Multi-Modal Medical Image Fusion Using Transfer Learning Approach

Ms. Shrida Kalamkar, Dr. Geetha Mary A*
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, Tamil Nadu, India

Abstract—**Multimodal imaging techniques of the same organ help in getting anatomical as well as functional details of a particular body part. Multimodal imaging of the same organs can help doctors diagnose a disease cost-effectively. In this paper, a hybrid approach using transfer learning and discrete wavelet transform is used to fuse multimodal medical images. As the access to medical data is limited, transfer learning is used for feature extractor and save training time. The features are fused with a pre-trained VGG19 model. Discrete Wavelet Transform is used to decompose the multimodal images in different sub-bands. In the last phase, Inverse Wavelet Transform is used to obtain a fused image from the four bands generated. The proposed model is executed on Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) datasets. The experimental results show that the proposed approach performs better than other approaches and the significance of the obtained fused image is measured using qualitative metrics.**

Keywords*—*Image fusion; discrete wavelet transform; computer vision; inverse wavelet transform

I. Introduction

The process of fusing multiple images of different imaging modalities to obtain an image with a large amount of information for increasing the clinical applicability of medical images is multimodal medical image fusion [1]. Medical Imaging techniques like Magnetic Resonance Imaging (MRI) give the human body structural characteristics. In contrast, Computed Tomography (CT) provides cross-sectional images from within the body, and Positron Emission Tomography (PET) is provided soft tissue. Single-photon Emission Computed Tomography (SPECT) provides 3D images of different human body organs[1], [2]. Different imaging techniques provide different characteristics and information about the same part of the human body. The purpose of the fusion is to obtain better contrast and fusion quality. The result of the fusion should meet the following conditions [3]:

1) The newly obtained fused image should preserve the original information of source images without any information loss [3].

2) The newly obtained fused image should not introduce any inconsistencies or artefacts, and

3) Should not add misregistration and noise in the newly fused image.

Medical Image Fusion finds its applicability in various diagnostic problems. CT and MRI images of the skull can be fused, providing doctors with both structural and cross-sectional information about the skull. This can help in identifying more clearly any skull-based tumors or diseases[4], [5]. Similarly, MRI and ultrasound images can be fused to confirm vascular blood flow. Many reputed medical institutes use very expensive image fusion tools to fuse multiple medical image modalities to diagnose images. Therefore, there is a need to develop a low-cost model for fusing such modalities of images that can benefit small clinics and hospitals [6]–[8]. Recently Deep learning architectures have become more successful for image processing due to the availability of many public datasets and optimization techniques. Deep neural networks have their role in various medical-related applications. Deep networks can be trained for specific clinical applications with the intended dataset, fast computing system, and a huge volume of data. Pre-trained networks can also be used for medical image processing towards specific clinical applications[9], [10]. This paper focuses on the survey and development of a novel transfer learning and wavelet-based technique for multimodal medical image fusion.

The contributions of the paper are stated as follows:

1) This paper reviews the different multimodal image fusion techniques in the spatial domain, transfer domain and deep learning domain.

2) Discrete wavelet transform (DWT) is used initially, which samples the approximation coefficients and detail coefficients at each level.

3) To handle the issue of limited multimodal medical images, a transfer learning approach based on the VGG19 Framework for any modality of a medical image without a training set is proposed.

The contents of the paper are organized as follows. Section II elaborates on the image fusion methods and the need for transfer learning in multimodal medical image fusion. Section III gives information about related research in the field. The problem of image fusion for multimodal medical images using the transform domain approach and VGG19 transfer learning framework and the detailed methodology has been described in Section IV. Section V illustrates experimentation details; datasets used, Fusion metrics and performance analysis used for evaluating the proposed methodology, qualitative metrics, and related comparative analysis in all three domains. Section VI portrays the concluding remarks and pointers to extend the current research work.

*Corresponding Author.

II. Background

The rapid growth of medical imaging technologies such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and single photon emission computed tomography (SPECT) has provided us much richer information on the physical condition[8], [11].

Table I briefly overviews different medical image modalities used to treat a disease.

TABLE I. Different Medical Imaging Modalities

Sr.No	Medical Image Modality	Methodology of taking an image	Type of image obtained
1	CT Scan[7], [8], [12]	Uses X-ray equipment to take images from within the human body from different angles.	Cross-sectional images of all types of body tissue and organs
2	MRI Scan[7], [8], [12]	Detailed images from within the human body using magnetic and radio waves.	Structural Characteristics of the human body are obtained. Gets Anatomic and contrast details between normal and abnormal tissues.
3	SPECT Scan[7], [8], [12]	Uses nuclear medicine for therapeutic and diagnostic procedures.	The body's temperature, blood flow details, etc., are obtained.
4	PET Scan[7], [8], [12]	It uses a radioactive drug to give information on how tissues and organs function.	Sometimes detect disease before it is detected using other imaging tests. It Delivers details about the area of disease.

In literature, image fusion can be carried out in three ways:

A. Pixel Level Image Fusion

In this mode of fusion, the original information in the source images is directly combined without extracting the relevant features [13]. Different Algorithms like the Color space model, Statistical Algorithms, multiresolution decomposition algorithms, and radiometric algorithms can be used to perform pixel-level fusion. Fig. 1 shows an overview of pixel-level image fusion.

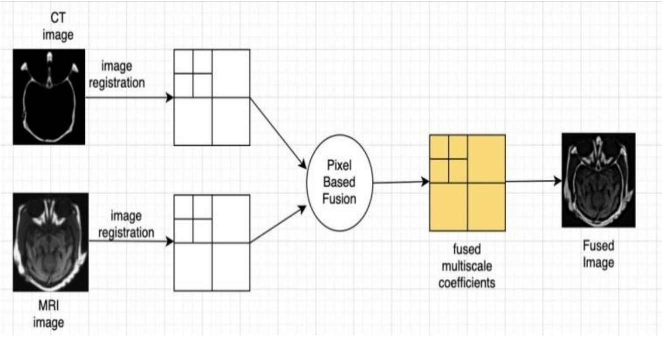

Fig. 1. Pixel-level image fusion.

B. Feature Level Fusion

Feature level fusion, features from the input images using some feature extraction algorithms are extracted to perform the fusion. A region-based fusion approach is used in this type of fusion[14]. Recently, transfer learning algorithms have proven to work well for extracting features from images. Fig. 2 shows an overview of feature-level image fusion.

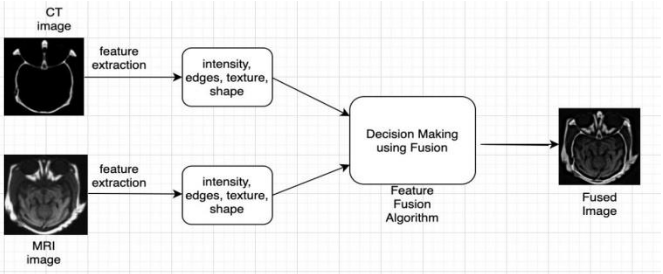
CT image
feature extraction
intensity, edges, texture, shape
feature extraction
intensity, edges, texture, shape
Decision Making using Fusion
Feature Fusion Algorithm
Fused Image
MRI image

Fig. 2. Feature-level image fusion.

C. Decision-level Fusion

Compared to pixel-level and feature-level fusion, decision-level fusion is accurate and supports real-time data fusion. The main disadvantage of this type of fusion is the high loss of information[10]. Algorithms include voting, Bayesian inference, evidence theory, and fuzzy integral [15].

III. Related Work

There are many different methods used in the medical field to fuse different modalities in medical images. A detailed review of multimodal medical image fusion is extensively covered in [4] Medical Image Fusion methods are divided into spatial domain, transform domain, and deep learning domain shown in fig 3.

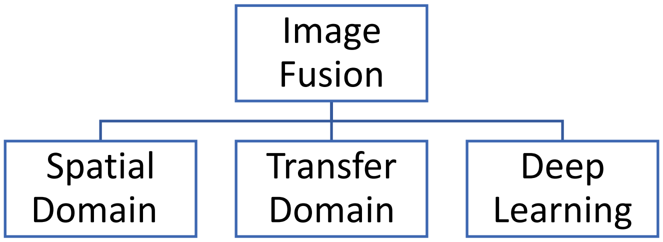

Fig. 3. Image fusion domains.

Table II shows a list of different algorithms in all three domains and systematically reviews different image fusion techniques in all three domains. The algorithms in the spatial domain were popular among the researchers. But spatial domain methods create spatial and spectral distortion of fused images. Hence, researchers used transform domain methods to perform a fusion of images for better fusion results[16]. The fusion methods in the transform domain transform the source image into the frequency domain and then perform reconstruction operations. The disadvantage of the transform domain is that it generates noise in images during fusion processing. With the advancement in deep learning technology, image fusion methods based on deep learning were introduced recently.

TABLE II. MEDICAL IMAGE FUSION METHODS

Domain	Description	Fusion Method	Advantages	Disadvantages
Spatial Domain[1]	Different fusion rules are applied to pixels in the source image.	1. High Pass Filtering 2. Principal Component Analysis 3. Saturation Method of Hue Intensity 4. Minimum and Maximum selection method 5. Brovey Method 6. Average Method	Simple Method. It can be combined as a part of the frequency domain to generate new research methods.	Spectral and Spatial Distortion of Fused Image
Transform Domain[1]	Based on multiscale transform (MST) theory. It transforms the source image from the time domain to the frequency domain. low and high-frequency coefficients are obtained	1. Pulse Coupled Neural Network (PCNN) 2. Discrete Wavelet Transform (DWT) 3. Nonsubsampled Contourlet Transform (NSCT) 4. Nonsubsampled Shearlet Transform(NSST)	Good structure and avoids distortion	Generate noise in image during fusion
Deep Learning[17]	Aim at learning feature hierarchies. Features from higher levels combined with lower-level features. Spatial and Transform Domain-based image fusion methods fail in automatic feature extraction and generation of fusion rules.	1. Convolution Neural Network (CNN) 2. U-Net 3. Stacked Auto-encoders(SAE) 4. Deep Boltzmann Machine(DBM)	Extract the relevant features from the image automatically	Some deep learning models neglect the spatial and temporal topologies of the multimodal data

IV. PROPOSED METHODOLOGY

A. Problem Formulation

We formulate the task of fusion of features from multiple modalities of images as a multimodal medical image fusion problem at the feature level. Several algorithms extract the features of interest from an image, ignore irrelevant features, and eventually apply rules to fuse the important feature during image fusion. These rules play an important role or are the key concept in the fusion process. The selection of proper fusion rules is crucial, resulting in a better fusion process. It is impossible to develop a generic fusion rule for all fusion applications. A more feasible approach in medical image fusion is to develop a new image fusion method that can merge the features extracted from different images into a single fused image.

B. Proposed Methodology for Multimodal Medical Image Fusion

Our proposed work aims to develop a hybrid approach by integrating the structural details provided by CT images with soft tissue details provided by MRI images to get a unique fused image. We have proposed a multimodal medical image fusion based on a blend of Discrete Wavelet Transform and transfer learning VGG19 architecture, which uses the concept of CNN.

The DWT of an image is calculated by passing it through a series of filters, i.e., low pass filter, and then through a high-pass filter to get the approximate and detailed features.

VGG networks are image style transfer methods based on features extracted from the content, style, and generated image. The VGG19 uses the ImageNet dataset for feature extraction and is a pre-trained network. This architecture of VGG19 can be effectively utilized in VGG19-based multimodal image fusion. Fig. 4 graphically represents the proposed CT-MRI medical image fusion architecture.

The following algorithms describe the methodology used for image fusion.

Algorithm 1: Generation of co-efficient from a medical image using DWT.

```
Input: CT and MRI image
Output: Approximate and detailed co-efficient (LL,LH,HL,HH)

DWT(image, DWT):
    1. Decompose CT image using DWT to generate detail and
       approximate co-efficient.
    2. Decompose MRI images using the DWT to generate detail
       and approximate co-efficient.
    3. Pass this co-efficient to VGG19 architecture for fusion.
    4. Repeat the function on image 2.
```

Algorithm 2: Fusion of generated co-efficient using VGG19

```
Input: Detailed and Approximate co-efficient from CT and MRI
images
Output: Fused Image

FUSION(LL,LH,HL,HH):
    1. Convert all images to YCbCr format
    2. Transfer all images to PyTorch tensors
    3. Perform fuse strategy
    4. Reconstruct fused image given RGB input images
    5. Return fused image
```

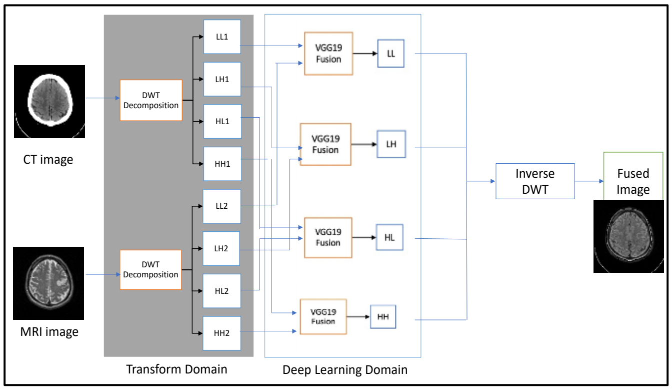

Fig. 4. Proposed CT-MRI Medical Image Fusion Architecture.

In this paper, we have used multimodal medical images, i.e., CT and MRI, the source images given as input to the pre-trained VGG19 model. Initially, we used the Discrete Wavelet decomposition technique of the transform domain to decompose the source images, i.e., CT and MRI images, to generate approximate coefficients of the handcrafted features.

CT images are decomposed to generate LL1 coefficient and horizontal, vertical, and diagonal coefficients, i.e., LH1 (horizontal), HL1 (vertical), and HH1 (diagonal). Similarly, MRI images are decomposed to generate LL2 coefficient and horizontal, vertical, and diagonal coefficients, i.e., LH2 (horizontal), HL2 (vertical), and HH2 (diagonal), as shown in Fig. 5 and Fig. 6.

Later the features from lower-level coefficients, i.e., LL1 and LL2, Horizontal coefficients, i.e., LH1 and LH2, vertical coefficients, i.e., HL1 and HL2, and diagonal coefficients, i.e., HH1 and HH2 of CT and MRI images, respectively are extracted and fused using VGG19 model.

LL, LH, HL, and HH bands are obtained after VGG19 on the input coefficients. The LL band has essential information of an original image. The LH, HL, and HH bands have horizontal, vertical, and diagonal information.

Finally, the frequency sub bands LL, LH, HL, and HH are combined using VGG19 architecture to generate the fused image after applying inverse discrete wavelet transform (IDWT)

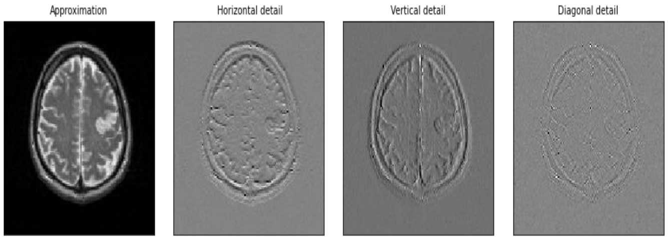

Fig. 5. Wavelet transform of MRI image.

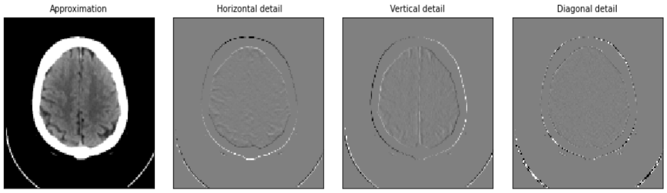

Fig. 6. Wavelet transform of CT Image.

V. Experimental Results on CT and MRI Images

Different fusion algorithms based on spatial, transform, and deep learning domains are used to compare the performance of the VGG19-based medical image fusion.

Spatial Domain methods like Principal Component Analysis (PCA), LPCA, and FCMPCA. Transform domain methods like Standard Wavelet Transform (SWT), Dual-Tree Complex Wavelet Transform (DTCWT), and Non-Subsampled Contourlet Transform (NSCT) are analyzed. Fusion is employed in the wavelet transformation of both CT and MRI images. Wavelet analysis provides a self-adaptive and localized analysis, which is applicable to the time domain and frequency domain. This analysis can focus on any details of the time domain and frequency domain. Multi-scale decomposition of the image based on DWT extracts low-frequency information, as well as horizontal, vertical and diagonal directions of the high-frequency details.

In the VGG19 transfer learning method, fusion is done by applying DWT to CT and MRI images. Both these images are then decomposed into four sub-bands. The sub-bands labelled LH1, HL1 and HH1 represent the detail images co-efficient while the sub-band LL1 represents the approximation image.

A. Evaluation Metrics

1) Peak Signal to Noise Ratio (PSNR): PSNR can be defined as the difference between x and y.

$$\text{PSNR in dB} = 10log_{10}\ \frac{255^2}{\sqrt{\sum_x\sum_y(x-y)^2}} \qquad (1)$$

PSNR is calculated using the two source images and the fused image.

$$PSNR_{avg} = \frac{1}{2}\ (PSNR(IM_1, FI) + PSNR(IM_2, FI)) \quad (2)$$

Where,

- IM_1 -> Image 1,
- IM_2 -> Image 2,
- FI -> Fused Image

A higher value of PSNR indicates better quality of the fused image.

2) Structural Similarity Index (SSIM): It contains the structural information present in the images. The loss of structural information leads to distortion of images.

$$\text{SSIM(x,y)}=(\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1})^{\alpha}(\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2})^{\beta}(\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3})^{\gamma}\ldots..(3)$$

Where,

- μ_x and μ_y are the mean of the images x and y and represents the luminance of the image.
- σ_x and σ_y are the variances. It represents the contrast of the image.
- σ_{xy} is the covariance. It illustrates the correlation between images.
- α, β, and γ - Adjust the relative significance of the above terms. The stability of the metric is C_1, C_2, and C_3- constants that maintain the metric's strength.

Average SSIM is calculated as:

$$SSIM_{avg} = \frac{1}{2}\ (SSIM(IM_1, FI) + SSIM(IM_2, FI)) \quad (4)$$

Where,

- $SSIM_{avg} \rightarrow Average\ SSIM$

- $SSIM(IM_1, FI) \rightarrow$ $SSIM\ between\ source\ image\ 1\ and\ fused\ image$
- $SSIM(IM_2, FI) \rightarrow$ $SSIM\ between\ source\ image\ 2\ and\ fused\ image$

The range of SSIM values is between 0 to 1; 1 indicates a perfect match between the reconstructed image and the original image.

The higher the value of SSIM better is the fused image.

3) Mutual Information (MI): In MI, how much information from two input source images (X, Y) is transferred to the fused image is calculated. The formula is as follows:

$$\text{MI} = \text{I}(x,f) + \text{I}(y,f) \qquad (5)$$

$$\text{I(x,y)} = \sum_{y\in Y}\sum_{x\in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \qquad (6)$$

Where,

- p(x) and p(y) - the edge probability density function of the two images,
- p(x,y) - the joint probability density function of the fused image and the source image X, Y.

Higher the MI value, more information is transferred from the original images to the fused image.

4) Entropy (EN): Information content in the image by taking values between 0 and 8.

$$\text{EN} = - \sum_{L=0}^{L-1} p_i\ X\ log_2 p_i \qquad (7)$$

Where,

L - number of grey levels and is a probability density function for each gray value i.

B. Experiment and Evaluation on the Group of MRI-CT Images

CT and MRI data were used in this experiment of fusion. The evaluation metrics used in this experiment have been discussed in the above section. The dataset used for this experimentation is available in The Whole Brain Atlas repository hosted by www.harvardmed.edu. All the source images are registered, which is the prerequisite for fusion. The source images are of spatial resolution 256×256×3.

We have conducted detailed experimental studies on different CT and MRI test images. The evaluation metrics used in this paper are PSNR, SSIM, MI and EN. The fusion results have been compared with both wavelet-based and neural network-based fusion techniques. Table III below presents the evaluation index data of different methods of MRI-CT fusion.

The performance of the multimodal medical image fusion methods of both wavelet-based and neural network-based fusion are evaluated using four qualitative metrics such as EN, MI, PSNR, and SSIM.

Deep learning techniques perform automatic and near-accurate feature extraction, so the Convolutional Neural Network algorithm achieves the best Entropy (EN), i.e., the amount of information contained in fused images is the largest. Our proposed DWT+VGG19 algorithm performed best on the PSNR and SSIM evaluation indicators. It can measure a large amount of structural information present in the images. High PSNR means a good quality image and a minor error introduced to the image. Also, the higher the value of SSIM better is the fused image. Hence our proposed algorithm for medical image fusion outperforms the other wavelet-based and neural network-based algorithms in the quality of the image.

The MI and Entropy metrics show that wavelet-based algorithms have poor performance if high noise is present in images. In future, we will use VGG19 architecture with averaging and maximum fusion rules and fuse the detailed contents.

TABLE III. EVALUATION INDEX DATA OF DIFFERENT METHODS OF MRI-CT FUSION

Fusion Methods	**MRI-CT evaluation metrics**			
	PSNR	SSIM	MI	EN
GFF	31.16	0.49	3.43	6.80
MSA	35.07	0.48	3.92	6.38
NSCT+SR	29.56	0.48	3.13	6.89
NSCT+PCNN	31.23	0.50	**5.01**	6.74
NSCT+LE	31.61	0.49	3.29	6.80
NSCT+RPCNN	31.68	0.50	4.42	6.77
NSST+PAPCNN	32.92	0.49	3.23	6.84
DWT	31.97	0.43	1.98	6.59
DWT+WA	30.98	0.49	4.75	6.30
U-Net	26.42	0.32	2.22	5.13
CNN	28.96	0.48	3.08	**7.01**
Proposed model	**40.63**	**0.67**	2.88	4.93

VI. CONCLUSION

Deep learning for multimodal medical image fusion has been a recent research topic. Many researchers are focusing on using deep learning in the fusion of different modalities of images. This paper majorly discusses the following points.

1) Image fusion methods in spatial, transform, and deep learning domains are discussed in this paper.

2) Automatic feature extraction using Deep learning models leads to the generation of the most compelling features, eventually improving model accuracy.

3) The VGG19 transfer learning method proposed in this paper can achieve multimodal medical image fusion. This method can effectively achieve image fusion and help in medical diagnosis. It can help in improving the accuracy of medical diagnoses.

4) Experimental results on CT-MRI data show that the proposed transfer learning method achieves state-of-art performance in terms of qualitative evaluation metrics.

REFERENCES

[1] W. Tan, P. Tiwari, H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," Neural Comput Appl, vol. 2, 2020, doi: 10.1007/s00521-020-05173-2.

[2] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. Abd El-Samie, “Survey study of multimodality medical image fusion methods,” Multimed Tools Appl, vol. 80, no. 4, pp. 6369–6396, 2021, doi: 10.1007/s11042-020-08834-5.

[3] V. Rajangam, N. Sangeetha, R. Karthik, and K. Mallikarjuna, “Performance analysis of VGG19 deep learning network based brain image fusion,” Handbook of Research on Deep Learning-Based Image Analysis Under Constrained and Unconstrained Environments, pp. 145–166, 2020, doi: 10.4018/978-1-7998-6690-9.ch008.

[4] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, “A Review of Multimodal Medical Image Fusion Techniques,” Comput Math Methods Med, vol. 2020, 2020, doi: 10.1155/2020/8279342.

[5] B. Meher, S. Agrawal, R. Panda, and A. Abraham, “A survey on region based image fusion methods,” Information Fusion, vol. 48, no. December 2017, pp. 119–132, 2019, doi: 10.1016/j.inffus.2018.07.010.

[6] A. Esteva et al., “Deep learning-enabled medical computer vision,” NPJ Digit Med, vol. 4, no. 1, pp. 1–9, 2021, doi: 10.1038/s41746-020-00376-2.

[7] J. M. Dolly and A. K. Nisa, “A Survey on Different Multimodal Medical Image Fusion Techniques and Methods,” Proceedings of 1st International Conference on Innovations in Information and Communication Technology, ICIICT 2019, pp. 1–5, 2019, doi: 10.1109/ICIICT1.2019.8741445.

[8] A. Dogra, B. Goyal, and S. Agrawal, “Medical image fusion: A brief introduction,” Biomedical and Pharmacology Journal, vol. 11, no. 3, pp. 1209–1214, 2018, doi: 10.13005/bpj/1482.

[9] V. Rajangam, N. Sangeetha, R. Karthik, and K. Mallikarjuna, “Performance analysis of VGG19 deep learning network based brain image fusion,” Handbook of Research on Deep Learning-Based Image Analysis Under Constrained and Unconstrained Environments, pp. 145–166, 2020, doi: 10.4018/978-1-7998-6690-9.ch008.

[10] G. Xiao, D. Prasad, B. Gang, and X. Zhang, Image Fusion.

[11] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. Abd El-Samie, “Survey study of multimodality medical image fusion methods,” Multimed Tools Appl, vol. 80, no. 4, pp. 6369–6396, 2021, doi: 10.1007/s11042-020-08834-5.

[12] G. Sreeja and O. Saraniya, Image Fusion Through Deep Convolutional Neural Network. Elsevier Inc., 2019. doi: 10.1016/b978-0-12-816718-2.00010-5.

[13] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, “Pixel-level image fusion: A survey of the state of the art,” Information Fusion, vol. 33, pp. 100–112, 2017, doi: 10.1016/j.inffus.2016.05.004.

[14] M. Ehatisham-Ul-Haq et al., “Robust Human Activity Recognition Using Multimodal Feature-Level Fusion,” IEEE Access, vol. 7, pp. 60736–60751, 2019, doi: 10.1109/ACCESS.2019.2913393.

[15] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. Abd El-Samie, “Survey study of multimodality medical image fusion methods,” Multimed Tools Appl, vol. 80, no. 4, pp. 6369–6396, 2021, doi: 10.1007/s11042-020-08834-5.

[16] M. Kaur and D. Singh, “Fusion of medical images using deep belief networks,” Cluster Comput, vol. 23, no. 2, pp. 1439–1453, 2020, doi: 10.1007/s10586-019-02999-x.

[17] Y. Li, J. Zhao, Z. Lv, and J. Li, “Medical image fusion method by deep learning,” International Journal of Cognitive Computing in Engineering, vol. 2, no. July 2020, pp. 21–29, 2021, doi: 10.1016/j.ijcce.2020.12.004.