The International Journal of Engineering and Information Technology

journal homepage:www.ijeit.misurata.edu.ly

MISURATA UNIVERSITY

A Survey on Deep Learning Techniques for Medical Image Fusion

Saad, A.
Computer Technology Tripoli (Tripoli – Libya)
amal.omar@cctt.edu.ly

Fgee, E.
University of Gharyan (Gharyan-Libya)
bahlulfgee@yahoo.com

Abstract— **Medical Image Fusion is a process that involves combining information from multiple medical images, which is essential for healthcare applications like diagnosis, treatment planning, and image-guided interventions. Deep learning techniques have shown significant promise in medical image fusion by integrating information and effectively capturing data from diverse modalities. Additionally, Data augmentation techniques have come to be an important tool for improving model performance and generalization. The objective of this paper is to give a comprehensive overview of deep learning techniques and data augmentation methods utilized in medical image fusion from 2018 to 2023. The paper covers a variety of topics such as image registration, feature extraction, fusion architectures, data augmentation techniques, and evaluation metrics. The survey also discusses the challenges, limitations, and future directions in the field.**

Index Terms—**medical image fusion, data augmentation, convolutional neural networks, generative adversarial networks, transfer learning.**

I. INTRODUCTION

Medical Image Fusion (MIF) is a process that creates a single image called a fused image by integrating two or more medical images. The images can be of a single type or multiple types (multimodality) [1, 2]. The fused image contains complementary information from each input image. There are two types of medical image modalities, anatomical and functional. The human body has a rich anatomical structure that can be captured through anatomical images. For example, computed tomography (CT) and magnetic resonance imaging (MRI), Single-photon emission computed tomography (SPECT). CT is effectively useful for imaging soft tissues, the low density of protons in bone tissue makes the bone image from MRI unclear. MRI is particularly effective for examining soft tissues, such as the brain and spinal cord [3]. Therefore, the biggest challenge in the medical field is to accurately identify diseases and provide better treatment. The accuracy of diagnosis can be improved by using multiple medical images through image fusion to increase the amount of information. MIF is an important technique in healthcare that is used in multiple clinical applications including diagnosis, medical assessment, treatment planning, and surgical operations. Using CT and MRI images provides both anatomical and functional information, leading to more accurate diagnosis and treatment planning. Likewise, by combining MRI and Positron Emission Tomography (PET) images, it is possible to visualize structural and metabolic information, which helps in the identification and characterization of tumors [4]. Recently, Deep Learning (DL) has been the latest paradigm shift in all fields. In the MIF field, DL methods have more powerful feature extraction capabilities than traditional methods. Convolutional neural networks (CNNs) algorithms have demonstrated remarkable abilities in feature extraction and data representation, making them suitable for MIF tasks. A multi-layer concatenation fusion network (MCF-Net) built by Liang et al [5] employs CNNs to retrieve features from diverse sources and then fuse them. This study shows how CNNs are capable of extracting informative features for MIF as an end-to-end Deep learning model. Another approach for medical image fusion is the generative adversarial networks (GANs). The model is a GAN proposed by Ma and colleagues in 2018 for combining infrared and visible images [6]. To combine the input images, a generator network is employed, and a discriminator network is utilized to distinguish between the fused image and the ground truth. The generator network is conditioned to reduce adversarial loss and mean squared error between the fused image and the ground truth. This assumes that the two sets of images contain complementary information that can be combined to produce a more informative and visually pleasing image. This study highlights the potential of GANs in enhancing the quality and accuracy of fused images. The use of DL techniques for MIF still involves some challenges, e.g., the lack of diverse datasets because of the difficulty of acquiring image datasets from medical imaging centers due to the privacy of patient data and the difficulty of getting a different type of images for the same patient. Furthermore, there is a lack of labeled datasets and difficulties in the interpretability and

Received 05 Apr , 2024; revised 06 May, 2024; accepted 15 Mar 2024.
Available online 08 Aug, 2024.

explain-ability of DL models. This leads to the need for improved evaluation metrics that align with clinical requirements, as well as the generalization of unseen data, computational complexity, time constraints, and resource requirements. Recent studies have proposed algorithms that prompt the use of data augmentation and transfer learning (TL) to face the challenges. One method is to use augmentations to increase the size of dataset training by inferring new data from the available dataset using the data augmentation technique [7]. In MIF, data augmentation produces new images by applying transformation techniques to the existing images. Another technique to overcome challenges is transfer learning. Transfer learning begins with a model that has already been trained on a large dataset instead of starting from scratch, transfer learning utilizes the learned representations from another model to enhance its performance and efficiency [8]. The purpose of this article is to give a complete overview of the present state of the art in this area. The survey aims to gather and simplify the current literature on the use of deep learning and data augmentation in MIF, with a focus on the achievements, challenges, and potential future direction. The contributions of this survey paper can be summarized as follows:

1. A comprehensive overview: The paper presents the most recent DL techniques used in MIF by focusing on different deep learning models, such as CNNs, and GANs, and their applications and performance in MIF tasks.
2. Analyzing of data augmentation techniques: The paper examines how data augmentation techniques are utilized in MIF, and the focus is on various data augmentation methods and their effect on improving the performance and generalization of DL models for MIF.
3. Evaluation and comparison: Comparing and evaluating the performance of various data augmentation methods and DL techniques for MIF. It discusses the strengths and limitations of each approach.
4. Discuss challenges and future work: Identify the challenges and open research questions in the field of deep learning-based medical image fusion.

The remaining sections of the paper are as follows: Section II provides a summary of research on image registration for MIF. In section III, feature extraction for MIF. In section IV, deep learning architectures for MIF are discussed in detail. In section V, data augmentation techniques for MIF are explained. In section VI evaluation metrics and datasets are presented. In section VII, the Discussion explains various research challenges and limitations. In Section VIII, applications and case studies are discussed. In Section IX, the conclusion is presented.

II. IMAGE REGISTRATION FOR MEDICAL IMAGE FUSION

Image registration [3, 37] involves aligning and matching two or more images of the same scene or object to establish a correspondence between their pixel coordinates. Traditional image registration techniques are categorized into intensity-based methods and feature-based methods. The intensity-based method is the spatial transformation that aligns the images best taking into account factors such as rotation, scale, translation, and deformation [9]. Features based on features such as color gradient, edges, geometric shape and contour, image skeleton, or feature points can be used to create correspondence between input images. [10,11] discuss the limitations of traditional registration techniques. the methods often require creating good transformation parameters, which can be challenging to obtain, especially in cases of significant differences or large deformations between the images. Another limitation is the researcher's reliance on handcrafted features or landmarks, which may not capture all the relevant information in the images and can result in errors or inconsistencies. Additionally, traditional registration techniques can be expensive to compute and time-consuming, especially for large datasets or complex registration tasks. In recent years deep learning techniques have been widely used to address the limitations of traditional image registration. Deep Neural Networks (DNN) [10] involve aligning and matching different medical images to enable comparison, analysis, and fusion of information. Data augmentation techniques in image registration can be beneficial for improving the robustness and generalization of the registration algorithm [11]. The accuracy and efficiency of medical image registration have been improved by building hybrid models from DL techniques and traditional ones that take advantage of both techniques. Fig. 1 shows the most important steps for registration and fusion, the registration and fusion are two main steps for enhancement of medical image. The pre-processing and registration steps are known as Feature extraction in MIF.

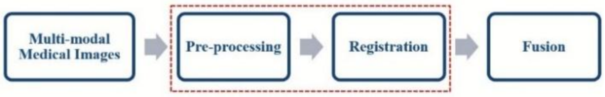

Figure 1. General diagram of the multimodal registration and fusion adopted from [12]

III. FEATURE EXTRACTION FOR MEDICAL IMAGE FUSION

Feature extraction [13] refers to the process of extracting relevant and discriminative features from the source images before fusing. The fusion of images requires this step to capture essential information from each modality and preserve important details. The fusion of images requires this step to capture essential information from each modality and preserve important details. Two categories of feature extraction methods can be broadly classified:

1. Transform-based methods use mathematical transforms such as wavelets, curvelets, and contourlets to extract features from the input images [13]. These are traditional handcrafted feature extraction techniques as well as deep learning-based approaches. The features are extracted from images based on intensity, texture, shape, or statistical properties [1], and they are then combined or fused

using fusion algorithms to generate the fused image manually.

2. Learning-based methods use machine learning algorithms like CNNs to learn features directly from the input images. CNNs utilize deep neural networks to learn hierarchical and abstract representations from the input images automatically. These networks can capture complex patterns and structures in the images by learning discriminative features directly from raw input data [9, 14].

The goal of feature extraction methods is to extract valuable features that represent the unique characteristics of each modality and preserve crucial details for accurate diagnosis and analysis. The methods of extracting features are based on the specific requirements of the fused image task and the characteristics of the input images. Selecting a method that is capable of effectively capturing relevant information and maintaining important features in the fused image is important.

A. Fusion Rules for Extraction Feature in Medical Image Fusion

Fusion rules are the methods or techniques employed to merge data from multiple input images into one fused image [15]. These rules determine how the pixel values or features from the input images are merged to build the fused representation. Fusion rules can be applied based on the specific application and requirements [4, 15], the following are common fusion rules:

1. *Averaging:* This rule determines the average of the pixel values in the input images and assigns it to this corresponding pixel in the fused image. This fusion rule is both simple and widely used.
2. *Maximum selection:* This rule identifies the highest pixel value among the corresponding pixels in the input images and assigns it to the corresponding pixel in the fused image. It is commonly employed in situations where the highest intensity or the most significant features require preservation.
3. *Weighted sum:* This rule combines different weights on the pixel values in the input images. The weights can be calculated by analyzing the quality or reliability of each input image.
4. *Transform-based fusion:* To extract features or coefficients from the input images, this rule employs a transformation or fusion algorithm, like wavelet transform or sparse representation. After combining these features or coefficients, it uses specific fusion rules to generate the fused image. To summarize, there are many techniques and algorithms available in the literature for image fusion. DL techniques for image fusion also utilize the fusion rules.

B. Transfer Learning and Pre-trained Models for Feature Extraction

Transfer learning (TL) is a technique in DL where a pre-trained model is a large dataset trained for a similar task that is used as a starting point for a new model but related task [16]. Models that leverage the knowledge and representations learned from the pre-trained model instead of training them from scratch. This led to significantly improved performance, especially when the new task had limited training data. Another important benefit is that TL can save time and hardware resources [17]. The authors of [18] explained that feature extraction in TL is the pre-trained model used to extract features. The model's weights are frozen, and the input data is passed through the model to obtain the learned features from the intermediate layers. The process of TL involves utilizing a model that has been trained to extract features. The model's weights are frozen, and the input data is routed through it to acquire the learned features from the intermediate layers. Fig. 2 illustrates how these features can be used as input for a new classifier or further processing in a new task. The transfer learning process consists of two main primary steps pre-training and Fine-tuning [8, 19].

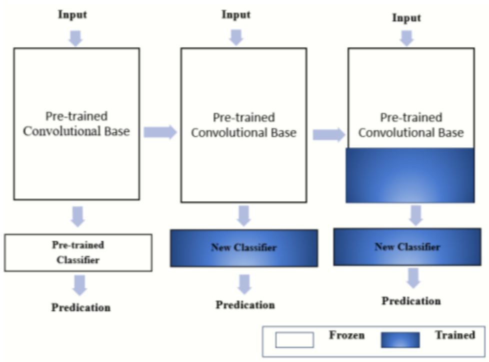

Figure 2. Top-level diagram of Transfer learning from a pre-trained adopted form [20]

Many studies in MIF used TL techniques, one recent study proposed a hybrid method that combined TL and the discrete wavelet transformation (DWT) to merge multi-modal medical images [21]. The proposed approach outperforms other approaches, as confirmed by the experimental results, and the significance of the image that has been fused is determined through qualitative metrics. Several transfer learnings based on CNN architecture have been used for MIF tasks, such as ResNet (Residual Network) [19], and U-Net [21], most of these models learned on the ImageNet dataset. ImageNet dataset has millions of natural images.

IV. MEDICAL IMAGE FUSION USING DEEP LEARNING ARCHITECTURE

The deep learning architectures for MIF are categorized into two types Non-End-To-End and End-to-End image fusion framework based on DL. Non-End-to-End image fusion framework involves multiple steps in the processing of image fusion. Initially, the source images are submitted to a DL network, like CNN or ResNet, to extract features. After that, fusion rules that are based on spatial transforms or decision maps are used to fuse extracted features. Finally, to obtain the final fused image, the fused features must be rebuilt [12]. In the End-to-End image fusion framework, the source images are directly inputted into a DL network, such as a U-Net, which performs both feature extraction and fusion in a single End-to-End process. The network is trained to extract important features from the source images and create the fusion image directly, without the necessity of explicit fusion rules [4]. Deep Learning has succeeded in fusing the medical image by building models that utilize algorithms such as CNNs [4, 20], GANs [4], and Transformer Networks [4, 28]. The use of these algorithms can result in the fusion of images from various modalities or enhance the information in a single modality image. The following explains the techniques used in these algorithms:

A. *Convolutional Neural Networks (CNNs)*

CNN models are used to extract features from images using convolutional layers. CNN has multiple layers of convolutional filters, followed by pools of layers and fully connected layers. The output of every layer is fed into the next layer, resulting in a set of features that can be applied for classification [22]. Medical image fusion can extract features from input images by utilizing basic CNN architectures that remove fully connected layers and use convolutional layers, as demonstrated in Fig. 3. The fused image can be generated by processing and fusing the extracted features [4, 23].

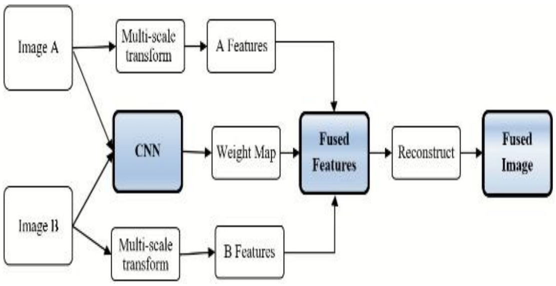

Figure 3.MIF framework based on CNN model adopted from [23]

The limitations of manually designing fusion rules can be overcome by using End-to-End training to learn the appropriate parameters of convolutional filters. Medical image fusion has been commonly performed using various basic CNN architectures. A CNN architecture called Siamese networks has two or more identical subnetworks that share the same weights and architecture. The use of these subnetworks involves processing various inputs in parallel and extracting their respective features [18]. Siamese networks are a popular choice for feature extraction and similarity computation tasks, like image matching, verification, and retrieval. To create Siamese networks, the main goal is to acquire a similarity metric that can measure the similarity or dissimilarity between inputs by establishing a link between weights and architecture.

B. *Generative Adversarial Networks (GANs)*

The initial proposal for Generative Adversarial Networks (GANs) was made in 2014 [23]. DNN frameworks known as GANs can learn from a training dataset and generate new data that matches the training dataset's characteristics [23]. GANs illustrated in Fig. 4 are made up of two neural networks, the generator, and the discriminator, that compete against each other. The generator is trained to generate fake data, and the discriminator is trained to recognize fake data from real examples. If the generator produces false data that the discriminator can easily recognize as implausible, like a faceless image, they will be punished. Over time, the generator acquires the ability to create more plausible examples. High-quality fused images can be generated using adversarial learning in image fusion methods based on GAN. The generator is designed to produce a merged image, the objective of the discriminator is to distinguish between the generated fused image and the actual fused image, while keeping the structural and functional information from the source images.

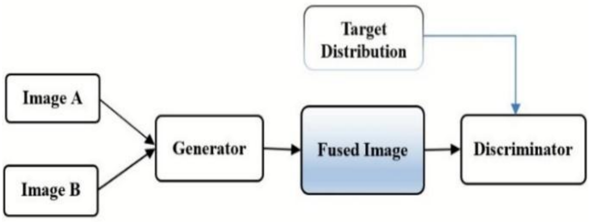

Figure 4. MIF framework based on the GAN model adopted from [23]

GANs can be used for multimodal [24] and multi-scale [23] fusion of images from multiple sources at different scales with high-quality fused images. Conditional generative adversarial network (cGAN) is a type of GAN that incorporates extra conditioning information to guide the image generation process. The generator of fused images can be conditioned on both the source images and additional information related to the fusion task using cGANs. One popular cGAN-based method for image-to-image translation and fusion is Pix2Pix. Pix2Pix was proposed by Isola et al [20] by using a cGAN architecture to learn the mapping between input and output images. Section VII will present the successful application of GANs in various MIF applications.

C. *Transformer Networks*

Transformer-based architectures have become popularly used for multimodal fusion tasks due their ability to capture complex relationships and dependencies between different modalities. They used self-attention mechanisms to model the interactions between modalities and regions of interest (ROI) or created fused representations. Each feature is assigned attention weights by this mechanism based on its relevance to other features across modalities. Weighted feature representations are calculated using attention weights, emphasizing important features and suppressing less relevant ones [1].

The application of Transformer Networks (TNs) has been successful in various computer vision tasks, including image fusion. TNs are utilized for capturing long-range dependencies and modeling the relationships between pixels in input images. TNs are constructed using an encoder-decoder architecture that has self-attention mechanisms. The encoder handles processing the input images and extracting their features, and the decoder produces the fused image using the encoded features by attending to different parts of the input images and gathering the relevant information for fusion [1, 3].

V. DATA AUGMENTATION TECHNIQUES FOR MEDICAL IMAGE FUSION

Data augmentation (DA) is a technique used to generate more training samples by creating new modified versions of the original data [25]. This technique is particularly useful when the original dataset is small, imbalanced, poor quality, or lacks diversity. The benefits of data augmentation in DL-based fusion include [8, 28]:

1. *Improved model performance:* Providing more training samples will enable the model to learn from a wider range of variations and patterns in the data.
2. *Increased model robustness:* By exposing the fusion model to different variations and transformations of the input data, data augmentation can improve the model's ability to handle variations and noise in real-world scenarios.
3. *Reduced overfitting:* Data augmentation helps to prevent overfitting by introducing variations in the training data, which reduces the model's reliance on specific features or patterns that may be present in the limited original dataset.
4. *Enhanced generalization:* By increasing the diversity of the training data, data augmentation enables the fusion model to generalize better to unseen data, improving its performance in real-world applications.

Data augmentation methods are categorized into two types, traditional techniques and advanced data augmentation. Traditional data augmentation methods refer to basic image transformation operations that are commonly used to perform geometric and photometric transformations on the training data. These methods include techniques such as rotation, flipping, cropping, scaling, blurring, color volatility, noise injection, and contrast adjustment [8]. however, advanced data augmentation methods include deeply learned augmentation strategies, meta-learning-based augmentation techniques [26], neural style transfer, generative modeling, and neural rendering. The existing DL models can improve their accuracy and generalization performance by utilizing (incorporating) these new images resulting from the data augmentation process.

A. *Domain-Specific Data Augmentation Strategies for Medical Images:*

Domain-specific data augmentation strategies for medical images refer to techniques that are specifically designed and tailored for the unique characteristics and requirements of medical imaging data. These strategies take into account the specific challenges and considerations in medical image analysis, such as limited data availability, class imbalance, and the need for robust and interpretable models. Some examples of domain-specific data augmentation strategies for medical images include:

1. *Patch-based augmentation:* This strategy involves extracting patches (sub-images) from medical images and applying various transformations to augment the dataset. Patch-based augmentation has the potential to increase the variety of training samples and enhance the model's ability to capture local image features [27].
2. *Intensity-based augmentation:* Medical images often exhibit variations in intensity levels due to different imaging modalities, acquisition protocols, and patient conditions.
3. *Intensity-based augmentation techniques:* histogram equalization, contrast adjustment, and intensity normalization, can help standardize the intensity distribution across images and improve the model's robustness to intensity variations [27, 28].
4. *Class-balanced augmentation techniques:* techniques aim to address this issue by generating additional samples for minority classes, thereby improving the model's ability to learn from and accurately classify rare or abnormal cases [28].

VI. EVALUATION MATRICES AND DATASETS

A. *Evaluation Metrics:*

The effectiveness of image fusion algorithms is evaluated through multiple parameter measures, with each method having its advantages. Multimodal MIF evaluation metrics can be divided into qualitative and quantitative metrics [10]:

1. *Qualitative evaluation metrics:* qualitative metrics include visual inspection and subjective assessment of the fused images. The fused images are analyzed by medical professionals with expertise or radiologists using these metrics to evaluate their quality, clarity, and information content. Color, spatial details, image size, and other parameters must also be taken into account when examining the fused image.
2. *Quantitative evaluation metrics*: The objective is to provide objective metrics for the performance of fusion algorithms. These metrics are used to analyze the similarity, information preservation, and spatial consistency of the fused image and its source images. Common quantitative evaluation metrics for multimodal MIF are listed in Table 1.

TABLE 1. Quantitative Evaluation Metrics

No	Name	Describe	Equation	Ref
1	Mutual Information (MI)	The amount of information that has been transferred between the fused and the source images.	$MI_F^{AB} = I_{FA} + I_{FB} \qquad (1)$ $the\ I_F^{AB} = \sum_{a\,\epsilon B} \sum_{b \in B} P\,(A,B)\, log \frac{P(A,B)}{p_{(A)P(B)}} \qquad (2)$ F is the result of combining, and A and B are the two input images.	[40]

2	Structural-Similarity-Index (SSIM)	Taking into account luminance, contrast, and structural information, the fused image and the source images have structural similarities.	$SSIM_{(A,B,F)} = 0.5 * \left(SSIM_{(A,F)} + SSIM_{(B,F)}\right)$ (3) $SSIM_{(A,F)} = \frac{(2\mu A\mu F+C1)(2\sigma AF+C2)}{(\mu_A^2+\mu_F^2+C1)(\sigma_A^2+\sigma_F^2+C1)}$, $SSIM_{(B,F)} = \frac{(2\mu B\mu F + C1)(2\sigma BF + C2)}{(\mu_B^2 + \mu_F^2 + C1)(\sigma_B^2 + \sigma_F^2 + C1)}$ (4)	[4]
3	Peak- Signal-to-Noise-Ratio (PSNR)	The quality of the fused image can be evaluated by comparing it to the source images in terms of signal-to-noise ratio.	$PSNR = 20 * \log_{10} \frac{Imax}{\sqrt{MSE}}$ (5) - I and max stand for the original image and the maximum pixel gray level. - MSE stands for mean square error. I and J are separate images that have been combined	[4]
4	Entropy	Quantifies the amount of information or randomness in the fused image	$EN = -\sum_{i-0}^{L-1} p(i) \log_2 p(i)$ (6) P(i) is the probability corresponding to the grey level i.	[29]

A. *Medical Image Fusion Datasets*

Datasets for deep learning research are collections of labeled or unlabeled data that are used to train, validate, and test deep learning models, these datasets are crucial for developing and evaluating the performance of DL algorithms and models across various domains and tasks.

Datasets for deep learning research are collections of labeled or unlabeled data that are used to train, validate, and test deep learning models, these datasets are crucial for developing and evaluating the performance of DL algorithms and models across various domains and tasks.

To learn and generalize effectively, deep learning models typically require a lot of data. The size, complexity, and specific problem addressed by the datasets used in deep learning research can vary widely. Patient privacy concerns have led to the lack of publicly available datasets in the field of medical image fusion. The researchers suggested that creating more high-quality public datasets for medical image fusion would be beneficial for future research, data augmentation is an important technique as mentioned in this section. Table 2 lists some of the medical image datasets that are multimodal and available. These datasets can be used for research and development purposes.

TABLE 2. Available multi-modal Medical Image datasets

No	Dataset Name	Year	Modality	Disease	Quantity	Ref
1	Harvard Medical School's Whole Brain Altas program	1999	MRI, CT SPECT, PET	Cerebrovascular disease, brain tumor, and Alzheimer's disease	13,000	http://www.med.harvard.edu/AANLIB/
2	**SMI** (Stanford Medical ImageNet)	2010–2017	MRI, CT SPECT, PET	Hypertensive encephalopathy of the brain	12,00000	https://aimi.stanford.edu/medical-imagenet
3	**BraTS** (Multimodal Brain Tumor Segmentation)	2015–2021	MRI (T1, T1 contrast - enhanced, T2)	Ischemic stroke	8000	BraTS Dataset
4	**MM-WHS** (Multimodal Whole Heart segmentation)		MRI, CT	whole heart segmentation	120 pairs	MM-WHS
5	**MSD** (Medical segmentation Decathlon)	2018	MRI, CT, PET	heart, liver, prostate, and brain	2,633 -3D	http://medicaldecathlon.com/dataaws/
6	**TCIA** (Cancer Imaging Archive)	2011-2023	MRI, CT, PET	lung, breast, brain, and prostate cancer.		https://www.cancerimagingarchive.net/

VII. APPLICATIONS AND CASE STUDIES

This section presents several recent case studies and applications for multimodal medical image fusion in diagnosis and treatment planning. Fusion of multi-scale and multi-modal imaging for disease detection, image-guided interventions and surgical navigation. Table 3 shows the name of each case study, the fusion methodology applied, image modality, computational time, fusion result, and the dataset used in the study.

TABLE 3. MIF applications and case studies

No	Study	Modality	Image location	Fusion method (Algorithm)	Computational time	Fusion result	dataset	Notes	Ref
1	**MCFNet** (multi-layer concatenation fusion network) (2019)	CT /MRI	256 × 256-pixel brain	an Encoder-decoder Network (CNN) transformer (loss function is based on MSE loss)	0.66 seconds Reduce time by getting max down-sampling from CT&MRI And Up-sampling to integrate features	Outperform	1- The Whole Brain Altas of Harvard. 2- ILSVRC 2013 ImageNet	optimizing the loss function	[1]

2	**IFCNN** (Image fusion framework based on CNN) (2020)	Multiple input image	-	Fully Connected CNN as three models (2 Conv- layers for extracted features, Fused image, 2 Conv-layer for reconstruct	Good time because they have good hardware resources	Outperform On 4 types of image dataset	RGB-D (NYU-D2) 100,000 pairs of RGB resized to 422x321 and it is augmented to 1,000,000 images	General model, dataset not specific for fusion	[21]
3	**MGMDcGAN** (multi-generator multi-discriminator conditional generative adversarial network) **(2020)**	MRI-PET, CT-SPECT (different resolutions)	brain-hemispheric	2 Models (GAN & cGAN2) GAN 1 (generate First) GCN 2 (cGAN and mask enhancement information for dense structure in the final fused image)	still, achieve comparable efficiency (multi-process but stile in average time)	Outperforms compared with the existing 9 fusion methods	Harvard dataset.	-	[15]
4	**DMC** (Deep multi-cascade fusion with Classifier-based Feature Synthesis for Medical Multi-modal Images) (2021)	CT/MRI MR-T1/ MR-T2 MRI/SPECT	Brain disease 256 x 256 pixels	Neural Network Autoencoder-decoder t-SNE (t-distributed Stochastic Neighbor Embedding)	-	Outperform (qualitative and quantitative)	Whole Brain Atlas (WBA) 74 pairs	Transferred Ms-COCO datasets and pre-trained	[2]
5	TransFuse (2021)	MRI/CT	Tumer Segmentation (Pixels)	Transformer-based architecture (CNN)	-	Superior Performance	-		[39]
6	Multi-modal Fusion of Imaging and Genomic Data (2023)	Genomic	Breast Cancer	DL-Based			Cancer Genome Atlas Spark Dataset		[4]

VIII. DISCUSSION

In this paper, we have provided information covering various aspects of MIF, including data collection, pre-processing, representation, classification models, evaluation methods, and datasets. Some models leveraged include CNN, GAN, and transform network, the CNNs extract features from input medical images, and the CNNs trained on a large labeled dataset of fused images. The big challenges for MIF are hardware resources which lead to building models during some hours or days, and limited datasets. data augmentation techniques tried to solve the limited dataset labeled. Overcome the problem of information loss in deep networks and lack of interpretability, these open problems need research in the future. In general, the following describes challenges and future work for medical image fusion:

A. Challenges and Limitations

This section presents the recent challenges and limitations of MIF based on some criteria as follows:

1. *Interpretability and Explainability fused image:* The fusion model based on DL for fused images must be easy to understand because they have the potential to impact patient care and clinical decision-making. Enhancing interpretability and explainability can be achieved through the use of visualization techniques, attention mechanisms, Layer-wise Relevance Propagation (LRP), and rule-based fusion [30]. The methods used by both approaches aim to determine the regions or features of interest in the most important input images and contribute the most to the fusion results. Saliency maps, Class activation maps, and gradient-based methods are just some examples of visualization techniques. In LRP relevance scores are assigned to input features to assist in understanding the significance of different features and elucidating the model's choices [31].
2. *Generalization and robustness of fusion models:* Challenges and limitations based on the generalization and robustness of fusion models across different datasets and modalities are explained in Table 4.

Table 4. Challenges and limitations based on the generalization and robustness of fusion models

The term	Describe	Ref
Dataset bias	Fusion models that were created for one dataset may not apply to new datasets due to differences in data distribution, imaging protocols, and acquisition parameters.	[33]

Modality mismatch	Fusion models that were trained in one modality may not perform well when applied to a different modality due to the specific characteristics and imaging principles of each modality.	[19]
Limited training data	To learn representative features and generalize well, DL-based fusion models need a lot of labeled training data. It can be a challenge to obtain labeled data for fusion tasks, especially in medical imaging, as it requires expert annotations and privacy concerns.	[34]
Lack of ground truth	It is difficult to objectively evaluate the generalization and robustness of fusion models across different datasets and modalities without ground truth	[30]

3. *Data augmentation limitations in MIF*: As explained by authors [32] these include limited availability of datasets, interpretability issues, computational complexity, and domain shift. A domain shift can occur between augmented training data and real-world testing data due to medical images not being fully captured by technicians. Applying this to unseen medical images can affect the performance of the fusion model.

In addition to these challenges, hardware resources are considered one of the most important obstacles to completing MIF because the model algorithms require a computer with high specifications in terms of processing speed and storage capacity. MIF algorithms require a graphics processor unit (GPU) and a graphics card with high specifications, which can be expensive.

B. *Future Directions*

Future work in multimodal medical image fusion will aim at addressing issues such as the lack of objective evaluation metrics to assess the quality and usefulness of fused images accurately. The creation of evaluation metrics that match human perception and subjective measures from experienced observers is crucial, particularly in medical diagnosis tasks. Additionally, the challenges of using deep learning networks for image fusion include the computational cost and data requirements which need to be addressed in future work by developing lightweight fusion networks that reduce computational overhead and dependence on large datasets. Creating large datasets that cover different modalities, pathologies, and imaging protocols is necessary for publicly available benchmark datasets. Therefore, extending the use of multi-modal image fusion to different clinical applications is a crucial area for future work. For example, diagnosis, treatment planning, and image-guided interventions. The effectiveness of fusion methods must be validated in different clinical scenarios and patient populations. On the other hand, regarding the development of interpretable and explainable deep learning fusion models, there are some recent studies on the future directions in this part, one of them is a research paper that has been published in the Journal of Intelligent Manufacturing identifies an integrated method that merges a deep learning object detection model, a clustering algorithm, and a similarity algorithm to produce an automated detection process that can be explicated [35]. According to the study, the proposed method addresses multiple challenges that are posed by automated inspection and digital transformation. Another study [36] offers a guide for novices to explainable deep learning, three easy dimensions defined by the guide, define the space of foundational methods that contribute to explainable deep learning, as well as possible future. Future work in data augmentation focuses on advanced data augmentation techniques tailored to medical images. Some potential areas of research include modality-specific augmentation, spatial transformation augmentation, GANs for augmentation, domain adaptation augmentation, and uncertainty-aware augmentation. Overall, Clinical diagnostic requirements are not limited to the fusion of structural and functional image data, as in thyroid tumor diagnosis, which demands CT, MRI, SPECT, and B-ultrasound. Fusion between multiple modes and algorithm compatibility is a challenging task and another important area for future work.

CONCLUSION

In conclusion, medical image fusion using deep learning algorithms is hot research. This survey paper aims to give researchers, and healthcare professionals a complete understanding of the present state of deep learning algorithms, data augmentation methods, and transfer learning techniques. Medical imaging requires data augmentation to increase the size and diversity of training datasets, improve model accuracy, and overcome overfitting.

The focus of future research should be on addressing challenges related to domain shift, computational complexity, and the realistic nature of synthetic data, while exploring advanced augmentation techniques designed for medical image fusion, leading to more accurate, informative, and clinically valuable fused medical images.

REFERENCES

[1] Azam., M, Khan., K, Ahmad., A et al. (2021). Multimodal Medical Image Registration and Fusion for Quality Enhancement. http://dx.doi.org/10.32604/cmc.2021.016131.

[2] Bharati, S., Mondal, M., Podder, P., Surya Prasath, V. B. (2022). Deep Learning for Medical Image Registration: A Comprehensive Review. *International Journal of Computer Information Systems and Industrial Management Applications*, *14*, pp.173-190. DOI: https://doi.org/10.48550/arXiv.2204.11341.

[3] Balakrishnan., G., Zhao, A., Sabuncu., M., Guttag., J., Dalca., A. (2018). An unsupervised learning model for deformable medical image registration. *Advances in Neural Information Processing Systems,* pp.*9252-9263.* DOI: https://arxiv.org/abs/1802.02604

[4] Zhou., T, Cheng., Q, Lu b., H, Qi Li, Q., Zhang, X., Qiu, X. (2023). Deep learning methods for medical image fusion: A review. Journal of Computers in Biology and Medicine, 160. DOI: https://doi.org/10.1016/j.compbiomed.

[5] Chiu, C., Chiang, H. & Chiu, E. (2023). Developing an explainable hybrid deep learning model in digital transformation: an empirical study. *Intel Manuf.* https://doi.org/10.1007/s10845-023-02127-y

[6] Guo., X, Nie., R, Cao., J, Zhou., D, Mei., L & He., K. (2019). FuseGAN: Learning to fuse multimodal medical images using generative adversarial networks. in IEEE

Transactions on Multimedia, 21(8), pp.1982-1996. DOI: 10.1109/TMM.2019.2895292

[7] Chopra., S, Hadsell., R & LeCun., Y (2005). Learning a similarity metric discriminatively, with application to face verification. *IEEE computer society conference on comput.*

[8] Guo., X, Nie., R, Cao., J, Zhou., D, Mei., L & He., K. (2019). FuseGAN: Learning to fuse multimodal medical images using generative adversarial networks. in IEEE Transactions on Multimedia, 21(8), pp.1982-1996. DOI: 10.1109/TMM.2019.

[9] Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 8798-8808). DOI: https://doi.org/10.48550/arXiv.1909.13719

[10] Chlap., P, Min., H, Vandenberg., N, Dowling., J, Holloway., L, Haworth., A. (2021). A review of medical image data augmentation techniques for deep learning applications, *Journal of Medical Imaging and Radiation Oncology,* 65(5), pp.545-563. DOI: 10.1111/1754-9485.13261

[11] Dyk DAV, Meng XL. (2001). The art of data augmentation. *J Comput Graph Stat*. https://doi.org/10.1198/ 10618600152418584.

[12] Ghaffari. A, Khorsandi., R, Fatemizadeh., E. (2012). Landmark and Intensity Based Image Registration Using Free Form Deformation. *IEEE EMBS International Conference on Biomedical Engineering and Sciences*, (7), pp.68-771. DOI: 10.1109/IECBES.2012.6498156.

[13] Song., T, Yu., X, Yu., S, Ren, Z., Qu., Y. (2021). Feature Extraction Processing Method of Medical Image Fusion Based on Neural Network Algorithm. DOI: https://doi.org/10.1155/2021/7523513.

[14] Goceri E. (2023). Medical image data augmentation: techniques, comparisons, and interpretations. *Artificial Intelligence Review,* 56, pp.12561–12605. https://doi.org/10.1007/s10462-023-10453-z

[15] Huang, J., Le, Z., Ma., Fan, Y., Zhang, F., and Yang, L. 2020. MGMDcGAN: Medical Image Fusion Using Multi-Generator Multi-Discriminator Conditional Generative Adversarial Network, (8), pp. 55145-55157. DOI: 10.1109/ACCESS.2020.2982016.

[16] Haribabu., M, Guruviah., V, Yogarajah., P. (2021). Recent Advancements in Multimodal Medical Image Fusion Techniques for Better Diagnosis: An Overview. 19(7), pp.673-694. DOI: http://dx.doi.org/10.2174/1573405618666220606161137

[17] He, K., Zhang., X, Ren., S et al. (2016). Deep residual learning for image recognition, *computer vision, and pattern recognition, IEEE conference*. Las Vegas, NV, USA.PP: 770-778. DOI: https://doi.org/10.1109/CVPR.2016.90

[18] Mumunia Kaur., T & Kumar., T. (2020). Deep convolutional neural networks with transfer learning for automated brain image classification. *Machine Vision and Applications*, 20 (104). DOI: https://doi.org/10.1007/s00138-020-01069-2

[19] Kim., H, Cosa-Linan., A, Santhanam., N, et al. (2022). Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*,22(69).DOI: https://doi.org/10.1186/s12880-022-00793-7

[20] Khademi., G & Simon., D (2019). Convolutional Neural Network for Environmentally Aware Locomotion Mode Recognition of Lower-Limb Amputees. *ASME Dynamic Systems and Control.* DOI: https://doi.org/10.1115/DSCC2019-9180.

[21] Kalamkar., S & Mary., G. (2022). Multi-Modal Medical Image Fusion Using Transfer Learning Approach. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(12). DOI: http://dx.doi.org/10.14569/IJACSA.2022.0131259.

[22] Khalifa.,N, Mohamed Loey.,M, Mirjalili.,M.(2022). A comprehensive survey of recent trends in deep learning for digital image augmentation. *Artificial Intelligence Review,*55, pp.2351–2377. https://doi.org/10.1007/s10462-021-10066-4.

[23] Li., Y, Zhao., J, Zhihan., L, Jinhua., L. (2021). Medical image fusion method by deep learning, *Cognitive Computing in Engineering, (*2), pp.21–29. DOI: https://doi.org/10.1016/j.ijcce.2020.12.004.

[24] Liang., X, Hu., P, Zhang., L et al (2019). MCFNet: multi-layer concatenation fusion network for medical image fusion. *IEEE Sensor.* J.19, pp.7107–7119. https://doi.org/10.1109/JSEN.2019.2913281.

[25] Medical Decathlon Dataset. (n.d.). Retrieved from https://medicaldecathlon.com/

[26] Ma., J, Pluim., P, & Reinhardt., M. (2019). Medical image registration based on feature extraction and matching: A survey. *IEEE Transactions on Medical Imaging*, 38(2), pp.404-420.

[27] Ma., J, Yu., W, Liang., P, et al. (2018). FusionGAN: A generative adversarial network for infrared and visible image fusion. Information Fusion, 48, pp.11-26. DOI: https://doi.org/10.1016/j.inffus.2018.09.004.

[28] Pajares, G. & Cruz, J. M. 2004. A wavelet-based image fusion tutorial. Pattern recognition, 37(9), pp.1855-1872. DOI: https://doi.org/10.1016/j.patcog.2004.03.010

[29] Roberts. W, Adrat., J, Ahmed., F. (2008). Assessment of image fusion procedures using entropy image quality and multispectral classification. *J. Appl. Remote Sens.* 2(1), pp.1–28. DOI: http://dx.doi.org/10.1117/1.2945910

[30] Böhle, M., Eitel, F., Weygandt, M., and Ritter, K. (2019). Layer-Wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-Based Alzheimer's Disease Classification. National Institutes of Health. DOI: https://doi.org/10.3389%2Ffnagi.2019.00194

[31] Saleh.,M, Ali.,A, Ahmed., K et al . (2023). A Brief Analysis of Multimodal Medical Image Fusion Techniques. *Image Fusion Techniques. Electronics*, 12(79).DOI:https://doi.org/10.3390/electronics12010097.

[32] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1) , pp.1-48. 2019. DOI: https://doi.org/10.1186/s40537-019-0197-0.

[33] Selvaraju, R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *In Proceedings of the IEEE International Conference on Computer Vision,* pp.618-626. DOI: 10.1109/ICCV.2017.74

[34] Vinuesaa, b., & Sirmacekc, B. (2021). Interpretable deep-learning models to help achieve the Sustainable Development Goals. DOI: https://doi.org/10.48550/arXiv.2108.10744

[35] Wang, Z., Xiongfei, Duan, H., Yanchi, C., Zhang, X., Guan, X. (2021). Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform. *Expert Systems with Applications. 171*. DOI: https://doi.org/10.1016/j.eswa.2021.114574

[36] Wei, M., Xi, M., Li, Y., Liang, M., and Wang, G. (2023). Multimodal Medical Image Fusion: The Perspective of Deep Learning. *Academic Journal of Science and Technology*, *5*(3), pp.202-208. DOI: https://doi.org/10.54097/ajst.v5i3.8013.

[37] Zuo., Q, Zhang., J, Yang., Y. (2021). DMC-Fusion: Deep Multi-Cascade Medical Multi-Modal Image Fusin. *IEEE Journal of Biomedical and Health Informatics*, 25(20), pp.3438-3449. DOI: https://doi.org/10.1109/JBHI.2021.3083752.

[38] Zitova, B., & Flusser., J. (2003). Image registration methods: a survey. Image and vision computing, 21(11), pp.977-1000. DOI: https://doi.org/10.1016/S0262-8856(03)00137-9.

[39] Zhang, Y., Li, O., Zhao, Y. 2021. Parallel Deep Learning Algorithms with Hybrid Attention Mechanism for Image Segmentation of Lung Tumors. in *IEEE Transactions on Industrial Informatics*,17(4), pp.2880-2889. DOI: 10.1109/TII.2020.3022912.

[40] Zhang, Y., Li, S., & Zhang, L. (2018). Medical image fusion using deep convolutional neural network. *Information Fusion*, 42, pp. 58-168. DOI: 10.1016/j.inffus.2017.10.006