GAMANAM- Global **A**dvances in **M**ultidisciplinary **A**pplications in **N**ext-Gen **A**nd **M**odern Technologies
Volume 1, Issue 2, April 2025

Review Paper

A Review of Multimodal Medical Imaging Fusion Methods

[1*]Ratna Kumari Jogi, [2]M.Usha Rani

[1*] *Research Scholar, Sri Padmavati Mahila Vishwavidyalayam, Tirupati, Andhra Pradesh, India*
[2]*Professor, Department of Computer Science, Sri Padmavati Mahila Vishwavidyalayam,Tirupati, Andhra Pradesh, India*

**Corresponding Author(s):*ratnakumarijogi@gmail.com

Received: 16/01/2025 Revised: 18/02/2025 Accepted: 11/03/2025 Published: 01/04/2025

Abstract: Due to its vital significance in many applications, including medical diagnosis and image enhancement, image fusion has emerged as one of the most promising fields in image processing in recent year. By combining two or more images from various modalities, Multimodal Medical Image Fusion (MMIF) enhances the quality of medical images by producing a fused image that is more lucid and instructive than the original images.. One of the main challenges in assessing image fusion techniques is choosing the optimum MMIF approach to provide the highest quality images. A thorough overview of MMIF procedures is provided in this study, including medical imaging modalities, the stages and levels of medical image fusion, and the methods for evaluating MMIF performance. Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and Single Photon Emission Computed Tomography (SPECT) are examples of common imaging modalities. The six primary types of medical image fusion approaches are sparse representation techniques, fuzzy logic approaches, morphological methods, transform domain methods, and spatial domain methods. The MMIF process functions at three levels: fusion at the pixel, feature, and decision levels. The measures used to evaluate the quality of fusion can be classified as either objective (quantitative) or subjective (qualitative). Additionally, the paper provides a detailed comparison of significant MMIF techniques, outlining their strengths and limitations to offer a comprehensive understanding of their performance.

Keywords: image modality, image fusion, multi-scale decomposition, deep learning, sparse representation

---

1. Introduction

The technique of combining features from several input images to create a single, more accurate and informative image is known as image fusion. By condensing crucial details from the source images, this method promotes visual clarity and usability for both human interpretation and computer analysis. Beyond data reduction, the main objective of picture fusion is to produce images with enhanced readability and diagnostic utility. Multi-sensor image fusion is a particular use of this idea that creates a single output by combining pertinent data from two or more images. Image fusion is a quickly developing area of image processing that uses a variety of techniques to produce sharper, more detailed images that may be used for analysis and decision-making. Utilizing various medical imaging modalities has become essential for enhancing medical diagnosis and treatment. The image fusion process typically consists of five essential steps: image registration, decomposition, applying fusion rule, reconstruction, and evaluation of the final fused image [1-2].

A sophisticated method called Multimodal Medical Image Fusion (MMIF) combines information from several medical imaging modalities. This method efficiently separates and combines important features, allowing for a more thorough comprehension of physiological processes and anatomical structures. MMIF improves image clarity by combining complimentary data, which makes it useful for automated analysis as well as clinical interpretation. In many applications, such as disease diagnosis, treatment planning, and post-treatment evaluation, the technique is essential. It facilitates reliable lesion detection and segmentation, which makes it a crucial tool for precise diagnosis and focused treatment. Since every imaging technique has unique benefits and drawbacks, combining information from several sources improves overall therapy efficacy and diagnostic precision. Multimodal image fusion is becoming increasingly important in medical imaging, as seen by the growing number of research studies on the subject. A survey of PubMed-indexed papers from 2000 to the third quarter of 2022 shows an increasing trend in the creation of fusion techniques, which reflects the ongoing improvements needed to satisfy the changing needs of MMIF[3-6].

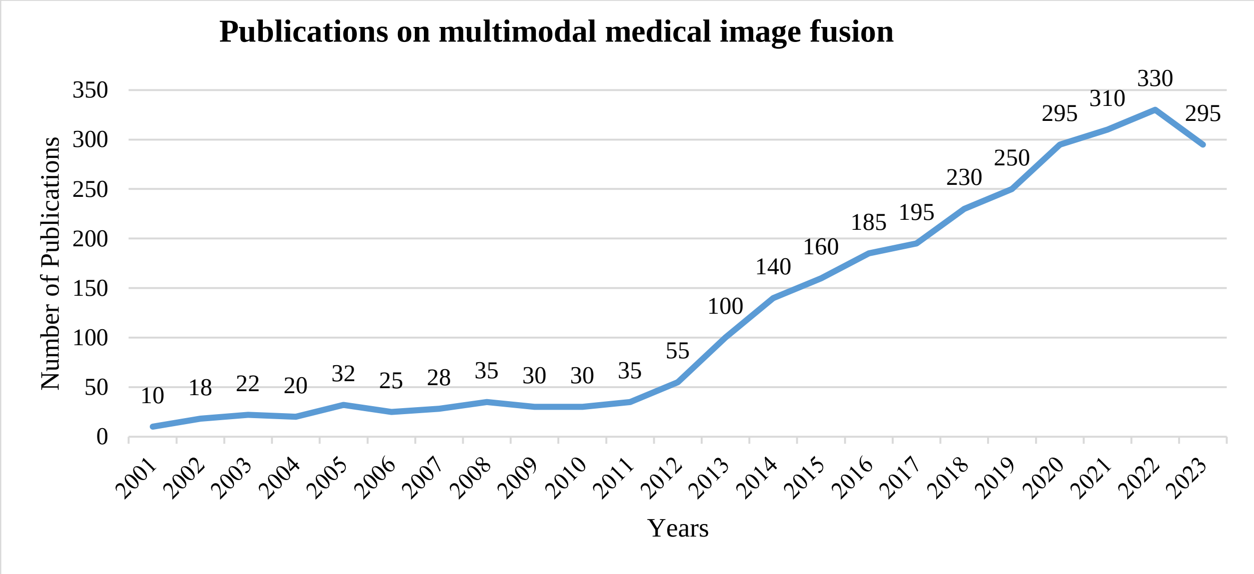


Figure 1: Publications on multimodal image fusion in each year sourced from PubMed

A more thorough comprehension of the image information for diagnosis and evaluation is made possible by medical image fusion, which enhances the clinical interpretability of medical scans. Complementary features from several images captured using various imaging modalities are combined in this procedure. Magnetic Resonance Angiography (MRA), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Computed Tomography (CT), Structural Positron Emission Tomography (SPET), and Single Photon Emission Computed Tomography (SPECT) are among the modalities that are frequently utilized in fusion. Numerous algorithms have been created to combine data from these imaging methods; each has unique advantages but also has drawbacks. As medical image fusion develops, scientists are improving current techniques and creating fresh strategies to deal with new problems [7-8]. The types of datasets can be used to classify image fusion; multi-sensor, multi-focus, multi-scale, multi-temporal and multi-spectral fusion are typical categories [9]. By combining several photos taken at various focus levels, multi-focus fusion creates a single image that preserves all of the important elements [10]. A common technique in medical diagnostics, multi-sensor fusion combines data from various imaging sources to provide a more complete and clear image, frequently displaying details that the human eye could miss [11]. By utilizing complimentary data from many modalities, this method is very useful for precisely identifying anomalies [8].

The objective of multi-scale fusion is to create a single, improved Low Dynamic Range (LDR) image by merging LDR photographs taken under various exposure circumstances. Multi-spectral fusion combines various spectrum data to create a more comprehensible image by using spectral-domain information and spatial-temporal correlations [12]. In order to ensure a more thorough evaluation of medical disorders, multi-temporal fusion is essential for reducing information loss while capturing significant clinical factors over time [13]. Each of the three stages of image fusion techniques—pixel-level, feature-level, and decision-level fusion—contributes to better image quality and interpretability [14]. The simplest method is pixel-level fusion, in which raw data of source images or their multi-resolution conversions are mixed directly. At an intermediate stage, feature-level fusion creates a more detailed output by extracting important characteristics from the input images, such as size, shape, edges, segments, and orientations. High-level fusion, also known as decision-level fusion, combines the outcomes of multiple methods to improve target identification and complete the fusion process [10].

Traditional medical image fusion techniques are generally classified into three main categories: spatial domain, transform domain, and hybrid transform approaches. A significant number of studies have concentrated on spatial domain fusion, utilizing techniques such as Principal Component Analysis (PCA) and Intensity-Hue-Saturation (IHS) [15]. However, spatial domain methods often lead to spectral and spatial distortions, limiting their effectiveness in preserving critical image details. To mitigate these issues, transform domain techniques have been developed, leveraging multi-scale transformations to enhance fusion quality. In this approach, the source images are first converted into the transform domain, where fusion is performed on the transform coefficients before applying an inverse transformation to reconstruct the final fused image. Common transform-based methods include Discrete Wavelet Transform (DWT), contour transforms, and pyramid transforms. These techniques generally maintain structural clarity and minimize distortions; however, they may introduce noise, making noise reduction an essential aspect of the fusion process [16]. In recent years, there has been growing interest in hybrid techniques that integrate both spatial and transform domain methods to achieve superior fusion performance. A notable example is the PCA-DWT approach, which combines the advantages of both domains to enhance image clarity while reducing distortions and artifacts.

Image fusion performance is assessed using both qualitative and quantitative metrics. The two primary evaluation mechanisms are objective and subjective. The subjective method involves visually comparing the source images with the final fused image, but it is often expensive, time-consuming, and not always practical for large-scale applications [18]. In contrast, the objective evaluation method compares the fused image's quality to a reference image, if available, and is widely used for assessing fusion accuracy. Despite the ongoing challenges in image fusion, there is a pressing need for reliable, accurate fusion techniques that work across different image types and applications. Fusion methods must also be resilient to unpredictable acquisition conditions and computationally efficient for real-time applications. Misregistration, a significant source of error in fusion, remains a major issue. This paper aims to review advancements in medical image fusion research and explore future directions, with a particular emphasis on multimodal medical image fusion, which combines images of the same anatomical region from various imaging modalities.

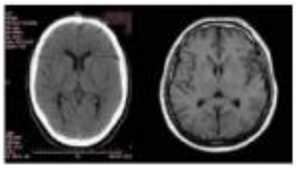
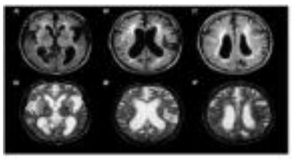
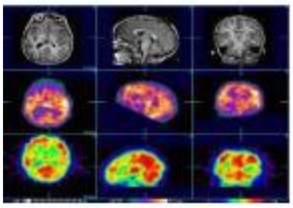
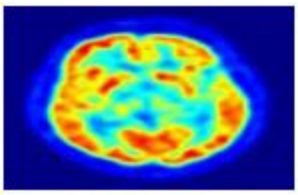
The rest of the document is organized as follows: Medical imaging modalities are summarized in Section 2. Image fusion steps are described in Section 3. The MMIF level classification is covered in Section 4. Different image fusion techniques are categorized in Section 5. The techniques for evaluating fused image quality are presented in Section 6. In section 7, the paper is finally concluded.

2. Review on medical imaging modalities

An overview of the several medical imaging modalities is given in this section. Each modality has unique qualities

and traits that make it possible to examine particular human organs, illnesses, diagnoses, patient states, and subsequent treatments. Radiology, visible light photography, microscopy, 3D reconstruction, and printed signals (waves) are a few examples of imaging modalities [19]. The development of medical diagnosis has been greatly aided by advances in medical imaging capture and enhancement technology. There are two types of medical imaging systems: structural and functional. Both kinds are crucial for identifying lesions. Functional and structural information from medical imaging can be integrated to produce more thorough and significant findings. Medical image fusion is especially useful when treating the same organ since it makes disease monitoring and analysis more accurate. Table 1 gives a quick summary of the several types of medical imaging.

Table 1: Summary of different medical imaging techniques

Imaging modality	Characteristics	Resolution	Application	Radiation Type
 Computed Tomography(CT)	Provides information on significant anomalies, stroke, hemorrhage, brain lesions, dense skulls, and bone structure. Outstanding at visualizing bones	Good spatial resolution	Brain, head, and neck cancer diagnosis as well as tumor identification	X-rays (ionizing)
 Magnetic Resonance Imaging (MRI)	Commonly used imaging technology Provide information on soft tissues that are diseased.	Good resolution	Evaluation of breast cancer and diagnosis of the brain, lung, and liver	Electro magnetic (Non-ionizing)
 Single-Photon Emission Computed Tomography	Analyze blood flow to determine which part of the brain is more active. SPECT can identify conditions like dementia, seizures, blocked blood arteries, and more.	Inadequate spatial resolution	diagnosis of head and neck cancer, treatment for lung, breast and bone cancer, Liver diagnosis and tumor detection	Photons (ionizing)
	Provides details on the operation of the human brain. Allows aberrant brain activity and the symptoms of many disorders to be recorded.	Good spatial resolution	Cancer therapies, unrefined tumor size detection, and esophageal cancer diagnosis	Positron (ionizing)

Positron Emission Tomography (PET)				
 Ultrasound	Diagnostic radiology uses frequencies between 2 MHz and 15 MHz Able to generate diagnostic data that is both qualitative and quantitative.	Good spatial resolution	Applicable to the functional and structural organs of body.	Sound Waves (Non-ionizing)
 X-rays	Utilized for diagnosing a human body's anatomical structure	Adequate spatial resolution	Extensively employed in the evaluation of breast cancer	X-rays (ionizing)

2.1 Structural Systems

High-resolution images with detailed anatomical information are produced by structural imaging modalities as MRI, CT, X-rays, and ultrasound (US). Bones and blood arteries are two examples of tissues with varying densities that can be distinguished using CT scans. However, because MRI is so good at collecting soft tissue structures, it is better suited for imaging the brain, muscles, and organs than bones.

2.1.1 Computed Tomography (CT): One of the primary imaging methods used in image fusion is CT. It generates narrow cross-sectional pictures by detecting X-ray attenuation. CT, a popular non-invasive diagnostic technique in contemporary medicine, has a number of benefits, such as reduced distortion, better visibility of dense structures like bones, and enhanced ability to detect minute changes in tissue composition.

2.1.2 *Magnetic Resonance Imaging (MRI)*: A medical imaging method called magnetic resonance imaging (MRI) uses radiofrequency signals and magnetic fields to produce finely detailed images of anatomical structures and track a number of physiological processes. By creating slices that mimic the human body using magnetic signals, it is especially helpful for delivering information about sick soft tissues.

2.1.3 *Single-Photon Emission Computed Tomography (SPECT)*: A nuclear medicine imaging method called Single-Photon Emission Computed Tomography (SPECT) frequently uses gamma rays to assess blood flow in organs and tissues. It gives 3D data, which is frequently displayed as slices through the body, and is used to evaluate the functional activity of interior organs. Since tissue locations can differ greatly, SPECT is frequently used to image tissues outside of the brain. SPECT-CT and MR-SPECT are two well-known fusion systems that integrate SPECT with other imaging modalities.

2.1.4 *Positron Emission Tomography (PET):* Positron Emission Tomography (PET) is a non-invasive imaging technique that provides important information about the functional activity of organs by assessing the metabolism of a particular tracer. PET scans are crucial for diagnosing a number of illnesses, monitoring changes in brain movement, and identifying anomalies in brain activity. PET is widely utilized in clinical practice and is especially crucial for the identification of full-body cancer. The remarkable sensitivity of PET is its primary benefit. MRI-PET, PET-CT, MRI-CT-PET, and MRI-SPECT-PET are fusion methods that use PET data.

2.1.5 *Ultrasound:* In order to create images, ultrasound imaging uses low-frequency vibrations produced inside the body. Due to its capacity to identify minute changes in tissue characteristics, a new imaging method called Vibro-acustography (VA), which uses ultrasound-induced acoustic emissions, is used in conjunction with mammography to improve the detection of breast cancer. The liver, breast, prostate, and arteries are among the major human tissues where ultrasound is being studied as a non-invasive imaging technique. Tumor density has no effect on ultrasound imaging, in contrast to X-ray mammography, which is unable to measure tissue thickness and depth. Because of this, it is very helpful for highlighting tumor lesions and assessing breast growth. More diagnostic data can be gathered from both imaging modalities by combining ultrasound and X-ray pictures using pixel-based or color-based fusion algorithms.

2.1.6 X-*Rays:* Internal organ imaging is made possible by the creation of "shadowgrams" of the human body using X-rays. X-ray radiation is usually not directly captured in radiography; rather, its intensity is measured and converted into an image, which helps disclose more intricate details about the object under study. Most often, X-rays are utilized to find anomalies and fractures in bones. X-rays are the primary imaging method used in mammography, which is essential for breast cancer screening. Vibro-acustography with X-ray mammography and ultrasound-X-ray are two examples of fusion techniques that combine X-ray data with other modalities to provide more comprehensive and additional diagnostic information.

3. Steps in fusion

The purpose of multimodal image fusion techniques is to combine several images from one or more imaging modalities while maintaining the complimentary information from each image in order to improve image quality and accuracy. Information of various kinds is provided by medical imaging modalities as MRI, CT, PET, US, and SPECT. While PET and SPECT have lower spatial resolution but still provide useful functional data like blood flow and soft tissue movement, MRI, US, and CT offer great spatial resolution and anatomical details of the body. A multimodal image, which provides more thorough information to support medical diagnosis, is produced by combining structural and functional images. The image fusion process combines several input images to produce a more detailed image. First, the images are aligned geometrically by medical image registration. The source photos are combined using an image fusion approach after alignment to create a new image with complementary information. For this process to be successful, the fused image must: (1) retain all pertinent medical information from the original photos; and (2) not add any new information not found in the original images.

Medical imaging uses fusion techniques for a variety of picture types, including multi-focus, multimodal (images from the same modality), and multi-sensor (images from distinct modalities). These are the usual steps in the multimodal fusion process: First, the area of the body or organ of interest is determined. An appropriate algorithm is then used to choose two or more imaging modalities for fusion. Performance measures are employed to evaluate the fusion method. Lastly, the combined image offers more detailed information about the scanned body area than the separate input photos. Figure 2 shows how these stages are frequently used in the spatial domain.

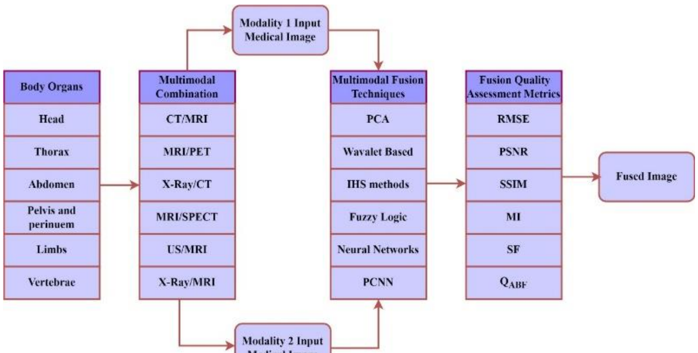


Figure 2: The overall process of fusing multimodal medical images

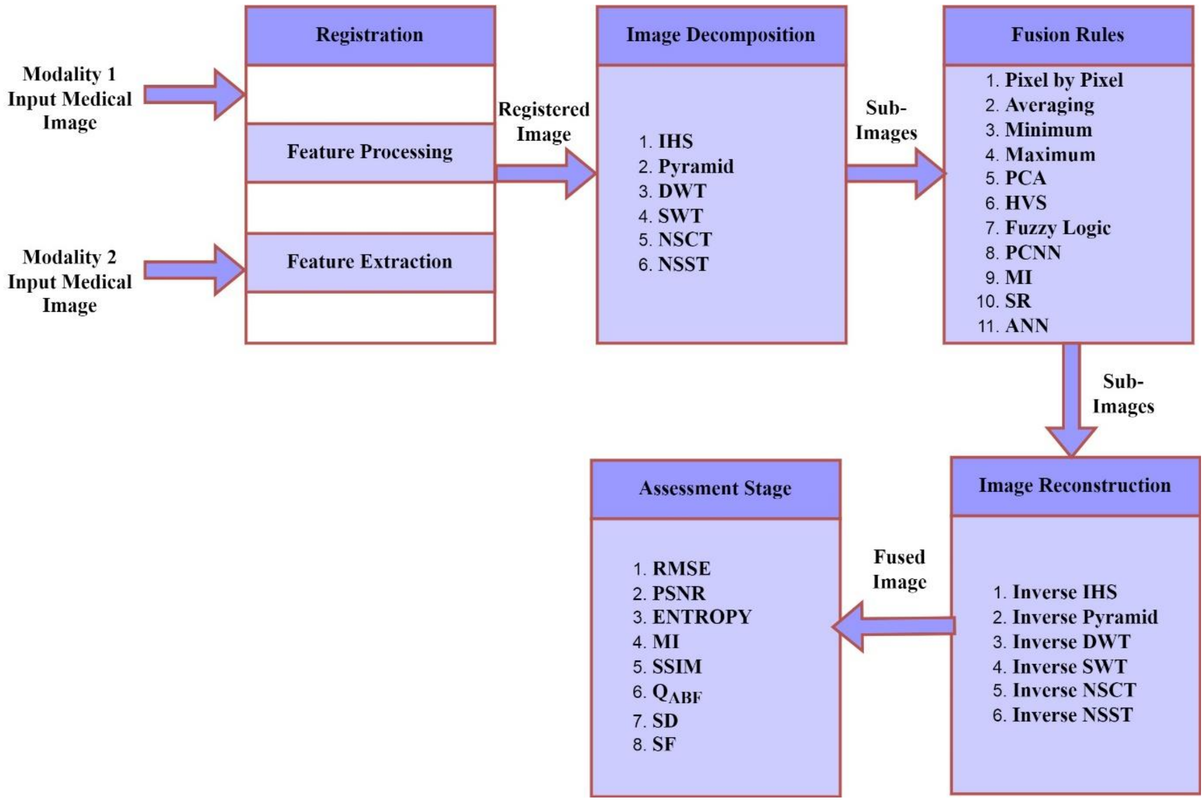


Figure 3: Step-by-step process of medical image fusion

Figure 3 shows the step-by-step process of medical image fusion using the transform domain method. The following are the five steps:

(1) *Image Registration:* to match the related images, the input source image is mapped with the reference image.

(2) *Image Decomposition:* Using decomposition algorithms, the source images are first split up into smaller images. Multiple features from these sub-images are then combined using fusion algorithms like Intensity–Hue–Saturation (IHS), Pyramid, Distinct Wavelet, Non-Subsampled Contourlet Transform (NSCT), Shearlet transform, Sparse Representation (SR), and others [20].

(3) *Fusion Guidelines:* Fuzzy logic, Mutual Information (MI) fusion, Principal Component Analysis (PCA), pulse-coupled neural networks (PCNNs) and Artificial Neural Networks (ANNs), Human Visual System (HVS) fusion are a few examples of fusion algorithms that are used to extract vital information and numerous features from sub-images.

(4) *Image Reconstruction:* the combined image is reconstructed in this stage. The process of assembling the sub-images using an inverse algorithm is known as image reconstruction.

(5) *Image Quality Evaluation Techniques:* Using both subjective and objective evaluation metrics, the image quality assessment is the final stage in determining the quality of the fusion outcome. The radiologists are requested to provide a subjective evaluation of the fusion outcome [21].

4. Multimodal medical image fusion levels

In order to facilitate categorization, we begin by providing a summary of image fusion at the pixel, feature, and decision levels. Techniques at the pixel, feature, and decision levels make up the three layers of picture fusion. Pixel-level picture fusion is the process of directly merging the original information from the source images or their multi-resolution transformations to create a final image that is more informative for visual perception. Feature-level fusion aims to extract key attributes from the source image, including length, form, edges, segments, and orientations. More important features are produced by combining the attributes extracted from the input photos, producing more detailed and evocative pictures. Decision-level fusion is an extensive level of fusion that pinpoints the actual target. To arrive at a final fusion judgment, it aggregates the results from numerous algorithms.

4.1 Pixel level fusion*:* Pixel-level-based fusion methods efficiently integrate images by deciding on the fusion choice based on individual pixels [22]. A spatial domain and a transform domain are two further divisions [23].

4.2 Feature level fusion: Fusion extracts objects of interest for several image modalities at the feature level using segmentation techniques. Components of comparison (such regions) from different visual modalities are then combined using factual methods. To address these issues concurrently, Fei et al. [24] proposed MMIF, which is based on a decision map and sparse representation. Prior to being sorted into vectors according to their location in the original photographs, the raw photos are first divided into patches. A decision map is made in the second stage. Using the decision map, select the vector from each group in the third phase. The remaining vector pairs are then fused using the sparse representation technique. Lastly, the method calculates the average for the overlapping patches after creating the fusion result using the decision map.

4.3 Decision level fusion*:* The input images are handled independently in decision-level fusion, which uses the interpreted/labeled data. The data is consolidated, basic translation is improved, and the observed objects are better understood through the application of selective principles. The primary benefit of this method is that, in accordance with the higher-level representations, multi-modality fusion is reinforced and made more dependable. To create fused images, three methods are often integrated at the decision level. These three methods are statistical methods, logical reasoning, and information theory. Examples of the three methods include voting, fuzzy decision rules, hybrid consensus techniques, joint measures, and Bayesian fusion techniques. Each input image is chosen based on predetermined criteria before fusing into the global optimum based on the validity of each conclusion to create the single fused image in decision-level fusion. To produce as much information as possible, a predetermined principles strategy is applied. Dictionary learning and Bayesian techniques are the most widely used approaches at the decision fusion level [18]. In order to combine data from several sensors, the Bayesian approach depends on the Bayes hypothesis, which is based on probabilities. Examples of Bayesian techniques are DWT Swarm Optimized, HWT Bayesian, and Nonparametric Bayesian. Several existing multimodal medical image fusion techniques in literature are listed in tables 2 and 3 by highlighting the contribution along with advantages and disadvantages of these methods.

Table 2: Major contributions of research papers in multimodal medical image fusion

Reference	Publication Year	Input Modalities Used	Tested Organs of Body	Disease/ Tumor	Method Type	Fusion Technique	Source of Dataset
He et al. [25]	2010	PET/MRI	Human Brain	Alzheimer's disease	Spatial Domain	PCA and IHS	AANLIB
Bhavana et al. [26]	2015	PET/MRI	Human Brain	Alzheimer's disease	DWT	Average	-
Kim et al. [27]	2016	MRI/CT MRI/PET	-	-	Sparse Representation	clustering-based dictionary learning	-
Shabanzade et al. [28]	2016	PET/MRI	-	-	Sparse Representation	SR and Clustering-Based Dictionary Learning in NSCT Domain	-
Xia et al. [29]	2018	CT/MRI MR-T1/PET MR-T2/PET	-	-	NSCT	NSCT-SR-PCNN	-
Ouerghi et al. [30]	2018	PET/MRI	Human Brain	Alzheimer's	PCNN	PCNN	AANLIB
Yin et al. [31]	2018	CT/MRI MRI/PET MRI/SPECT	-	-	NSST	PA-PCNN in NSST Domain SR in NSST Domain	-
Zhu et al. [32]	2019	MRI/PET CT/MRI SPECT/MRI	-	-	NSCT and Laplacian Pyramid	local Laplacian energy	-
Bashir et al. [33]	2019	MRI/CT	Human Brain	-	Spatial Domain	PCA	-
Liu et al. [34]	2020	MRI/CT	Human Brain lesions	lesions	SR and Laplacian Pyramid	local Laplacian energy	AANLIB
Rehal et al. [35]	2021	PET/MRI	Human Brain	-	Spatial Domain	2-DHT and HIS	-
Polinati et al. [36]	2021	MRI/CT	Human Brain	neoplastic,cerebrovascular, degenerative, and infectious diseases	Sparse Representation	adaptive sparse representation	-
Tirupal et al. [37]	2022	MRI/CT MRI/SPECT	Human Brain	-	Fuzzy	interval-valued intuitionistic fuzzy set (IVIFS)	-

Table 3: Merits and Demerits of several approaches

Method Type	Merits	Demerits
Spatial Domain	✓ The most basic image fusion technique ✓ Provides highly focused image output from the input image ✓ Processing is incredibly quick, computationally efficient, and quicker; it is also very clear, easy to understand, and simple to use.	1. Image contrast is reduced by the blurring effect. 2. There is no guarantee that the resulting fused image will be sharp. 3. Spatial domain fusion is usually the cause of spectral degradation. 4. Other possible outcomes include color distortion and spectrum degradation.
DWT	Provides spectral content of image	1. May result in edge artifacts. 2. Experience a blocking artifact.
Pyramidal	Provides directional information	1. Due its shift variant nature, it may result in artifacts around edges. 2. It is computationally costly and requires a lot of memory.
Multi-scale Decomposition	✓ Better in reducing spectrum distortion than conventional fusion techniques; greater signal-to-noise ratio as compared to a strategy based on the pixel-level method. ✓ The fused image has great spatial resolution and high-quality spectral components. ✓ Multilevel fusion produces better outcomes when an image is merged twice in the medical profession using the proper fusion procedure. ✓ Strategies offer exceptional detailed image quality for multi-focus images.	1. A more intricate process is needed for the fusion algorithm than for pixel-level methods. 2. Produce output images that are nearly identical for better outcomes; this requires for a competent fusion procedure. 3. Generate resultant images that are almost identical.
Sparse Representation	✓ The most important elements that improve the ultimate fusion performance are SR coefficients. ✓ Increase the contrast of the image and enhance the retention of visual information. ✓ Preserve details about the images' composition and the thoroughness of the content included in the original images.	Often produces visual aberrations that impact the reconstructed image. Its two primary shortcomings are a high susceptibility to mis-registration and a restricted capacity to retain information.
Deep Learning	✓ Optimizing the image fusion process is made considerably easier by neural networks' learning environment.	1. Based on dynamic processes with complex parameter settings, there are a number of problems that must be fixed, such as misidentification, local extreme,

	✓ Various input data emphasize integrating high-dimensional data to generate a feasible solution. ✓ The approach can be modified to meet the needs of the application. ✓ When there is a lot of input data, they yield better results than other fusion procedures.	and the rate of training convergence. 2. Training the fusion model requires more work and sophisticated technologies. 3. For small image collections, it does not yield dependable results.

5. Quality assessment metrics

Every fusion technique has benefits, and the efficacy of image fusion algorithms is evaluated by taking into account a number of parameter measurements that have been reported in the literature [38-39]. Subjective/qualitative and objective/quantitative evaluation methods are two categories for fusion quality evaluation measures. Based on visual inspection, the subjective quality assessment contrasts the final fused image with the original input images. When examining the fused image, a number of factors need to be taken into account, including color, spatial features, image size, etc. However, the lack of fully fused ground truth images makes current techniques of evaluating quality expensive, time-consuming, and unpleasant.

Two ways are used to categorize the objective method. When the reference image is available, the first approach is applied. When the reference image is unavailable, the second technique is applied. The ground truth image served as the reference image for the fusion algorithm's validation. Only in extremely rare instances is the ground truth medical image accessible, or it might be created by hand. The resulting fused image and the original medical images are utilized to compute the quality metric in the event that the reference image is unavailable. Peak Signal to Noise Ratio (PSNR), Root Mean Square Error (RMSE), Structural Similarity (SSIM), Universal Quality Index (UQI), Correlation Coefficient (CC) and Mutual Information (MI) are objective quality assessment metrics of the reference image.

The Gradient-Based Index ($Q_{AB/F}$), Standard Deviation (SD), Entropy (EN), and Spatial Frequency (SF) are objective quality assessment metrics that do not require a reference image. In Table 3, the findings of the objective quality evaluation are presented. The availability of a reference image determines the classification of the objective approaches. The computations required for each quantitative metric are displayed in the ensuing subsections [40].

5.1 **Metrics Requiring a Reference Image**

5.1.1 *Root Mean Square Error Ratio (RMSE):* RMSE assesses the quality of the final medical image by contrasting the ideal or actual fused image with the ground truth image. For the best combined image results, its value should be near zero. I_R denotes input, I_F denotes fused pictures, and (i,j) denote horizontal and vertical pixels, respectively. The following equation is used to determine RMSE.

$$RMSE = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I_R(i,j) - I_F(i,j))^2} \qquad (1)$$

5.1.2 *Peak Signal to Noise Ratio (PSNR):* It is widely used to gauge how well image fusion reconstruction works. It is the ratio of an image's maximum value to the amount of background noise. When there are similarities between the reference and fused images, the PSNR rating is high. A higher value suggests better fusion. The initial picture and maximum pixel grey level are represented by I and I_{max}, respectively. The abbreviation for mean square error is MSE. I and J are two distinct images that have been merged. The following is the formula used to determine PSNR.

$$PSNR = 20log_{10}(\frac{I_{max}}{\sqrt{MSE}}) \qquad (2)$$

5.1.3 *Structural Similarity Index Measure (SSIM):* The SSIM metric is used to determine the structural similarity between a fused image and a source image. Its rating falls between 0 and 1, with 1 denoting a precise match and 0 denoting a total lack of resemblance to the original image. Since the SSIM value shows how similar the fused image is to the source image, a higher value results in a greater fusion outcome. Eq. (3) can be used to calculate SSIM.

$$SSIM(A,F) = \frac{(2\mu_A\mu_F + C1)(2\sigma_{AF} + C2)}{({\mu_A}^2 + {\mu_F}^2 + C1)({\sigma_A}^2 + {\sigma_F}^2 + C2)} \qquad (3)$$

Where μ_A is the mean of A and μ_F is the mean of F. σ_{AF} is the covariance of A and F, ${\sigma_A}^2$ is the variance of A, and ${\sigma_F}^2$ is the variance of F. To avoid instability that can result from a division with a value near to zero, two constants, $C1$ and $C2$, are used. SSIM readings can range from 0 to 1, with 0 signifying poor quality and 1 signifying excellent quality. A higher MSSIM score results in less distortion within the fused image.

5.1.4 *Mutual Information (MI):* The information that the two images have in common determines how similar they are. For improved fusion, its value should be high. where A and B represent the two input images and F represents the merged image. MI is simple to compute with Equations (4) to (6).

$$MI_F^{AB} = MI(A,F) + MI(B,F) \qquad (4)$$

$$MI(A,F) = \sum_{z\in Z}\sum_{y\in Y} p(A,F)log_2\frac{p(A,F)}{p(A)p(F)} \qquad (5)$$

$$\mathrm{MI}(B,F) = \sum_{z\in Z}\sum_{y\in Y} p(B,F)log_2\frac{p(B,F)}{p(B)p(F)} \qquad (6)$$

5.1.5 *Correlation Coefficient (CC):* It is used to display the relationship between the fused image and the reference image. The values should be close to +1 since it represents spectrum information. C_{rf} is an acronym for the correlation coefficient between the source and the composite image. The optimum value is one when the reference and fused images are exactly the same; as the similarity drops, it falls below one. The following is the formula used to determine CC.

$$CC = \frac{2C_{rf}}{C_r + C_f} \tag{7}$$

5.2 **Metrics that do not require a Reference Image**

5.2.1 *Standard Deviation (SD):* It is used to evaluate the contrast in the fused image. A high standard deviation value results in a considerable contrast in the fused image F having mean value as μ. The following is the formula used to calculate SD.

$$SD = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - \mu)^2} \tag{8}$$

5.2.2 *Entropy (EN):* Entropy is used to measure the fused image with average information content. The high entropy value indicates that the fused image has a high degree of information richness. Entropy is measured in bits per pixel. When the probability associated with grey level i is denoted by $P(i)$, the following formula is used to determine EN.

$$EN = -\sum_{i=0}^{L-1} P(i) log_2 P(i) \tag{9}$$

5.2.3 *Gradient-Based Index(Q_{ABF}):* The quantity of edge information transferred from the input source images to the fused image is measured by Q_{ABF}. After this measure, which has a range of 0 to 1, reaches its ideal value of 1, all of the edges of the original images are transferred to the fused image. A value of 0 indicates the loss of all edge information.

$$Q_{ABF} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(Q_{AF}(i,j)W_A(i,j) + Q_{BF}(i,j)W_B(i,j))}{\sum_{i=1}^{M}\sum_{j=1}^{N}(W_A(i,j) + W_B(i,j))} \tag{10}$$

5.2.4 Spatial Frequency (SF): Spatial frequency is a metric that quantifies the total amount of activity in a fused image by reflecting different contrasts and surface changes. A high SF value results in a better fused image. The following is the formula used to determine SF.

$$SF(i,j) = \sqrt{|RF(i,j)|^2 + |CF(i,j)|^2} \tag{11}$$

$$RF(i,j) = \sqrt{\frac{1}{MxN}\sum_{i=2}^{M}\sum_{j=2}^{N}[I(i,j) - I(i,j-1)]^2} \tag{12}$$

$$CF(i,j) = \sqrt{\frac{1}{MxN}\sum_{i=2}^{M}\sum_{j=2}^{N}[I(i,j) - I(i-1,j)]^2} \tag{13}$$

6. Conclusion

Multimodal Medical Image Fusion (MMIF) is widely employed to enhance the visual quality of fused images, making them more effective for diagnosis and treatment. Over the years, researchers have developed various fusion methods to improve the integration of medical images. This paper presents a comprehensive review of MMIF, covering key aspects such as the classification of medical imaging modalities used in fusion, a detailed explanation of MMIF procedures, and an overview of the three primary fusion levels: pixel-level, feature-level, and decision-level fusion. It also contrasts several MMIF domains, such as spatial fusion and transform domain approaches including wavelet-based fusion, pyramidal fusion, and multi-scale decomposition methods (e.g., NSCT, NSST, PNCC). Additionally investigated are sophisticated methods based on deep learning, sparse representation, morphological processing, and fuzzy logic. Additionally, the study looks at evaluation metrics for image fusion, classifying them into two categories: subjective quality assessment and objective quality assessment. The latter is further subdivided into metrics that are reference-based and metrics that are not. Additionally, a comparison of measures for evaluating image quality among current MMIF approaches is provided.

This paper provides a detailed comparison of recently proposed MMIF methods, presenting numerical evaluations using multiple objective metrics. However, identifying a universally superior technique remains a challenge, as each method has its own strengths and limitations. Experimental findings indicate that spatial domain fusion techniques are computationally efficient and straightforward but often result in spectral distortions, reducing the overall quality of the fused image. In contrast, frequency domain methods provide better fusion results by minimizing spectral distortions and achieving higher signal-to-noise ratios, making them preferable for high-quality image fusion. To further enhance fusion performance, researchers frequently integrate spatial and frequency domain techniques, leveraging the advantages of both approaches to optimize the final fused image.

Sparse representation-based methods are also widely recognized for improving image contrast and preserving structural details of source images. However, they have certain limitations, such as challenges with mis-registration and limited capacity for preserving fine details. Recently, deep learning-based approaches, particularly those using convolutional neural networks (CNNs), have gained popularity for significantly enhancing the quality of fused images. These methods excel when handling large datasets with high dimensionality and diversity, leveraging dynamic processes with numerous parameters to train fusion models. Despite their advantages, deep learning methods are computationally intensive, requiring substantial training time and specialized GPUs. Furthermore, they often struggle to deliver accurate results when applied to smaller image datasets.

References

[1] Vajgl, M.; Perfilieva, I.; Hod'áková, P. Advanced f-transform-based image fusion, Adv. Fuzzy Syst. 2012, 2012, 4.

[2] Blum, R.S.; Xue, Z.; Zhang, Z. An Overview of lmage Fusion, In Multi-Sensor Image Fusion and Its Applications; CRC Press: Boca Raton, FL, USA, 2018; pp. 1–36.

[3] Ganasala, P.; Kumar, V. Feature-motivated simplified adaptive PCNN-based medical image fusion algorithm in NSST domain, J.Digit. Imaging 2016, 29, 73–85.

[4] Tawfik, N.; Elnemr, H.A.; Fakhr, M.; Dessouky, M.I.; El-Samie, A.; Fathi, E. Survey study of multimodality medical image fusion methods, Multimed. Tools Appl. 2021, 80, 6369–6396.

[5] Venkatrao, P.H.; Damodar, S.S. HWFusion: Holoentropy and SP-Whale optimisation-based fusion model for magnetic resonance imaging multimodal image fusion, IET Image Process. 2018, 12, 572–581.

[6] PubMed. Available online: https://www.ncbi.nlm.nih.gov/pubmed/.

[7] James, A.P.; Dasarathy, B.V. Medical image fusion: A survey of the state of the art. Inf. Fusion 2014, 19, 4–19.

[8] Kaur, M.; Singh, D. Multi-modality medical image fusion technique using multi-objective differential evolution based deep neural networks, J. Ambient Intell. Humaniz. Comput. 2021, 12, 2483–2493.

[9] Bhat, S.; Koundal, D. Multi-focus image fusion techniques: A survey, Artif. Intell. Rev. 2021, 54, 5735–5787.

[10] Zhou, Y.; Yu, L.; Zhi, C.; Huang, C.;Wang, S.; Zhu, M.; Ke, Z.; Gao, Z.; Zhang, Y.; Fu, S. A Survey of Multi-Focus Image Fusion Methods, Appl. Sci. 2022, 12, 6281.

[11] Li, B.; Xian, Y.; Zhang, D.; Su, J.; Hu, X.; Guo, W. Multi-sensor image fusion: A survey of the state of the art, J. Comput. Commun. 2021, 9, 73–108.

[12] Bai, L.; Xu, C.;Wang, C. A review of fusion methods of multi-spectral image, Optik 2015, 126, 4804–4807.

[13] Madkour, M.; Benhaddou, D.; Tao, C. Temporal data representation, normalization, extraction, and reasoning: A review from clinical domain, Comput. Methods Programs Biomed. 2016, 128, 52–68.

[14] Ghassemian, H. A review of remote sensing image fusion methods. Inf. Fusion 2016, 32, 75–89.

[15] Dinh, P.-H. Multi-modal medical image fusion based on equilibrium optimizer algorithm and local energy functions, Appl. Intell. 2021, 51, 8416–8431.

[16] Zhang, H.; Xu, H.; Tian, X.; Jiang, J.;Ma, J. Image fusion meets deep learning: A survey and perspective. Inf. Fusion 2021, 76, 323–336.

[17] Jose, J.; Gautam, N.; Tiwari, M.; Tiwari, T.; Suresh, A.; Sundararaj, V.; Rejeesh, M. An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion, Biomed. Signal Process. Control 2021, 66, 102480.

[18] Meher, B.; Agrawal, S.; Panda, R.; Abraham, A. A survey on region based image fusion methods, Inf. Fusion 2019, 48, 119–132.

[19] MITA. Available online: https://www.medicalimaging.org/about-mita/modalities

[20] Dinh, P.-H. A novel approach based on three-scale image decomposition and marine predators algorithm for multi-modal medical image fusion, Biomed. Signal Process. Control 2021, 67, 102536.

[21] Chang, L.;Ma,W.; Jin, Y.; Xu, L. An image decomposition fusion method for medical images, Math. Probl. Eng. 2020, 2020, 4513183.

[22] Daniel, E. Optimum wavelet-based homomorphic medical image fusion using hybrid genetic–grey wolf optimization algorithm,IEEE Sens. J. 2018, 18, 6804–6811.

[23] Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art, Inf. Fusion 2017, 33, 100–112.

[24] Fei, Y.;Wei, G.; Zongxi, S. Medical image fusion based on feature extraction and sparse representation, Int. J. Biomed. Imaging 2017, 2017, 3020461.

[25] He, C.; Liu, Q.; Li, H.; Wang, H. Multimodal medical image fusion based on IHS and PCA, Procedia Eng. 2010, 7, 280–285.

[26] Bhavana, V.; Krishnappa, H. Multi-modality medical image fusion using discrete wavelet transform, Procedia Comput. Sci.2015, 70, 625–631.

[27] Kim,M.; Han, D.K.; Ko, H. Joint patch clustering-based dictionary learning for multimodal image fusion. Inf. Fusion 2016, 27, 198–214.

[28] Shabanzade, F.; Ghassemian, H. Multimodal image fusion via sparse representation and clustering-based dictionary learning algorithm in nonsubsampled contourlet domain. In Proceedings of the 2016 8th International Symposium on Telecommunications (IST), Tehran, Iran, 27–28 September 2016; pp. 472–477.

[29] Xia, J.; Chen, Y.; Chen, A.; Chen, Y. Medical image fusion based on sparse representation and PCNN in NSCT domain, Comput. Math. Methods Med. 2018, 2018, 2806047.

[30] Ouerghi, H.; Mourali, O.; Zagrouba, E. Non-subsampled shearlet transform based MRI and PET brain image fusion using simplified pulse coupled neural network and weight local features in YIQ colour space, IET Image Process. 2018, 12, 1873–1880.

[31] Yin, M.; Liu, X.; Liu, Y.; Chen, X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain, IEEE Trans. Instrum. Meas. 2018, 68, 49–64.

[32] Zhu, Z.; Zheng, M.; Qi, G.;Wang, D.; Xiang, Y. A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain, IEEE Access 2019, 7, 20811–20824.

[33] Bashir, R.; Junejo, R.; Qadri, N.N.; Fleury, M.; Qadri, M.Y. SWT and PCA image fusion methods for multi-modal imagery, Multimed. Tools Appl. 2019, 78, 1235–1263.

[34] Liu, F.; Chen, L.; Lu, L.; Ahmad, A.; Jeon, G.; Yang, X. Medical image fusion method by using Laplacian pyramid and convolutional sparse representation, Concurr. Comput. Pract. Exp. 2020, 32, e5632.

[35] Rehal, M.; Goyal, A. Multimodal Image Fusion based on Hybrid of Hilbert Transform and Intensity Hue Saturation using Fuzzy System, Int. J. Comput. Appl. 2021, 975, 8887.

[36] Polinati, S.; Bavirisetti, D.P.; Rajesh, K.N.; Naik, G.R.; Dhuli, R. The Fusion of MRI and CT Medical Images Using Variational Mode Decomposition. Appl. Sci. 2021, 11, 10975.

[37] Tirupal, T.; Chandra Mohan, B.; Srinivas Kumar, S. Multimodal medical image fusion based on interval-valued intuitionistic fuzzy sets. In Machines, Mechanism and Robotics; Springer: Berlin/Heidelberg, Germany, 2022; pp. 965–971.

[38] Eskicioglu, A.M.; Fisher, P.S. Image quality measures and their performance. IEEE Trans. Commun. 1995, 43, 2959–2965.

[39] Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. Science 2015, 349, 255–260.

[40] S. M.V., S. Kethavath, S. Yerram, S. Kalli, Nagasirisha.B, and J. Brahmaiah Naik, “Brain Tumor Detection through Image Fusion Using Cross Guided Filter and Convolutional Neural Network”, ECTI-CIT Transactions, vol. 18, no. 4, pp. 579–590, Oct. 2024.