Perioperative Precision Medicine
ISSN: 2957-5443

Zentime

Review Article https://doi.org/10.61189/617079irudnn

A review of multimodal medical image fusion: Developments in traditional, model-based and learning-based approaches

Zhaopeng Zhou[1], Jiaen Wu[1], Jiaxun Jiang[1], Miao Zhou[2], Wenhui Guo[3], Yongchu Hu[4]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China. [2]Jiangsu Cancer Hospital, Nanjing 213164, Jiangsu Province, China. [3]The Department of Anesthesiology, Naval Medical University, Shanghai 200433, China. [4]The Department of Anesthesiology, Second Affiliated Hospital of Navy Medical University, Shanghai 200003, China.

Corresponding author: Yongchu Hu.

Declaration of conflict of interest: None.

Received April 10, 2025; Accepted July 11, 2025; Published December 31, 2025

Highlights

- Significant Advantages of Multimodal Fusion: Multimodal medical image fusion enhances diagnostic accuracy and medical value by integrating multi-source data such as CT, MRI, and PET images, outperforming single-modality technologies.
- Benefits of Deep Learning: Deep learning technologies significantly advance multimodal medical image fusion, enabling more efficient and accurate fusion results.
- Perioperative clinical significance: Multimodal image fusion can provide important support for perioperative preoperative planning, intraoperative guidance, and postoperative evaluation, thereby improving surgical accuracy and patient safety.
- Future Research Directions: Future research should focus on improving model interpretability, enhancing modality alignment, and achieving breakthroughs in practicality, applicability, and efficiency.

Abstract

Multimodal medical image fusion technology optimizes image content by integrating images from diverse modalities, such as Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and Single Photon Emission Computed Tomography (SPECT), while retaining critical information. With the rapid advancements in medical imaging technology, single-modal approaches have limitations in capturing comprehensive anatomical or functional characteristics. As a result, researchers are increasingly turning to multimodal fusion methods to enhance diagnostic accuracy and provide richer data for classification, detection, and segmentation tasks. In particular, during the perioperative period, multimodal image fusion plays a crucial role in surgical planning, intraoperative navigation, and postoperative evaluation, enabling precise localization of lesions and improving clinical decision-making. This paper presents a survey of the latest literature on medical image fusion, covering three major approaches: traditional methods, model-based methods, and learning-based methods. It discusses the advantages and limitations of each approach, with a particular emphasis on traditional image processing techniques, model-based fusion methods, and the integration of emerging deep learning (DL) technologies. Comparative experimental analysis highlights performance differences among these methods in terms of information retention, computational efficiency, and clinical applicability. Finally, the paper reviews performance evaluation metrics for multimodal fusion and provides recommendations for future research to further promote the widespread adoption of this technology in clinical diagnostics and intelligent healthcare.

Keywords: Multimodal medical image fusion, multimodal database, perioperative period, deep learning, evaluation metrics

Address correspondence to: Yongchu Hu, The Department of Anesthesiology, Second Affiliated Hospital of Navy Medical University, No.415 Fengyang Road, Shanghai 200003, China. E-mail: liuyang1268@smmu.edu.cn.

Introduction

Medical imaging plays a crucial role in diagnosis and clinical treatment. One of the greatest challenges in medicine is achieving accurate disease identification and effective treatment [1]. The exploration of medical imaging began in 1895 with the discovery of X-rays by German physicist Wilhelm Conrad Roentgen. This groundbreaking discovery not only revolutionized the scientific community but also provided a new tool for medical diagnosis. X-ray imaging enables non-invasive observation of the internal structure of the human body, quickly becoming an indispensable part of medical diagnostics.

With the continuous advancement of science and technology, more imaging techniques have emerged, such as MRI, CT, ultrasound (US), and PET. These technologies have greatly enhanced clinical diagnosis, disease monitoring, and surgical planning, improving both the level of medical care and the quality of life for patients. However, despite their effectiveness, single-modality imaging technologies—such as X-rays, MRI, or ultrasound—each have limitations [2]. For example, X-ray imaging is excellent for visualizing bone structures but less effective for soft tissues. MRI excels in soft tissue imaging but has long scan times and is sensitive to metal implants. CT is ideal for imaging bones and high-density tissues, while PET provides valuable functional and metabolic data. Moreover, single-modality imaging often cannot provide both rich anatomical and functional information simultaneously, which hinders early disease diagnosis and treatment.

As a solution to these limitations, multimodal medical image fusion (MMIF) technology has emerged. Image fusion was first proposed in 1985 and has since rapidly advanced, particularly in fields such as infrared and visible light image fusion, material analysis, remote sensing, multi-focus images, and medical image fusion [3].

In recent decades, medical image fusion technology has significantly advanced medical imaging. It can be broadly classified into single-modality fusion and multimodal fusion. Single-modality fusion typically provides less information, which is why most research focuses on multimodal image fusion [4]. MMIF combines images from multiple modalities, integrating their strengths [5]. This technology offers not only more comprehensive and accurate diagnostic information but also enhances treatment decisions and improves outcomes. Furthermore, it holds great potential for early disease detection, treatment evaluation, and prognosis monitoring, making it an indispensable tool in modern medicine. The core principle is to integrate information from various imaging modalities, overcoming the limitations of single-modality imaging by leveraging the strengths of each. Different imaging modalities provide complementary anatomical and functional data [6]. By effectively fusing these different image modalities, doctors can obtain more precise and comprehensive information, enhancing diagnostic accuracy and increasing the likelihood of early detection.

With the advancement of computer vision, artificial intelligence, and deep learning technologies, multimodal fusion methods based on image processing and machine learning (ML) have matured. These methods have become especially valuable in neuroimaging, oncology, and cardiovascular diseases, improving diagnostic efficiency and treatment outcomes [3, 7, 8]. For example, in brain tumor diagnosis, MRI delineates the tumor boundary and brain tissue structure, while PET reveals areas of metabolic activity. The fusion of these images allows for precise determination of the tumor's location, size, and activity, guiding surgical resection and radiotherapy planning [9]. In cancer treatment, PET/CT fusion plays a critical role in accurate positioning, staging, diagnosis, and radiotherapy planning by combining anatomical and functional metabolic information, significantly enhancing diagnostic accuracy and optimizing treatment strategies [10]. Furthermore, this technology provides crucial perioperative support: it enables comprehensive preoperative assessment by displaying lesion and anatomical information; it enables intraoperative navigation through real-time data integration; and it assists postoperative assessment by accurately monitoring residual lesions and tissue recovery. The integration of multimodal information facilitates preoperative planning and prognostic assessment by providing more detailed and accurate pathological data, thereby improving the level of diagnosis and treatment and prognosis.

Table 1. Comparison of several common imaging methods

Imaging Technology	Imaging Principle	Main Advantages	Main Disadvantages	Applications
MRI	Strong magnetic field + RF signals	High soft tissue contras no ionizing radiation, multi-parametric	Long scan time, high cost, metal implants restricted	Neurology, musculoskeletal, soft tissues
CT	X-ray tomography	Fast scanning, high spatial resolution, suitable for emergencies	Ionizing radiation, lower soft tissue contrast	Chest, abdomen, bones, emergencies
PET	Radioactive tracer metabolic imaging	Provides functional & metabolic info, often combined with CT	High radiation dose, low resolution, high cost	Oncology, cardiovascular, neurology
US	High-frequency sound wave reflection	No radiation, real-time imaging, low cost, portable	Low resolution, operator-dependent, poor penetration through bone/gas	Fetal, cardiac, abdominal examinations
SPECT	Radioactive tracer functional imaging	Shows blood flow & metabolism, can combine with CT	Long imaging time, low resolution, requires radioactive material	Cardiology, brain disorders
OCT	Low-coherence optical imaging	Ultra-high resolution, non-invasive	Limited imaging depth, high cost, motion-sensitive	Ophthalmology (retina), cardiovascular

However, MMIF also faces several challenges, including dealing with modality differences, improving computational efficiency of fusion algorithms, and enabling real-time clinical applications. This paper aims to comprehensively review existing MMIF technologies, focusing on traditional methods, model-based methods, and learning-based methods. Through a systematic analysis of the advantages and disadvantages of these methods and a review of commonly used performance evaluation metrics, this paper explores their prospects and challenges in clinical applications. The ultimate goal is to offer valuable references for researchers to further promote the development of multimodal fusion technology, improving diagnostic accuracy and clinical outcomes.

Imaging technology

Each imaging modality in medical diagnosis has its own unique characteristics and provides specific types of information. Different medical imaging modalities utilize various bands of the electromagnetic spectrum to diagnose different medical conditions in the human body [7].

Below are some common medical imaging modalities, briefly summarized in **Table 1** [11-16].

Magnetic resonance imaging (MRI)

MRI utilizes strong magnetic fields and radiofrequency signals to produce high-resolution images, making it ideal for soft tissue imaging, such as the brain, spinal cord, and muscles. It offers excellent soft tissue contrast, no ionizing radiation, and multi-directional imaging capabilities, making it suitable for detailed observation of lesions. However, MRI has several disadvantages, including long scan times, high costs, and limitations for patients with metal implants.

Computed tomography (CT)

CT uses X-rays to scan the body and quickly generate 3D cross-sectional images. It is particularly useful for imaging the chest, abdomen, and bones and plays a crucial role in emergency diagnoses. CT offers fast imaging and high spatial resolution, especially for bone and lung imaging, providing precise views of internal structures and lesions. However, it uses ionizing radiation, which can harm the body with frequent exposure, and it has lower soft tissue contrast.

Positron emission tomography (PET)

PET uses radioactive tracers to observe metabolic activity within the body and is often combined with CT (PET-CT) to provide both anatomical and functional information. PET is valuable for studying tumors, cardiovascular diseases, and neurological conditions, offering detailed insights into metabolism and function. However, it involves radiation exposure, has lower spatial resolution, and is expensive.

Ultrasound imaging

Ultrasound imaging uses high-frequency sound waves to generate real-time images, commonly used for examining fetuses, the heart, and abdominal organs. It is non-invasive, portable, cost-effective, and provides rapid imaging. However, the quality of ultrasound imaging depends on the operator's skill, and its resolution is lower than that of other imaging modalities. Additionally, ultrasound cannot penetrate gas or bone well, limiting its effectiveness in imaging deep tissues.

Single photon emission computed tomography (SPECT)

SPECT uses radioactive tracers to visualize blood flow or metabolic activity and is often employed in diagnosing heart disease and brain disorders. It provides functional and metabolic information and is usually combined with CT (SPECT-CT) for integrated imaging. However, SPECT has longer imaging times, lower resolution, and involves radioactive materials, which pose a risk of radiation exposure.

Optical coherence tomography (OCT)

OCT uses light waves to provide high-resolution imaging, primarily in ophthalmology and increasingly in cardiovascular fields. It offers excellent resolution for non-invasive tissue observation, especially in eye exams. However, its imaging depth is limited to superficial tissues, and it is relatively expensive. Additionally, motion can interfere with imaging quality.

Multimodal imaging in clinical scenarios

Various imaging modalities have their unique advantages and limitations, making them suitable for different medical scenarios. For example, brain tumor diagnosis is a key area for imaging technologies. A meta-analysis on the diagnostic value of PET/CT for brain tumors showed that PET, using high-efficiency tracers, provides valuable diagnostic information. The combination of PET/CT improves diagnostic accuracy and better delineates brain tumor boundaries, while MRI provides clear anatomical structure. Thus, PET/CT and MRI are often combined in brain tumor diagnostics to obtain more precise information [17]. At the same time, OCT plays an important role in early cancer detection [18]. While OCT has limitations in scanning deep tissue sites, its high resolution, superior to other technologies, makes it ideal for cancer screening. It is also used to guide biopsies, provide real-time feedback during surgeries, and monitor tumor responses to treatment.

As mentioned, single-modality imaging technologies have inherent limitations in providing comprehensive spatial, temporal, and functional information, making it difficult to fully capture tissue structure, function, and metabolic activity. Multimodal imaging technology addresses these limitations by fusing data from different modalities. In multimodal image fusion, combining the strengths of these imaging technologies provides more comprehensive and detailed diagnostic information. With continuous technological advancements, multimodal fusion can effectively enhance diagnostic accuracy and improve disease treatment outcomes.

In this section, we discuss some commonly used multimodal medical databases, including MIMIC-IV, Open-i, TCGA, the Cancer Imaging Archive, Open Access Series of Imaging Studies (OASIS), PhysioNet, and recent large-scale databases, such as MedTrinity-25M and CLIMB (Clinical Large-scale Integrative Multimodal Benchmark). These databases cover various data modalities—electronic health records (EHR), medical images, physiological signals (electrocardiogram, electroencephalogram), genomic data, pathological sections, radiology reports, and clinical records—providing researchers with rich multimodal information. They play an important role in clinical research, disease prediction, imaging genomics, and the development of AI-based diagnostic tools.

MIMIC-IV (https://physionet.org/content/mimiciv) primarily includes patient hospitalization data, diagnoses, EHR, vital signs, lab test results, and medical images, making it a crucial resource for clinical prediction and decision support systems [19]. Open-i (https://openi.nlm.nih.gov) is a platform for medical imaging and text integration, offering X-rays, CT, MRI images, and radiology reports, making it valuable for image-text joint learning [20].

TCGA (https://portal.gdc.cancer.gov) contains CT, MRI images, and genomic data for cancer patients, widely used in imaging genomics

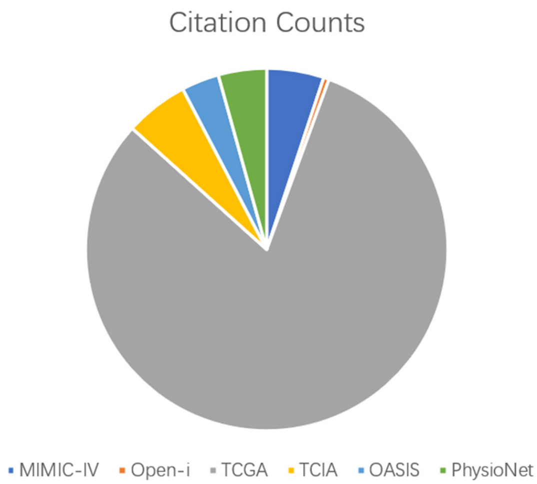

Figure 1. Citations of six databases in PubMed in the past five years.

research [21]. The Cancer Imaging Archive, (https://www.cancerimagingarchive.net) focuses on cancer-related medical imaging, providing high-quality pathology slides, CT, and MRI data to support tumor imaging analysis and radiation oncology research [22]. OASIS (https://www.oasis-brains.org) is dedicated to neuroimaging, containing MRI images and cognitive data from individuals with Alzheimer's disease and normal aging, offering significant value for brain structure research and cognitive disorder prediction [23]. PhysioNet Database (https://physionet.org) includes various physiological signals such as electrocardiogram, electroencephalogram, and cardiac ultrasound, combined with patient medical records, supporting cardiovascular disease research and the development of intelligent diagnostic systems [24].

MedTrinity-25M (https://huggingface.co/datasets/UCSC-VLAA/MedTrinity-25M) provides 25 million images across 10 imaging modalities (CT, MRI, ultrasound, gastroscopy, etc.), with fine-grained annotations for 65 types of lesions, ideal for cross-modal classification, segmentation, and report generation tasks [25].

CLIMB (https://github.com/DDVD233/climb) Integrates 2D images, 3D sequences, monitoring time series, physiological signals, EHR text, and brain network diagrams from 4.51 million patients (19TB of data), covering 96 clinical conditions and 13 fields, with a multi-task evaluation pipeline [26].

The establishment and sharing of these databases, along with advancements in AI and big data technologies, have significantly promoted the cross-modal integration of medical data. This has supported AI innovation in the medical field and fostered research in medical image analysis, intelligent diagnosis, and personalized treatment, etc. **Figure 1** shows the approximate citation counts of the first six databases mentioned above in PubMed, restricted to the past five years. This was achieved by searching for the keyword, such as "MIMIC-IV" in the PubMed database, within a five-year period. **Figure 2** shows several brain images from the OASIS database [27].

MMIF procedure

Multimodal medical image fusion usually involves several key steps. First, medical images from different modality sources undergo preprocessing, including denoising and normalization, to enhance the effect of subsequent processing. Next, image registration is conducted. The source image is aligned with the reference image, and this process is completed based on matched features to provide support for the subsequent fusion step [28]. Then, features are extracted from the image to capture the structural, textural, and functional features in images of different modalities. Finally, fusion rules are employed to highlight key information and features from sub-images, which support later stages of image reconstruction. Image fusion technique is applied by overlapping two input source images to produce a resultant image with excessive and complementary information [29]. Performance metrics are necessary to evaluate and validate the effectiveness of the selected fusion algorithm [30]. The overall MMIF procedure is shown in **Figure 3** [31].

Two prerequisites need attention in the image fusion step. First, all reasonably required medical information present in the input image must appear in the fused image. Second, the fused image should not contain information that is not present in the original input image. Fusion methods are applicable across multi-sensor images from diverse modalities, multi-focus images from the same modality, and are widely employed in multimodal medical images.

MMIF techniques

In MMIF, methods are generally divided into three categories: traditional methods, model-based methods, and learning-based meth-

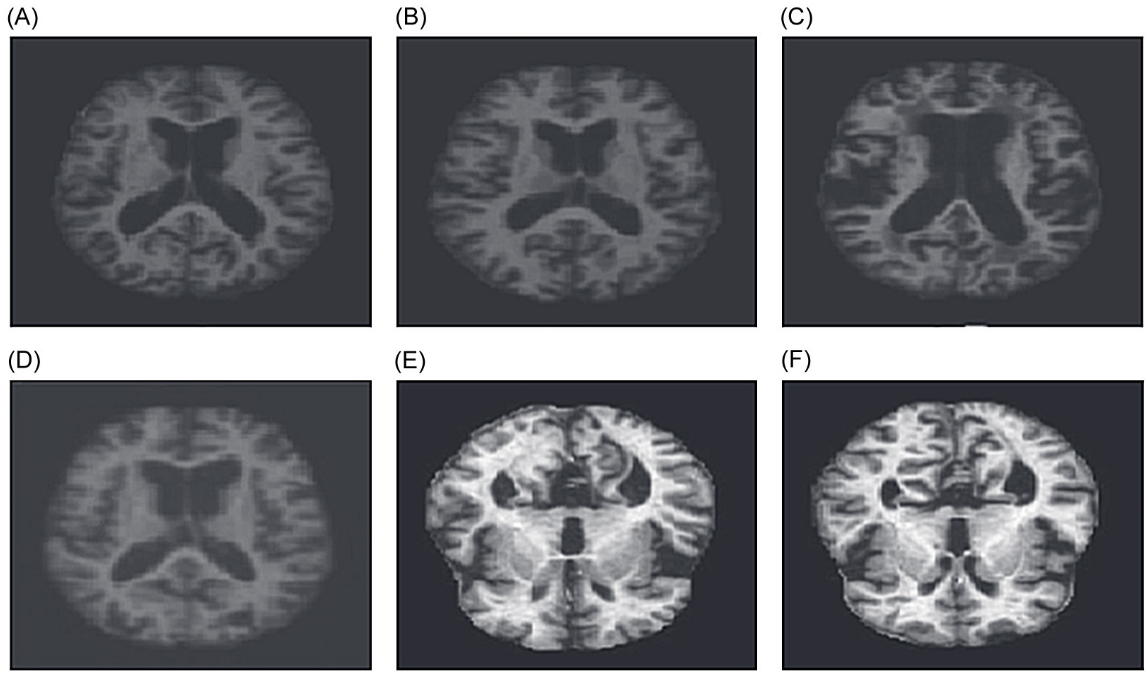

Figure 2. MRI images from OASIS dataset: (A) non-demented aging disease, (B) high-grade mild dementia, (C) mild dementia disease, (D) less extreme dementia disease, (E) Alzheimer's Disease, and (F) healthy control [26].

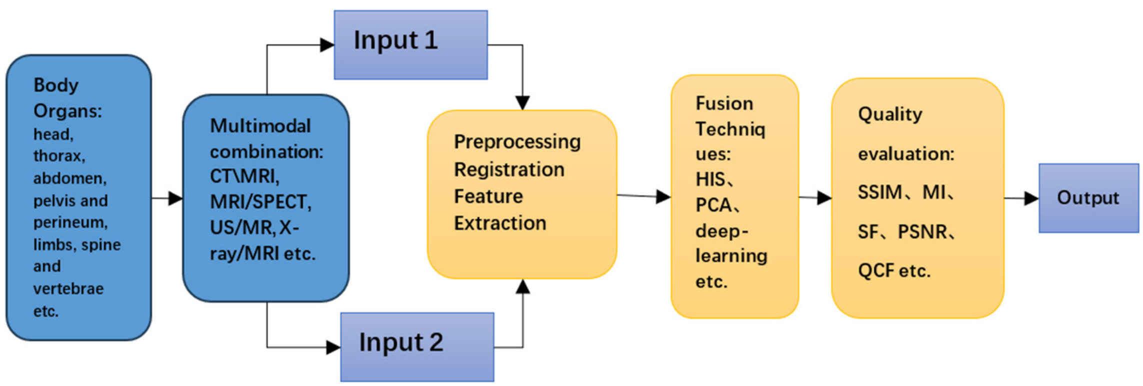

Figure 3. Overall multimodal medical fusion procedure.

ods. Below is a brief introduction to these approaches.

Traditional methods

Traditional methods primarily rely on classic image processing and analysis techniques, often using simple mathematical operations and transformations. Pixel-level fusion is the basic category within these methods, including techniques such as IHS (Intensity-Hue-Saturation) transformation, principal component analysis (PCA), and multi-resolution analysis techniques like Laplace pyramid, wavelet transform, and Contourlet transform [32]. Among these, IHS transformation, wavelet transform, and Laplacian pyramid fusion methods are particularly valuable in practice.

Pixel-level fusion methods combine images at the individual pixel level to make fusion decisions [33, 34]. Since this is applied directly to the initial images, the resulting images exhibit less spatial distortion and a lower signal-to-noise ratio (SNR). In the frequency domain method, images are first converted into the frequency domain using Fourier transform. The fusion process is then performed on the frequency components, followed by an inverse transformation to obtain the final fused image.

(A)

$$\begin{bmatrix} I_0 \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -\sqrt{2}/6 & -\sqrt{2}/6 & 2\sqrt{2}/6 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix}$$

(B)

$$\begin{bmatrix} R_{new} \\ G_{new} \\ B_{new} \end{bmatrix} = \begin{bmatrix} 1 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & -1/\sqrt{2} & -1/\sqrt{2} \\ 1 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} I_{new} \\ v_1 \\ v_2 \end{bmatrix}$$

Figure 4. Transformation formula: (A) RGB to IHS (B) corresponding inverse transformation.

The IHS method has garnered significant attention among traditional methods of MMIF due to its ability to control image details, simplicity in computation, suitability for real-time applications, and status as a classic technique in image fusion [35]. Below is a brief overview of this method.

IHS transformation is a color space conversion method primarily used to convert the RGB (Red-Green-Blue) color space into the IHS (Intensity-Hue-Saturation) color space. The core idea is to separate the image's color information (hue H, saturation S) from the brightness information (intensity I), allowing for independent processing of each component during image processing, ultimately enhancing the visual effect and information fidelity after fusion.

I represents the brightness or grayscale value of the image (the overall brightness of the color). H represents the type of color, such as red, green, or blue. S represents the purity of the color (the depth of the color). The mathematical equations for IHS transformation are as follows:

$$I = \frac{R + G + B}{3} \tag{1}$$

$$H = \arctan\left(\frac{\sqrt{3}\left(G - B\right)}{2R - G - B}\right) \tag{2}$$

$$S = 1 - \frac{\min(R, G, B)}{I} \tag{3}$$

Inverse transformation (from IHS back to RGB):

$$R = I + S\cos(H) \tag{4}$$

$$G = I + S\cos(H - 120^{\circ}) \tag{5}$$

$$B = I + S\cos(H + 120^{\circ}) \tag{6}$$

Figure 4 further shows the RGB-IHS transformation and its inverse transformation process in matrix form.

For instance, Sajid Ullah Khan et al. proposed a novel structural and spectral feature enhancement method in the NSST (Non-Subsampled Shearlet Transform) domain for MMIF [36]. In this domain, the image is decomposed into sub-bands of multiple scales and directions, enabling image processing operations at different scales and directions, such as image fusion, denoising, and feature extraction. First, two pairs of images are generated using the IHS method. The input image is then decomposed using the NSST to obtain low-frequency and high-frequency subbands. The proposed structural information (SI) fusion strategy is applied to the low-frequency subband (LFS), enhancing structural information such as texture and background. Next, PCA is used as the fusion rule for the high-frequency subband (HFS) to extract pixel-level information. Finally, the fused image is obtained through inverse NSST and IHS.

In medical imaging, PET displays metabolic changes in pseudo-color, while MRI provides anatomical structures in grayscale. The fusion of PET and MRI brain images creates a composite image that integrates anatomical structures with metabolic changes, helping doctors effectively diagnose potential diseases. Chen et al. proposed a new fusion method based on IHS and log-Gabor wavelet transform [37]. First, the intensity components of MR and PET images are decomposed using the log-Gabor wavelet transform at an appropriate decomposition scale. Then, a "maximum selection" fusion rule is applied to high-frequency subbands, while a two-stage fusion rule based on a weighted average scheme and visibility measure is used for low-frequency subbands. Finally, an inverse log-Gabor wavelet transform is applied to the fused high- and low-frequency subbands to obtain new intensity components, along with the original hue and saturation components of the PET image, which are then transformed into a fused color image. This method effectively presents both anatomical structures and color changes, while significantly reducing color distortion.

Additionally, Mozhdeh Haddadpour et al. proposed another PET/MRI fusion technology based on the two-dimensional Hilbert transform (2-D HT) and the IHS method [38]. This approach aims to fuse the anatomical details of MRI images with high spatial resolution and the metabolic characteristics of PET images to facilitate more accurate and convenient clinical diagnosis and treatment. Both methods seek to maximize the preservation and enhancement of the spatial and spectral (functional) information from the input images, ultimately providing reliable support for clinical applications.

The IHS method, however, has several limitations: (i) Spectral Distortion: It may lose spectral information, especially in high-dimensional multispectral images. (ii) Non-linear Instability: It lacks stability for images with strong non-linear characteristics. (iii) Color Distortion: Large color discrepancies between input images can cause color shifts after IHS fusion. To address these issues, researchers often integrate IHS with wavelet transforms, PCA, deep learning, and other techniques to improve fusion performance.

Model-based methods

Model-based methods use mathematical and statistical models to capture the characteristics of images, leading to more accurate fusion. Sparse representation fusion methods are widely used within this category. These methods assume that images can be represented by a small number of non-zero coefficients under a redundant dictionary. During the fusion process, a common dictionary is first learned, followed by sparse encoding of the source image. The sparse coefficients are then fused, and the fused image is reconstructed using the dictionary.

For example, PET and MRI images fused using the nonparametric Bayesian-based sparse representation (SR-NPB) technique outperformed three other sparse representation techniques—Traditional Sparse Representation (SR), Adaptive Sparse Representation (ASR), and Joint Sparse Representation (JSR)—in both visual and quantitative comparisons [39]. The new approach was also faster in execution, with experiments using 20 sets of PET and MRI scans. Bhardwaj et al. applied a fractional-order BSA (Bird Swarm Algorithm)-based Bayesian fusion strategy using the BRATS brain MRI dataset [40]. This method simplifies the fusion process by using wavelets generated from the original images using the HWT (Haar Wavelet Transform).

In Nayak's study, a bird swarm optimization (BSO) technique was used to develop a novel Bayesian fusion algorithm on the BRATS dataset [41]. The BSO, inspired by the flocking behavior of birds, optimizes the fusion process by simulating the cooperation and information sharing of birds during foraging and migration. This new medical image fusion scheme, BSO, demonstrated superior performance. Discrete wavelet transform was applied to images of different modalities, extracting decomposition coefficients from the four frequency bands: LL (low–low, capturing overall structure), LH (low–high, representing horizontal details), HL (high–low, representing vertical details), and HH (high–high, capturing diagonal details). Bayesian fusion theory, optimized by the BSO algorithm, effectively fused the feature information of different modalities, enhancing image directional sensitivity and reducing redundant coefficients. This method significantly improved the information richness and diagnostic value of the fused image.

In the study of multimodal medical image fusion, various algorithms and methods are constantly being proposed to improve fusion quality. For example, in the study by Das and Bhattacharya, researchers found that multi-resolution wavelet decomposition provides more image information than single-resolution methods, particularly in image fusion and denoising [42]. Meanwhile, Li et al. proposed a medical image fusion approach based on joint low-rank and sparse decomposition with dictionary learning for image denoising and enhancement [43]. This method's dictionary learning model includes low-rank and sparse regularization terms, constructing the fused image by combining the low-rank and sparse

Table 2. Comparison of learning methods in multimodal medical image fusion

Method Category	Representative Models	Key Advantages	Typical Applications
Traditional ML Methods	SVM, RF, KNN, PCA	Interpretable, easy to implement, effective for small datasets	Early-stage studies, simple or low-noise image fusion
CNN-based Methods	U-Net, FusionNet	Preserves spatial and texture features, supports hierarchical learning	Fusion of structurally aligned images (e.g., CT+MRI)
GAN-based Methods	FusionGAN, CycleGAN	High-quality outputs, handles nonlinear relationships	High-detail, cross-modality image fusion (e.g., PET+MRI)
Autoencoder-based Methods	Fusion AE, VAE	Good for unsupervised learning, extracts latent representations	Fusion tasks with limited or no annotations
Hybrid Deep Models	Wavelet-CNN, SwinFusion	Multi-scale extraction, blends transform and deep features	Structural reconstruction, diagnosis-oriented fusion
Emerging Methods	Transformer-based	Strong global context modeling, adaptable across modalities	Large-scale fusion, complex multimodal integration

Support vector machines (SVM), random forests (RF), K-nearest neighbors (KNN) and principal component analysis (PCA). Variational Autoencoder (VAE).

components of the source image, thereby improving image quality.

Learning-based methods

Learning-based MMIF methods use machine learning and deep learning techniques to learn features and patterns from large datasets for efficient image fusion. These methods can automatically extract complex features, offer strong generalization performance, and adapt flexibly to the fusion needs of different types of images. Prior to the widespread use of deep learning, traditional machine learning methods such as support vector machines, random forests, K-nearest neighbors and PCA were commonly used. These methods achieved effective information integration through statistical analysis, feature optimization, and pixel-level fusion strategies. However, as the complexity and dimensionality of medical imaging data grew significantly, traditional methods began to show limitations in feature representation and computational efficiency. This has driven the advancement of deep learning technology.

In recent years, deep learning has gained significant traction in fields like text processing, computer vision, and natural language processing, making MMIF more efficient and intelligent [44]. Deep learning techniques, such as convolutional neural networks (CNNs), generative adversarial networks (GANs), and autoencoders, have proven particularly effective. CNNs, such as FusionNet and U-Net, extract local features from medical images via multilayer convolution operations, making them highly suitable for fusion tasks like CT-MRI and PET-MRI [45-49]. GANs, such as FusionGAN and CycleGAN, generate realistic, detailed images through adversarial training, effectively addressing complex, nonlinear fusion challenges. Autoencoder frameworks, like Fusion Autoencoder and Variational Autoencoder, excel at latent feature extraction, compression, and high-quality reconstruction, particularly in unsupervised learning scenarios. Additionally, hybrid models like Wavelet-CNN combine traditional transform-domain techniques with deep learning for improved multiscale feature extraction and enhanced fusion efficiency. Their characteristics and applicability are summarized in **Table 2**.

Li et al. introduced a novel deep learning architecture designed specifically for infrared and visible image fusion [50]. Unlike traditional convolutional networks, their encoding network consists of convolutional layers, fusion layers, and dense blocks, where the output of each layer connects to all other layers. This architecture aims to extract more useful features from source images during encoding and uses two fusion layers (fusion strategies) to merge these features. Finally, the decoder reconstructs the fused image, optimizing information integration before reconstruction and improving feature extraction during encoding.

To further enhance the detail quality and realism of the fused image, the Generative Adversarial Network (GAN) is used to achieve high-quality medical image fusion through adversarial training between the generator and discriminator [45-48, 50-53].

Despite the impressive results achieved by current fusion methods, there are still areas for improvement. For example, GAN-based fusion of infrared and visible light images is time-consuming and lacks ground truth, making it difficult to design a comprehensive input for the discriminator to perform high-level tasks. Two existing solutions to this problem involve using visible images as references or learning a generative model through a dual-discriminator approach, where both infrared and visible images serve as input [54, 55]. To address these challenges, Hou et al. proposed a semantic segmentation-based infrared and visible light image fusion GAN (SSGAN) [56]. This method is a deep learning-based, end-to-end fusion network that does not require manual rule design. It uses semantic masks to separate images into foreground (salient semantic features) and background (fine texture details). The foreground represents salient semantic objects, while the background contains finer texture details. Distinct generator pathways extract the semantic features, while the discriminator combines the infrared foreground and visible background to generate images that preserve both semantic clarity and textural detail.

The instability of GANs has prompted the exploration of Transformer architectures. The Transformer's attention mechanism excels at capturing global multimodal relationships and has demonstrated great potential in accurately integrating cross-modal features. With its success in natural language processing and computer vision, researchers have started to explore its application in the field of medical image fusion.

Current fusion rules employed by convolution-based networks for multimodal medical image fusion are often not fine-grained enough. Deep learning-based methods generally concatenate source images or depth features from different modalities and input them into neural networks, which may lead to insufficient utilization of the source image information and limit the comprehensive use of available modalities. To address this, Zhang et al. proposed Transfusion, a transformer-based end-to-end medical image fusion framework [57]. This innovative framework incorporates the Transformer as a fusion strategy, utilizing its self-attention mechanism to integrate the global contextual information of multimodal features and fully fuse them. Additionally, the designed branch networks interact through multi-scale fusion transformers, enabling better utilization of information from different modalities.

Simultaneously, self-supervised learning (SSL) has been developing rapidly [58-62]. SSL does not require a large amount of manually annotated data but instead leverages the structural information within the data to train models and enhance cross-modal alignment and feature representation.

Fang et al. proposed a multimodal brain tumor segmentation framework that utilizes hybrid fusion of modality-specific features via a self-supervised learning strategy, employing fully convolutional networks [63]. Their multi-input architecture, along with the novel hybrid attention fusion scheme, learns independent features from multimodal data, adapts to varying numbers of multimodal inputs, and captures the correlation between them through the attention mechanism.

Inspired by contrastive learning, Zhang et al. developed an SSL framework based on contrastive autoencoding and convolutional information exchange for multimodal medical fusion tasks [64]. Since extracting features from source images in pairs often causes information redundancy, the researchers combined Transformers and CNNs in parallel as autoencoders. This combination preserves both global and local features based on prior knowledge and uses a contribution estimation model to assess the optimal fusion contribution of paired source images. The Transformer-CNN hybrid encoder and contribution estimation model treat paired source images as positive and negative results of reconstructed images, improving fusion quality by mitigating information redundancy through refined feature exchange.

Currently, self-supervised MMIF methods pretrain networks using large natural image datasets via the original image reconstruction task, then achieve image fusion by integrating fusion rules into the trained network. However, they often overlook the domain differences between

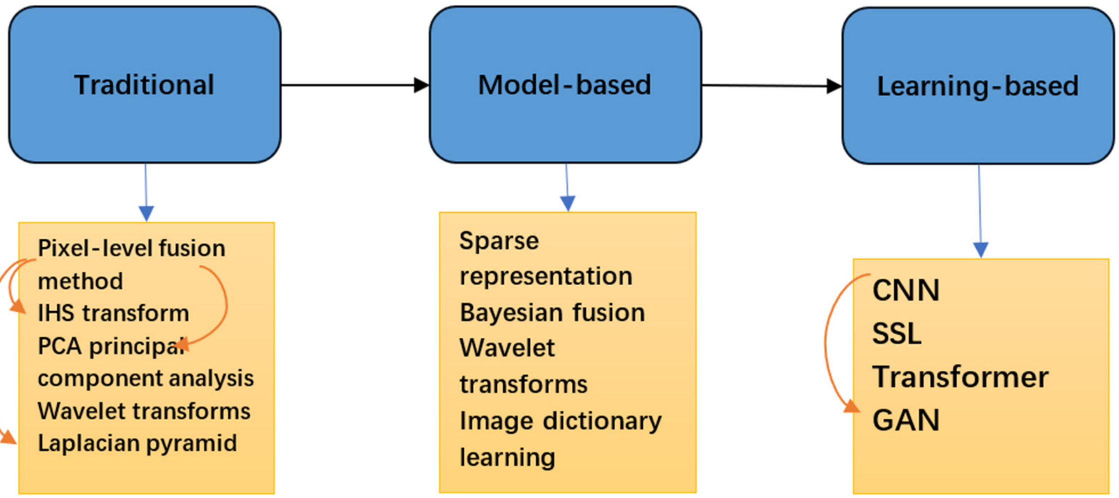

Figure 5. Three fusion approach categories: Traditional methods provide a foundation for model methods, which in turn provide a development foundation for learning methods. "Laplacian Pyramid", "IHS" and "PCA" are pixel-level fusion technologies; GAN is developed based on CNN.

the training and test data. Additionally, most existing MMIF methods are built upon traditional CNN architectures, limiting their ability to model long-range contextual relationships in the source modalities. To overcome this, Zhang et al. proposed a self-supervised transformer for MMIF [65]. Under this framework, random image super-resolution (RSR) is used to train the network. Both RSR and MMIF share similar multi-resolution inputs and objectives: image restoration and enhancement, which helps reduce domain discrepancies between the training and test data. Furthermore, a transformer-based fusion network is introduced to capture both local and global information from the source modalities. By leveraging RSR as a pretext task aligned with fusion objectives, this model reduces domain differences and effectively captures both global and local modality information.

In summary, image fusion methods can be classified into three categories: traditional methods, model-based methods, and learning-based methods. Each method has distinct characteristics and application scenarios, with developmental relationships among them. As shown in **Figure 5**, traditional image fusion methods (e.g., Laplacian pyramid, IHS transform, PCA) are primarily based on pixel-level operations or direct transform-domain processing, laying the theoretical and technical foundation for model-based methods. These model-based methods continue to evolve, ultimately driving the development of learning-based fusion technologies, with deep learning as the core. **Table 3** lists some fusion techniques.

With the continuous growth of medical imaging data and advancements in computing power, learning-based methods are moving toward smarter and more efficient solutions. Future research will likely focus on multimodal feature alignment across different domains, improving the interpretability of fusion methods, and leveraging large-scale pretrained medical imaging models. The development of these technologies will not only enhance diagnostic efficiency but also play a key role in areas such as telemedicine, disease detection, and personalized treatment.

Image fusion quality assessment metrics

The evaluation of MMIF quality is crucial to ensure that the fusion results meet clinical diagnostic needs. Fusion quality assessment methods are typically divided into two categories: subjective and objective evaluation [66]. Subjective evaluation, as the name suggests, relies on human judgment, often performed by clinical experts. It typically involves comparing the fused image with the original images to score the fusion result. The advantage of subjective evaluation is that it directly reflects the practical utility of the fused image and the observer's intuitive assessment. However, it has limitations, including dependence on per-

Table 3. Comparison of several methods

Method Name	Type	Advantages	Disadvantages	Computational Complexity	Image Quality
IHS	Traditional	Intuitive, low computational cost	Spectral loss, unstable for high dimensions	Low	Medium
PCA	Traditional	Dimensionality reduction, feature extraction	Color distortion, sensitive to noise	Medium	Medium
Wavelet Transform	Traditional	Multi-resolution, image detail preservation	High complexity, long processing time	High	High
GAN	Learning-based	High quality image generation, handles complex tasks	Unstable training, requires large data	High	High
CNN	Learning-based	Automatic feature extraction, strong recognition	High computational cost, long training	High	High
Transformer	Learning-based	Integrates multimodal data, self-attention mechanism	Long training, complex model	High	High
Sparse Representation	Model-based	Effective feature representation, robust to noise	Needs high-quality dictionary, complex computation	Medium	Medium

sonal experience and subjective judgment, variability between different observers, and being time-consuming, labor-intensive, and difficult to quantify. As a result, objective evaluation indicators are commonly introduced for automated and repeatable quality assessment. Objective evaluation analyzes the quality of the fused image from multiple perspectives by examining the statistical and structural features of both the fused and source/reference images.

Common objective evaluation indicators focus on different aspects, including the (SSIM), peak signal-to-noise ratio (PSNR), mutual information (MI) and cross entropy (CE).

(1) SSIM assesses the similarity of image structure, brightness, and contrast. Its value range is from 0 to 1, with 1 indicating that the structure of the fused image closely matches that of the original image. The formula is as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

(2) PSNR measures the difference in pixel intensity between the fused image and the reference image. A higher PSNR value indicates better image quality, as it signifies a lower level of distortion between the fused and reference images. The formula is as follows:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (8)$$

(3) MI measures the amount of information the fused image retains from the two source images. A larger MI value indicates that the fused image contains more of the original information. The formula is as follows:

$$MI(X,Y) = \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \quad (9)$$

(4) CE evaluates the similarity between the light intensity distribution of the fused image and the source image. A smaller cross-entropy value indicates less information loss [67].

In general, the above objective indicators can quantify fusion quality to some extent, but each focuses on different aspects, making it challenging to fully capture the subjective quality evaluation criteria of the human eye. To improve the consistency between objective evaluation and human perception, recent research has proposed enhanced fusion quality assessment methods. For instance, Wang et al. proposed a super-pixel-level structural similarity evaluation index system [68]. This method first performs superpixel segmentation on the image, dividing it into smaller areas that are perceptually uniform. It then extracts features from the source and fused images within the corresponding superpixel areas to calculate regional structural similarity. This approach helps reduce errors caused by direct pixel comparison of images from different modalities, enabling a more accurate assessment of how well the fused image retains the source image's structural information. Compared to traditional pixel-by-pixel SSIM, the superpixel-based evaluation is less sensitive to registration errors and offers greater accuracy in fusion quality assessment.

In the future, by integrating human visual perception models with artificial intelligence technologies, fusion quality evaluation methods will more accurately and comprehensively reflect the strengths and weaknesses of fusion results, providing stronger support for the advancement of MMIF technology.

Discussion

MMIF effectively integrates complementary features from different imaging modalities, overcoming the limitations of single-modality imaging and enhancing diagnostic accuracy. Initially, traditional fusion techniques dominated, focusing mainly on spatial and frequency domain methods. However, these methods have limitations in preserving information and accurately representing key diagnostic features. As a result, model-based methods gradually became mainstream, offering greater flexibility and robustness in processing complex image data and integrating probabilistic information, thereby improving diagnostic accuracy.

In recent years, deep learning methods have driven further progress in MMIF. Architectures based on CNNs, GANs, and Transformers have significantly improved MMIF performance with their powerful feature extraction and adaptive learning capabilities. CNNs excel in capturing multi-scale features through their hierarchical learning structure. GANs enhance fusion quality by synthesizing high-fidelity images, bridging modality differences, and improving visual clarity. Transformers, known for their excellent feature extraction and global context modeling, handle long-range dependencies and enhance modality alignment.

Despite the ongoing development of MMIF techniques, challenges remain, including addressing cross-modality differences, optimizing algorithms for efficiency, and achieving clinical applicability. Continued research is crucial for improving these technologies, optimizing fusion performance, and ultimately enabling clinical application.

Looking ahead, MMIF technology holds broad application potential. Possible future development directions include:

- Improving interpretability is key to enhancing user trust and optimizing clinical decision-making. With growing demands for transparency in AI products, enhancing the interpretability of deep learning models is crucial for clinical decision support. Future developments should focus on explaining model decisions. For instance, incorporating attention mechanisms could highlight specific regions of medical images prioritized during fusion, providing visual cues for clinicians. Additionally, feature visualization techniques (e.g., Grad-CAM or saliency maps) could be used to display internal feature representations, clarifying how the model processes multimodal data.

- Future research should aim to improve fusion accuracy by addressing modality differences, enhancing feature alignment, and reducing information redundancy. Researchers could design custom loss functions (e.g., mutual information-based or redundancy-penalizing losses) to minimize overlapping or irrelevant information during fusion.

- The focus is on developing computationally efficient and lightweight models for clinical usability. However, bottlenecks such as high computational complexity, slow inference speed, and resource-intensive models limit real-time clinical applications. Researchers should consider applying model compression techniques, such as pruning (removing redundant neurons) and quantization (reducing weight precision), to create lightweight models without compromising accuracy. Additionally, utilizing advanced hardware or optimizing algorithms with efficient convolution operations and approximate complex functions can help, with custom implementations tailored to clinical hardware limitations.

- To address data scarcity and labeling costs, self-supervised learning and synthetic data augmentation methods are increasingly employed, enabling robust model training without large labeled datasets. For example, self-supervised learning can be used to pre-train a model on an unlabeled multimodal dataset and then fine-tune it on a smaller labeled dataset.

• Research should focus on developing general models that can adapt to various medical imaging modalities and clinical scenarios. This includes applying transfer learning to adapt pre-trained models to new modalities and developing multi-task learning frameworks to enhance the adaptability and versatility of MMIF models.

Overall, continuous innovation and interdisciplinary collaboration are crucial to advancing multimodal medical image fusion technology. Combining medical image fusion technology with genomics, physiological signal monitoring and big data analysis will further promote the implementation of personalized diagnosis and treatment strategies, achieve early diagnosis and precise treatment of diseases, and significantly improve the accuracy of clinical diagnosis and the efficiency of medical decision-making.

Conclusion

The rapid advancement of multimodal image fusion technology has greatly benefited the medical field and human well-being. This paper summarizes the evolution of MMIF technology, beginning with an introduction to mainstream imaging technologies, followed by an analysis of the advantages and limitations of single-modality technologies. It then introduces multimodal image fusion technology, followed by an overview of three MMIF methods: traditional methods, model-based methods, and learning-based methods, with a particular focus on the application of deep learning in recent years. Finally, performance evaluation indicators for multimodal fusion are analyzed, and suggestions for future research directions are provided.

Author contributions: Zhaopeng Zhou: Provide medical theory background support, guide research direction, and participate in discussions on review framework and writing details. Miao Zhou: Provide medical theory background support, participate in the discussion of the review framework and writing details. Wenhui Guo: Guide the overall conception of the review and preside over the revision of the paper. Yongchu Hu: Responsible for the overall conception and review of the paper, and complete the revision and finalization of the paper.

References

[1] George ASH, Shahul A, George AS. Artificial Intelligence in Medicine: A New Way to Diagnose and Treat Disease. Partn Univers Int Res J 2023;2(3):246-259.

[2] Caraiani C, Dong Y, Rudd AG, et al. Reasons for inadequate or incomplete imaging techniques. Med Ultrason 2018;20(4):498-507.

[3] Huang B, Yang F, Yin M, et al. A Review of Multimodal Medical Image Fusion Techniques. Comput Math Methods Med 2020;2020:8279342.

[4] Karim S, Tong G, Li J, et al. Current advances and future perspectives of image fusion: A comprehensive review. Inform Fusion 2023; 90:185-217.

[5] Du J, Li W, Lu K, et al. An overview of multimodal medical image fusion. Neurocomputing 2016;215:3-20.

[6] Sossi V. Multi-modal Imaging and Image Fusion. In: Kiessling F, Pichler BJ, editors. Small Animal Imaging: Basics and Practical Guide. Berlin, Heidelberg. Springer Berlin Heidelberg, 2011. p. 293-314.

[7] Li Y, El Habib Daho M, Conze PH, et al. A review of deep learning-based information fusion techniques for multimodal medical image classification. Comput Biol Med 2024;177:108635.

[8] Haribabu M, Guruviah V, Yogarajah P. Recent Advancements in Multimodal Medical Image Fusion Techniques for Better Diagnosis: An Overview. Curr Med Imaging 2023;19(7):673-694.

[9] Werner P, Barthel H, Drzezga A, et al. Current status and future role of brain PET/MRI in clinical and research settings. Eur J Nucl Med Mol Imaging 2015;42(3):512-526.

[10] Wang SY, Chen XX, Li Y, et al. Application of Multimodality Imaging Fusion Technology in Diagnosis and Treatment of Malignant Tumors under the Precision Medicine Plan. Chin Med J (Engl) 2016;129(24):2991-2997.

[11] Schwenck J, Sonanini D, Cotton JM, et al. Advances in PET imaging of cancer. Nat Rev Cancer 2023;23(7):474-490.

[12] Hussain S, Mubeen I, Ullah N, et al. Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review. Biomed Res Int 2022;2022:5164970.

[13] Fendler WP, Eiber M, Beheshti M, et al. PSMA PET/CT: joint EANM procedure guideline/SNMMI procedure standard for prostate cancer imaging 2.0. Eur J Nucl Med Mol Imaging 2023;50(5):1466-1486.

[14] Crișan G, Moldovean-Cioroianu NS, Timaru DG, et al. Radiopharmaceuticals for PET and SPECT Imaging: A Literature Review over the Last Decade. Int J Mol Sci 2022;23(9):5023.

[15] Greffier J, Villani N, Defez D, et al. Spectral CT imaging: Technical principles of dual-energy CT and multi-energy photon-counting CT. Diagn Interv Imaging 2023;104(4):167-177.

[16] Piovesan A, Vancauwenberghe V, Van De Looverbosch T, et al. X-ray computed tomography for 3D plant imaging. Trends Plant Sci 2021;26(11):1171-1185.

[17] Treglia G, Muoio B, Trevisi G, et al. Diagnostic Performance and Prognostic Value of PET/CT with Different Tracers for Brain Tumors: A Systematic Review of Published Meta-Analyses. Int J Mol Sci 2019;20(19):4669.

[18] Wang J, Xu Y, Boppart SA. Review of optical coherence tomography in oncology. J Biomed Opt 2017;22(12):1-23.

[19] Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023;10(1):1.

[20] Demner-Fushman D, Antani S, Chachra S, et al. Open-i: Imaging, informatics, natural language processing, and multi-modal information retrieval. 2016.

[21] Liñares-Blanco J, Pazos A, Fernandez-Lozano C. Machine learning analysis of TCGA cancer data. PeerJ Comput Sci 2021;7:e584.

[22] Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26(6):1045-1057.

[23] Marcus DS, Wang TH, Parker J, et al. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 2007;19(9):1498-1507.

[24] Jiang S, Ashar P, Shandhi MMH, et al. Demographic reporting in biosignal datasets: a comprehensive analysis of the PhysioNet open access database. Lancet Digit Health 2024; 6(11):e871-e878.

[25] Xie Y, Zhou C, Gao L, et al. MedTrinity-25M: A Large-scale Multimodal Dataset with Multigranular Annotations for Medicine. ArXiv 2025; arXiv:2408.02900.

[26] Dai W, Chen P, Lu M, et al. CLIMB: Data Foundations for Large Scale Multimodal Clinical Foundation Models. ArXiv 2025;arXiv:2503.07667.

[27] Azam MA, Khan KB, Aqeel M, et al. Analysis of the MIDAS and OASIS Biomedical Databases for the Application of Multimodal Image Processing. Intell Technol Appl 2020;581-592.

[28] Mohammed HA. The image registration techniques for medical imaging (MRI-CT). Am J Biomed Eng 2016;6(2):53-58.

[29] El-Gamal FE-ZA, Elmogy M, Atwan A. Current trends in medical image registration and fusion. Egypt Inf J 2016;17(1):99-124.

[30] Princess MR, Kumar VS, Begum M. Comprehensive and Comparative Study of Different Image Fusion Techniques. Int J Adv Res Electr, Electron Instrum Energy 2014;3:11800-11806.

[31] Azam MA, Khan KB, Salahuddin S, et al. A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. Comput Biol Med 2022;144: 105253.

[32] Kaur H, Koundal D, Kadyan V. Image Fusion Techniques: A Survey. Arch Comput Method E 2021;28(7):4425-4447.

[33] Daniel E. Optimum Wavelet-Based Homomorphic Medical Image Fusion Using Hybrid Genetic–Grey Wolf Optimization Algorithm. IEEE Sens J 2018;18(16):6804-6811.

[34] Li S, Kang X, Fang L, et al. Pixel-level image fusion: A survey of the state of the art. Information Fusion 2017;33:100-112.

[35] Basu S, Singhal S, Singh D. A Systematic Literature Review on Multimodal Medical Image Fusion. Multimed Tools Appl 2024;83(6):15845-15913.

[36] Khan SU, Khan F, Ullah S, et al. Multimodal medical image fusion in NSST domain with structural and spectral features enhancement. Heliyon 2023;9(6):e17334.

[37] Chen CI. Fusion of PET and MR Brain Images Based on IHS and Log-Gabor Transforms. IEEE Sens J 2017;17(21):6995-7010.

[38] Haddadpour M, Daneshvar S, Seyedarabi H. PET and MRI image fusion based on combination of 2-D Hilbert transform and IHS method. Biomed J 2017;40(4):219-225.

[39] Shabanzade F, Khateri M, Liu Z. MR and PET Image Fusion Using Nonparametric Bayesian Joint Dictionary Learning. IEEE Sens Lett 2019;3(7):1-4.

[40] Bhardwaj J, Nayak A. Haar wavelet transform-based optimal Bayesian method for medical image fusion. Med Biol Eng Comput 2020; 58(10):2397-2411.

[41] Nayak A, Bhardwaj J. Discrete Wavelet Transform and Bird Swarm Optimized Bayesian Multimodal Medical Image Fusion. Helix 2020; 10(01):7-12.

[42] Arpita D, Bhattacharya M. Evolutionary algorithm based automated medical image fusion technique: Comparative study with fuzzy fusion approach. 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC) 2009; 269-274.
[43] Li H, He X, Tao D, et al. Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. Pattern Recognit 2018;79:130-146.
[44] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436-444.
[45] Bhatnagar G, Wu QMJ, Liu Z. Directive Contrast Based Multimodal Medical Image Fusion in NSCT Domain. IEEE Trans Multimedia 2013; 15(5):1014-1024.
[46] Gupta J, Pathak S, Kumar G. Deep Learning (CNN) and Transfer Learning: A Review. Journal of Physics: Conference Series 2022;2273(1): 012029.
[47] Salehi AW, Khan S, Gupta G, et al. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability 2023;15(7):5930.
[48] Bhatt D, Patel C, Talsania H, et al. CNN variants for computer vision: History, architecture, application, challenges and future scope. Electronics 2021;10(20):2470.
[49] Liu JM, Sun HM, Katto J. Learned image compression with mixed transformer-cnn architectures. ArXiv 2023;arXiv:2303.14978.
[50] Li H, Wu XJ. DenseFuse: A Fusion Approach to Infrared and Visible Images. IEEE Trans Image Process 2019;28(5):2614-2623.
[51] Karthika S, Durgadevi M. Generative Adversarial Network (GAN): a general review on different variants of GAN and applications. 2021 6th International Conference on Communication and Electronics Systems (ICCES) 2021;1-8.
[52] Andresini G, Appice A, De Rose L, et al. GAN augmentation to deal with imbalance in imaging-based intrusion detection. Future Gener Comp Syst 2021;123:108-127.
[53] Lim W, Yong KSC, Lau BT, et al. Future of generative adversarial networks (GAN) for anomaly detection in network security: A review. Comput Secur 2024;139:103733.
[54] Ma J, Yu W, Liang P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion. Inform Fusion 2019;48:11-26.
[55] Ma J, Xu H, Jiang J, et al. DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion. IEEE Trans Image Process 2020;29:4980-4995.
[56] Hou J, Zhang D, Wu W, et al. A Generative Adversarial Network for Infrared and Visible Image Fusion Based on Semantic Segmentation. Entropy 2021;23:376.
[57] Zhang J, Liu A, Wang D, et al. Transformer-Based End-to-End Anatomical and Functional Image Fusion. IEEE Trans Instrum Meas 2022;71:1-11.
[58] Zhai X, Oliver A, Kolesnikov A, et al. S4L: Self-Supervised Semi-Supervised Learning. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) 2019;1476-1485.
[59] Liu X, Zhang F, Hou Z, et al. Self-Supervised Learning: Generative or Contrastive. IEEE Trans Knowl Data Eng 2023;35(1):857-876.
[60] Jaiswal A, Babu AR, Zadeh MZ, et al. A survey on contrastive self-supervised learning. Technologies 2020;9(1):2.
[61] Hendrycks D, Mazeika M, Kadavath S, et al. Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty. ArXiv 2019;arXiv:1906.12340.
[62] Chen T, Kornblith S, Swersky K, et al. Big self-supervised models are strong semi-supervised learners. Adv Neural Inf Process Syst 2020;33:22243-22255.
[63] Fang F, Yao Y, Zhou T, et al. Self-Supervised Multi-Modal Hybrid Fusion Network for Brain Tumor Segmentation. IEEE J Biomed Health Inf 2022;26(11):5310-5320.
[64] Zhang Y, Nie R, Cao J, et al. Self-Supervised Fusion for Multi-Modal Medical Images via Contrastive Auto-Encoding and Convolutional Information Exchange. IEEE Comput Intell Mag 2023;18(1):68-80.
[65] Zhang J, Liu Y, Liu A, et al. Multimodal Image Fusion via Self-Supervised Transformer. IEEE Sens J 2023;23(9):9796-9807.
[66] Xun S, Li Q, Liu X, et al. Charting the path forward: CT image quality assessment - an in-depth review. J King Saud Univ-Com 2025; 37(5):92.
[67] Mao A, Mohri M, Zhong Y. Cross-Entropy Loss Functions: Theoretical Analysis and Applications. Proceedings of the 40th International Conference on Machine Learning 2023;202: 23803-23828.
[68] Wang E, Yang B. Superpixel-level Structural Similarity Metric for Image Fusion Quality Assessment. 2020 7th International Conference on Information Science and Control Engineering (ICISCE) 2020;1371-1376.