87.8328	79.6724
#06	107.5226	94.2983	90.0641	81.4145	#06	101.1981	92.2863	87.3012	60.3735
#07	100.4245	89.8537	60.5955	53.0444	#07	98.1761	85.2306	77.1598	56.1837
#08	81.6243	60.2342	45.8220	40.2959	#08	103.5314	90.0695	80.5575	67.0287
#09	103.5965	92.0664	88.8559	82.7361	#09	96.9686	86.8438	80.8439	52.8795
#10	99.6979	88.2386	63.1378	57.7499	#10	100.5466	90.6073	87.4639	64.3873

In conclusion, our proposed fusion method demonstrates superior performance across multiple metrics, including correlation, deviation, peak signal-to-noise ratio (PSNR), and standard deviation, affirming its effectiveness in preserving structural details from original modalities. The consistently high correlation values, particularly with a peak of 0.9200 in the MRI-TET subset, indicate that our method excels in maintaining key features and enhancing image quality compared to alternative techniques like CV2, which showed the lowest correlation and highest variability. Additionally, our method achieves the lowest deviation values across all samples, significantly minimizing distortion and ensuring essential features are preserved. The PSNR results further highlight our method's capability for effective noise reduction and detail preservation, outshining DWT, PCA, and CV2 in nearly all cases. Lastly, the consistently lower standard deviation values signify its stability and reliability, making it a compelling choice for medical imaging applications where accurate diagnosis and analysis are paramount. Overall, our fusion method stands out as the most robust approach among the evaluated techniques, offering enhanced image quality and consistency critical for clinical settings.

5. CONCLUSION

Fusing multimodal medical images is crucial for disease identification, analyzing treatment options, and improving the performance and precision of computer-assisted systems. In this paper, a multimodal medical image fusion algorithm is proposed based on the combination of the results of fused images obtained using the three methods CV2, DWT, and PCA with multimodal source images, In order to benefit from their results for the improvement of the final fusion quality. The proposed integrated all critical properties from multiple fusion results with source images into a single composite image for more accurate diagnosis and treatment. Experimental results illustrated that the proposed method exhibits significantly better fusion than the three classical methods (CV2, PCA and DWT). The performance of our method is significantly better compared to various distinguished equivalent metrics.

Abbreviations

The following abbreviations are used in this manuscript:

CNN Convolutional Neural Networks

Corr Correlation

CS Component Substitution

CT Computed Tomography D Deviation

DWT Discrete Wavelet Transform

MMIF Multimodal Medical Image Fusion

MRI Magnetic Resonance Imaging

PCA Principal Component Analyses

PSNR Peak Signal-to-Noise Ratio

SPECT Single-Photon Emission Computed Tomography
STD Standard Deviation

REFERENCES

- [1] Huang, B.; Yang, F.; Yin, M.; Mo, X.; Zhong, C. A review of multimodal medical image fusion techniques. *Computational and mathematical methods in medicine* 2020, 2020.
- [2] Hermessi, H.; Mourali, O.; Zagrouba, E. Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Processing* 2021, 183, 108036.
- [3] Azam, M.A.; Khan, K.B.; Salahuddin, S.; Rehman, E.; Khan, S.A.; Khan, M.A.; Kadry, S.; Gandomi, A.H. A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. *Computers in biology and medicine* 2022, 144, 105253.
- [4] Saleh, M.A.; Ali, A.A.; Ahmed, K.; Sarhan, A.M. A Brief Analysis of Multimodal Medical Image Fusion Techniques. *Electronics* 2022, 12, 97.
- [5] Kaur, H.; Koundal, D.; Kadyan, V. Image fusion techniques: a survey. *Archives of computational methods in Engineering* 2021, 28, 4425–4447.
- [6] Tan, W.; Tiwari, P.; Pandey, H.M.; Moreira, C.; Jaiswal, A.K. Multimodal medical image fusion algorithm in the era of big data. *Neural Computing and Applications* 2020, pp. 1–21.
- [7] Alseelawi, N.; Hazim, H.T.; Salim ALRikabi, H.T. A Novel Method of Multimodal Medical Image Fusion Based on Hybrid Approach of NSCT and DTCWT. *International Journal of Online & Biomedical Engineering* 2022, 18.
- [8] Xie, G.; Lu, W. Image edge detection based on opencv. *International Journal of Electronics and Electrical Engineering* 2013, 1, 104–106.
- [9] Jiang, D.; Zhuang, D.; Huang, Y.; Fu, J. Survey of multispectral image fusion techniques in remote sensing applications. *Image fusion and its applications* 2011, pp. 1–23.
- [10] Bhardwaj, J.; Nayak, A. Haar wavelet transform-based optimal Bayesian method for medical image fusion. *Medical & Biological Engineering & Computing* 2020, 58, 2397–2411.
- [11] Li, H.; He, X.; Tao, D.; Tang, Y.; Wang, R. Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recognition* 2018, 79, 130–146.
- [12] Kulkarni, S.C.; Rege, P.P. Pixel level fusion techniques for SAR and optical images: A review. *Information Fusion* 2020, 59, 13–29.
- [13] Tawfik, N.; Elnemr, H.A.; Fakhr, M.; Dessouky, M.I.; Abd El-Samie, F.E. Survey study of multimodality medical image fusion methods. *Multimedia Tools and Applications* 2021, 80, 6369–6396.
- [14] Nirmala, D.E.; Vaidehi, V. Comparison of Pixel-level and feature level image fusion methods. In *Proceedings of the 2015 2nd international conference on computing for sustainable global development (INDIACom)*. IEEE, 2015, pp. 743–748.
- [15] Li, Y.; Zhao, J.; Lv, Z.; Li, J. Medical image fusion method by deep learning. *International Journal of Cognitive Computing in Engineering* 2021, 2, 21–29
- [16] Rajalingam, B.; Priya, R. Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis. *International Journal of ChemTech Research* 2018, 11, 160–176.
- [17] Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Information Fusion* 2017, 33, 100–112.
- [18] Yang, B.; Jing, Z.L.; Zhao, H.t. Review of pixel-level image fusion. *Journal of Shanghai Jiaotong University (Science)* 2010, 15, 6–12.
- [19] Balachander, B.; Dhanasekaran, D. Comparative study of image fusion techniques in spatial and transform domain. *Asian Research Publishing Network (ARPN)* 2016, 11, 5779–5783.
- [20] Ma, T.; Ma, J.; Fang, B.; Hu, F.; Quan, S.; Du, H. Multi-scale decomposition based fusion of infrared and visible image via total variation and saliency analysis. *Infrared Physics Technology* 2018, 92, 154–162.
- [21] Dogra, A.; Goyal, B.; Agrawal, S. Medical image fusion: A brief introduction. *Biomedical & Pharmacology Journal* 2018, 11, 1209.
- [22] AG, N.J.; Vasudevan, B.; Krishnan, R.; Suriya, M. Fusion of Images to Detect Brain Tumors Using DWT, IDWT and VGG19 Algorithms. In *Proceedings of the 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2023, pp. 1540–1548.
- [23] Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M.; et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629* 2018.

- [24] Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging* 2013, 26, 1045–1057.
- [25] Metwalli, M.R.; Nasr, A.H.; Allah, O.S.F.; El-Rabaie, S. Image fusion based on principal component analysis and high-pass filter. In *Proceedings of the 2009 International Conference on Computer Engineering & Systems*. IEEE, 2009, pp. 63–70.
- [26] He, C.; Liu, Q.; Li, H.; Wang, H. Multimodal medical image fusion based on IHS and PCA. *Procedia Engineering* 2010, 7, 280–285.
- [27] Naidu, V.; Raol, J.R. Pixel-level image fusion using wavelets and principal component analysis. *Defence Science Journal* 2008, 58, 338.
- [28] Yang, B.; Zuo, Y.; Yang, S.; Deng, G.; Zhu, S.; Cai, Y. M-US-EMRs: A Multi-modal Data Fusion Method of Ultrasonic Images and Electronic Medical Records Used for Screening of Coronary Heart Disease. In *Proceedings of the International Symposium on Bioinformatics Research and Applications*. Springer, 2022, pp. 88–99.
- [29] Huang, K.; Lin, B.; Liu, J.; Liu, Y.; Li, J.; Tian, G.; Yang, J. Predicting colorectal cancer tumor mutational burden from histopathological images and clinical information using multi-modal deep learning. *Bioinformatics* 2022, 38, 5108–5115.
- [30] Johnson, K.A. The whole brain. Accessed 04/01/2023.