Scientific Reports

https://doi.org/10.1038/s41598-025-26916-y

Article in Press

Multimodality medical image fusion using directional total variation based linear spectral clustering in NSCT domain

Received: 5 May 2025

Accepted: 31 October 2025

Published online: 04 February 2026

Cite this article as: Khan M.Z., Diwakar M., Srivastava P. *et al.* Multimodality medical image fusion using directional total variation based linear spectral clustering in NSCT domain. *Sci Rep* (2025). https://doi.org/10.1038/s41598-025-26916-y

Mohammad Zubair Khan, Manoj Diwakar, Prakash Srivastava, Prabhishek Singh, Neeraj Kumar Pandey, Mohammad Mahyoob Albuhairy & Jeehaan Algaraady

Multimodality medical image fusion using Directional Total Variation Based Linear spectral clustering in NSCT Domain

Mohammad Zubair Khan[1,2], Manoj Diwakar[3,*], Prakash Srivastava[4], Prabhishek Singh[5], Neeraj kumar Pandey[6], Mohammad Mahyoob Albuhairy[7], Jeehaan Algaraady[8]

[1]Department of Computer Science and Information, Taibah University, Madinah 42353, Saudi Arabia
[2]King Salman Center for Disability Research, Riyadh 11614, SaudiArabia
[3,4,6]Department of CSE, Graphic Era (Deemed to be) University Dehradun Uttarakhand, 248001 India
[5]School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India
[7]Languages and Translation Department, Taibah University, SaudiArabia
[8]Languages and Translation College, Taiz University, Taiz, Yemen

[1,2]mkhanb@taibahu.edu.sa, [3]manoj.diwakar@gmail.com, [4]prakash2418@gmail.com, [5]prabhisheksingh88@gmail.com, [6]dr.neerajkpandey@gmail.com, [7]mqassem@taibahu.edu.sa, [8]jihan.amu@gmail.com

*Corresponding author: Manoj Diwakar (e-mail: manoj.diwakar@gmail.com)

Abstract

In medical science, there is a challenge to find out critical information from the medical images by low vision disability medical experts. As a solution, we can enhance the medical images by fusing different modality images viz., CT-MRI which can be more informative. This article presents a new multi-modal medical image fusion architecture in non-subsampled contourlet transform (NSCT) domain which is shift-invariant over noisy medical images. Initially noise from medical images is reduced using a convolution neural network (CNN) approach. Furthermore, NSCT is applied in denoised source multi-modal images to obtain approximation and detailed parts. In approximation parts of both input images, the fusion operation is performed using Direction Total Variation enabled linear spectral clustering. Simlarly in detailed parts of both input images fusion operation is performed using sum modified laplacian (SML) approaches. By performing inverse operation on both modified approximation and detailed parts, final fused image is obtained. From qualitative and quantitative result analysis, it can be concluded that the proposed method is an essential means of ensuring that multi-modality images provide more reliable analytical results to analyze experimental outcomes and comparative research.

Keywords: Multi-modal image fusion, Medical imaging, Non-subsampled contourlet transform, Convolution Neural Network

1 Introduction

In today's scenario, medical image information has a massive impact on the diagnosis on detecting critical diseases like cancer, bleeding, and Alzheimer. To analyze such type of diseases, different modality of images are required and every modality of medical image contains significant information. By fusing different modality images, fused image can be obtained which is more informative and enhanced [1]. The main reason for taking computed tomography (CT) and magnetic resonance imaging (MRI) images is to convert them into a particular image that plays a critical role in medical diagnosis. However, the more advanced modalities are too expensive for the individual to afford. However, we have some practical fusion-based approaches to capture and merge the entire information features from the multiple medical images into a single fused image to diagnose the disease efficiently.

Various medical images for clinical diagnosis of disease, surgery, and therapy through radiation are not sufficient [2]. The sensors of different modalities of images in the medical field are required a single modality medical image which contains maximum critical information. For example, computed tomography (CT) shows rigid structures like bones and hard tissue with minor deformities in the structure. In contrast, MRI (magnetic resonance imaging) shows the structure of the soft tissues [3]. Furthermore, T1-MRI images show human structure detailing of the tissues, while T2-MRI images show the normal and pathological tissues [4]. In addition, single-photon emission computed tomography (SPECT) images reflects clinically remarkable metabolic changes. Thus, a single sensor image usually does not give sufficient information to a doctor in an authentic clinical state. Generally, it is essential to collaborate on sensor images of different modalities to get more extensive information regarding diseased parts or organs. A robust method is used to join the image fusion technique, which amalgamates multi-modal sensor medical images [5]. The fusion-based images provide a more exact representation of a selected object and decrease the oddness and error produced by the sensor in the image. The fusion-based images enhance the efficacy of image-directed diagnoses and the analysis of medical problems [6]. The image fusion approach can be differentiated into three steps viz., pixel-level, feature-level, and decision level [7]. The pixel-level fusion-based approach is mostly accomplished directly on the actual obtained image [9]. The spatial domain-based method is applied to the original image using local spatial functionality [10].

The transform domain-based multi-scale image fusion has recently become the most common fusion approach [11]. Image fusion has been accomplished using the traditional gradient pyramid (GP) transform and discrete wavelet transform (DWT) [12-13]. Due to limitation of shift-invariance, in DWT, some advanced wavelet transform such as dual-tree complex wavelet transform (DTCWT) are more influential for image fusion applications [14]. To conquer these limitations

of the DWT, [15] introduced the dual-tree complex wavelet transform (DTCWT), exhibits both shift-invariance and directional selectivity. Thus, the DTCWT developed through 2-D wavelet can define isotropic movement in terms of line and curve singularities [16-17].

Compared with discrete wavelet transform (DWT), contourlet transform (CT) has various properties such as localization, different directions and anisotropy. Consequently, the contourlet transform (CT) provides edges and other singularities along curves that are more systematic [18-19]. Thus, the up-sampling and down-sampling of contourlet transform (CT) resulting in a loss of shift invariance and pseudo-Gibbs. Nonsubsampled contourlet transforms also maintain shift-invariance and successfully overcome the pseudo-Gibbs existence [20] [21]. However, the NSCT is more acceptable for image fusion. In [31], a new approach based on clustering has been performed for medical image fusion. This approach introduced a new fusion framework using Adaptive Firefly Optimization based Convolutional Neural Network (AFFOCNN) and modified convolutional network. Initially noise has filtered and then segmentation and feature learning has been done for further fusion process. In [32], ambiguous D-means fusion clustering algorithm (ADMFCA) based medical image fusion has been performed. Initially, edge detection has been done to analyze the COVID detection and further fusion operation has been performed. Recent medical image fusion (MIF) research reflects a shift from handcrafted rules toward deep learning and hybrid optimization. Liang [38] showed that deep neural networks can directly learn complementary features from multimodal inputs, while Luo et al. [39] improved edge and scale consistency using a cross-scale transformer with explicit edge enhancement.

Several works leverage transfer learning for efficiency and robustness. Dinh proposed bilateral texture filtering with ResNet-101 [40], optimization with VGG19 [41], and coupled neural P systems for adaptive fusion [42]. These methods exploit pre-trained models to capture modality-invariant features without heavy retraining. Optimization-driven approaches also remain influential. Examples include coati optimization with difference-of-Gaussians [43], local energy–based fusion [44], spectral total variation with neural P systems [45], and advanced decomposition–optimization pipelines [46]. Earlier, metaheuristic search with the grasshopper optimization algorithm [47] provided groundwork for later refinements. Collectively, these studies aim to preserve edges, suppress noise, and improve diagnostic clarity across diverse modalities.

With the motivation of NSCT based medical image fusion, a new multi-modality medical image fusion is proposed using SML and Direction Total Variation in the NSCT domain. The main contributions of this paper are given below:

- A new Proposed approach is presented for medical image fusion where modified Laplacian algorithms and Direction Total Variation based linear spectral clustering are utilized in the NSCT domain followed by CNN.

• In order to demonstrate the effectiveness of the proposed work, both qualitative and quantitative analysis has been carried out.

The rest of the paper is organised as follows: In section 2, a brief NSCT description is discussed. The proposed methodology is derived in Section 3. The results and discussion are included in section 4. Finally, conclusions are drawn in section 5.

2 Non-Subsampled Counterlet Transform(NSCT)

For medical image fusion, we have many transforms such as wavelet transform, shearlet transform. However, for medical image fusion, NSCT (Non-subsampled Contourlet Transform) is most popular transform because of its major characteristics such as directional filtering, Shift Invariance, Multi-scale Decomposition, handling elongated structures and edges with varying orientations. NSCT is a multi-scale and multi-directional computing transform of discrete pixels based on the contourlet transform [7] which contains 2 phases: a non-sampled pyramid (NSP) and an unsampled heading filter. Using a two channel non-sampling channel, a filter bank can generate low and high-frequency images for any NSP decomposition stage. The next steps are taken to deconstruct the availability of a low-frequency component. The decomposition of the NSP is done to decompile the low frequency available to capture the variations of the signal. NSP produces a low-frequency and high-frequency image that includes images of the same size as a source image and reflecting the number of stages in the decomposition.

The non-subsampled contourlet transform [7] can be performed on input images $X(i,j)$ to obtain low and high-frequency sub-bands i.e. $X_L(j)$ and $X_H(j)$, respectively.

$$NSCT(X) = \begin{cases} X(j^{2^{k-1}P}) \prod_{i=0}^{k-2} X_L(X^{2^iP}), 1 \le k \le i \\ \prod_{i=0}^{k-2} X_H(X^{2^iP}), k = i + 1 \end{cases} \quad (1)$$

3 Proposed multi-modal medical image fusion

This section focuses on the proposed fusion framework on the noisy input medical images in NSCT domain. Assuming that both input medical images are noisy due to low transmission. In proposed framework initially a preprocessing operation is performed over the Gaussian noisy input medical images before applying the fusion approach. This denoising approach is performed using CNN.

A: Convolutional Neural Network (CNN)

In proposed work of image fusion, a preprocessing step has been performed to improve the quality of medical images in terms of noise using CNN. By applying noise reduction filters in convolutional layer, the clean features can be obtained. The layers in this model do not differ from other layers but

execute three complex variants of calculations. The first convolutional layer, combined with ReLU activation, processes input data within a D-depth sliding window, where the image acts as the source of nonlinearity across 64 ReLU-activated feature maps [21]–[27]. In the convolutional neural network, three layers are downsampled using a max-pooling operation. For the expanding path, three layers are upsampled using strided convolutions. Initially, a vertex block with three convolutional layers was constructed, followed by a strided-convolution layer to complete the image feature extraction process. Before applying non-linear activations, the data underwent instance normalization to ensure that all variables were scaled to comparable magnitudes. Each layer of the network employed a rectified linear unit (ReLU) activation function, except for the output layer, which did not use any activation or normalization, thereby producing outputs with an unrestricted numerical range [28]–[30].

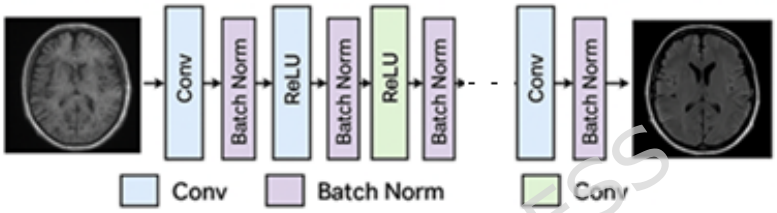


Figure 1: Convolutional Neural Network (CNN) architecture

Figure 1 illustrates a Convolutional Neural Network (CNN) architecture specifically designed for medical image denoising with edge preservation. The input is a noisy medical image (brain MRI), which is passed through a series of processing layers. The architecture includes multiple convolutional layers (Conv) for feature extraction, interleaved with Batch Normalization layers to stabilize and accelerate training, and ReLU activation functions to introduce non-linearity and enhance feature learning. This structured pipeline allows the network to suppress noise while preserving essential structural details, particularly edges. The final output is a denoised and enhanced version of the medical image, showing clearer anatomical features, demonstrating the effectiveness of the CNN in improving image quality for diagnostic purposes.

Training sets were created using the noisy simulated images and the matching "ground truth" (full-count) images included in the supervision. A total of 1200 samples in [33] were used in each network-training phase [33]. The samples were created after the image data were randomly split into 64 patches of identical size and pooled together. It was decided to ignore the axial extremities of each sleeping position's noisy slices throughout the learning process [34]. Before training, we reduce the mean and variance of all patch samples to zero and one, respectively, to prepare them for learning. The normalization factors were retained because they were required to scale the matching noise label by the same quantities as the normalization factors. Despite the minor variance, the final result would be quantitatively correct because of the close closeness of the final result to the less noisy alternative.

To minimize loss, CNN training was stopped when the loss curve of the validation set became flat.

The CNN architecture design is to optimize the problem formulated as Eq (2) follows:

$$\begin{cases} \arg\min_{z} & L(X_{\lambda}, D_{train}, D_{valid}) \\ \text{s.t.} & z \in Z \end{cases} \quad (2)$$

where $Z_z=\{z_1,\ldots,z_n\}$, $Z=Z1\times\cdots\times Zn$, Z_z denotes the CNN architecture parameter setting Z_z , and $L(\cdot)$ measures the performance of Z_z on the validation data valid which has been trained on the training data Dtrain. To optimize CNN model, the following loss function is utilized for minimizing the error using Eq (3):

$$l(\theta) = \frac{1}{2N}\sum_{i=1}^{N} \| R(y_i; \theta) - (y_i - x_i) \|_F^2 \quad (3)$$

N ∈ ℕ: Total number of training samples.; x_i ∈ ℝ^{H×W}: Clean (ground-truth) reference image for the i-th sample.; y_i ∈ ℝ^{H×W}: Observed (noisy or degraded) image corresponding to the i-th sample.; y_i – x_i: True residual (noise or degradation component).; R(y_i; θ) ∈ ℝ^{H×W}: Residual predicted by the CNN model R(·;θ), parameterized by θ.; θ: Set of trainable parameters of the CNN (convolutional kernels and biases).; ||·||_F^2: Squared Frobenius norm, i.e., sum of squared pixel values in a matrix.; L(θ): Average loss across the entire training dataset.

The next two layers inflate the map scale attribute with Conv+BN and ReLU, repeatedly merged after mapping, padded elements. The development and integration of alternated vectors of convolution continue while a 3x3x64-sized filter reconstructs the denoted images in the last layer. Every layer is the product of the previous layer. The residual learning can be implemented and only extracts the remains of the subordinate clean image. The method produces reliable results that are favorable to image quality.

B. Proposed Fusion approach:

The model suggested focuses on the Direction Total Variation based fusion and SML within the NSCT region, which generates a fused F image using a series of source images represented by A and B as shown in figure 2. The second part of the proposed image fusion method, with two source images, A and B, consists of the following steps to perform image fusion without subsampling contours:

Image decomposition using NSCT: Input images are taken from the source images, i.e. CT scan and MRI images of the brain. As a general

rule, images of 512 x 512 are selected for evaluation to get low and high-frequency sub-images at each level using the NSCT transform.

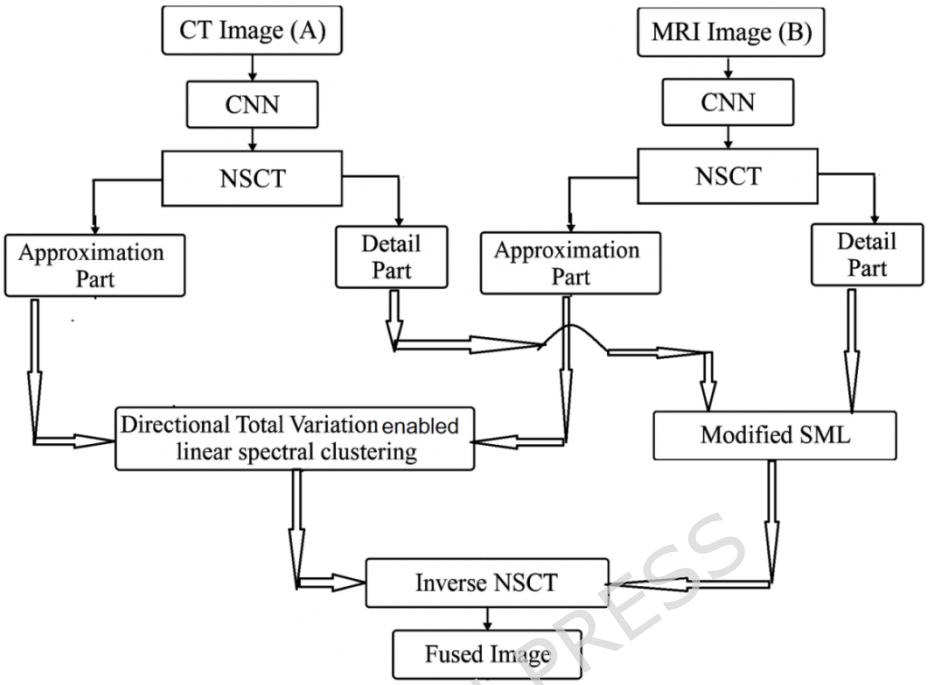


Figure 2: Proposed framework

Directional Total Variation enabled linear spectral clustering on low-frequency NSCT coefficients: In this novel methodology, the mathematically dependent local fusion rule is used. The directional total variation fusion law outperforms fusion laws that depend on a single coefficient, such as the absolute maximum fusion rule or the edge-preserving fusion rule, in terms of efficiency and efficacy. The choice of a certain component in the local fusion law relies on both that factor and its neighboring factor. Boundary selection is advantageous for the directional total variation of coefficients. The comfort level of the center coefficient and its surrounding coefficients may be higher in the vicinity of the margin when the directional total variation is significant. The effectiveness of regional image fusion is contingent upon the correct segmentation of images, as highlighted in the first section, since inaccuracies in division often lead to the formation of certain entities. The evaluation time of image segmentation is also influenced by the feasible congestion of the area fusion approach. The proposed improved linear spectral clustering (LSC) in this part is constrained by these challenges. The LSC approach, introduced in [30], allows for the representation of each pixel in a color image using one of five dimensional vectors (l, a, b, x, y). However, it should be noted that normalized cuts in actual pixel space exhibit a high degree of similarity to weighted K-means clustering in ten-dimensional space. The evidence is considered complete if the weighted K-means in equation (4) are comparable to the

standardized cuts, as shown in a prior study [30]. Each pixel's weight is represented by t(m), and T(m,n) represents the weight that is the same for two pixels m and n. The symbol "γ" represents the map function, which enhances linear separability by converting the pixels into a spatial feature with a large number of dimensions. The properties of the LSC superpixel algorithm may be enhanced [30].

$$\forall m,n \in V, \gamma(m) * \gamma(n) = \frac{T(m,n)}{t(m)t(n)}$$

$$\forall m \in V, t(m) = \sum_{m \in V} t(m,n) \quad (4)$$

At first, we discovered the widely used pixel-by-pixel calculation based on the Euclidean distance. In a color image, the two pixels that follow $= (l_m,a_m,b_m,x_m,y_m,z_m)$ and $n = (l_n,a_n,b_n,x_n,y_n,z_n)$, We may express the same calculation for each m × n pixel as in Equation. (5). Where it is detailed how to calculate proximity to a given location using color and depth data by $\hat{T}_S$, $\hat{T}_c$ and $\hat{T}_d$, respectively. Although, K_S,K_c and K_d are known to supervise the relative characteristic of spatial information, color and depth features, respectively.

$$\hat{T}(m,n) = K_c^2[F(l_m - l_n) + 2.55^2(F(a_m - a_n) + F(b_m - b_n)] + K_S^2 [F(x_m - x_n) + F(y_m - y_n)] + K_d^2[F(z_m - z_n)] \quad (5)$$

$$F(t) = 1 - t^2, \quad t \in [-1,1]$$

In summary, it is clear that we have defined a twelve-dimensional space feature. As an example, the normalized cuts of the input space do not possess a clear correspondence to the weighted K-means clustering of this space. Observing within the precise parameters specified in the equation. The difficulty of assessing the existence approach will be influenced by the inclusion of more condensed elements, such as the matrix kernel for weighted kernel K-means and the matrix affinity in the standardized cuts. Alternatively, using weighted K-means in a twelve-dimensional space with standardized cuts as the goal function may provide expedited and optimal outcomes. N seed pixels, which are equally spaced and sampled vertically, are used to determine the centers of the image, in accordance with the designated number of superpixels. The first sampling of comparable clusters may be found by using the vector properties of the N seed pixels. In the k-means method, we iteratively improve the differentiation of pixel and cluster sampling tasks until convergence is achieved. The enhanced spectral linear clustering are discussed below:

- The mapping of every pixel $m = (l_m,a_m,b_m,x_m,y_m,z_m)$ to a twelve dimensional vector γ(p) are in the space feature.

- A stable horizontal and vertical intervals v_x and v_y are the sampled K seeds over the uniform image.
- Proceed every seed to its less adjacent gradient in the 3×3 adjacent.
- Starting the parameters weighted mean $mean_K$, search centre of every cluster equivalent seed S_K , label of every pixel $m : Label(m) = 0$, Distance of each pixel : $D(m) = \infty$
- Perform again for every weighted means m_K and center search c_K **do**

for each point m in the $v_x \times v_y$ adjacent of c_K in the image plane

do

$D =$ Euclidean distance between $\gamma(p)$ and m_K in the feature space .

if $D < d(m)$ then
$d(m) = D$
$lable(m) = K$
end if
end for
end for

- Improve weighted means and search centers for containing all clusters up to the point when weighted means of converge K cluster.
- Integrate to their adjacent pixel.

Consequently, we associate and choose coefficients from NSCT sources with higher Directional Total Variation content. This lets us pick the image coefficients that are transformed. These coefficients are then used to convert the fused signal in reverse. This fusion law is very unlikely to be a mistake compared with other single coefficient fusion laws. Then, the Directional Total Variation of each coefficient is calculated using the formula is given below Equations.

$$ToJ^{(B)}(J^m) = \sum_{cd} \alpha(c,d)\left(\left|\nabla_y J^m(c,d)\right|\right) + \alpha(c,d)(|\nabla_z J^m(c,d)|) \quad (6)$$

J^m is coefficient at spatial location in iteration m.; α(c,d): adaptive weight at location that balances directional sensitivities (typically set to 1 for uniform weighting).

The equations presented above can be rewritten as follows:

$$v^{m+1} = \arg\min_V \frac{\mu}{2}||V - G||^2 + \frac{\lambda}{2}||f_y^m - \nabla_y^{CbDWF} V - e_y^m||^2 + \frac{\lambda}{2}||f_z^m - \nabla_z^{CbDWF} V - e_z^m||^2 \quad (7)$$

where

$$f_y^{(m+1)} = \arg\min_{f_y} |f_y| + \frac{\lambda}{2}||f_y^m - \nabla_y^{CbDWF} V^{m+1} - e_y^m||^2$$

$$f_z^{(m+1)} = \arg\min_{f_z} |f_z| + \frac{\lambda}{2}||f_z^m - \nabla_z^{CbDWF} V^{m+1} - e_z^m||^2$$

v^{m+1}: fused NSST coefficient block at iteration m+1.
μ: fidelity weight that controls adherence to the observed coefficients.
λ: regularization parameter controlling the effect of directional constraints.
∇_y^{CbDWF}: directional derivative operators along y computed under the **Complex balanced Directional Weighted Filter (CbDWF)** framework.
‖·‖: Euclidean (ℓ2) norm

The Direction Total Variation similarity-based fusion for the approximation of source images is represented by sub-images, where the average merged image cannot be obtained. So we use a Direction Total Variation based fusion rule shown in Eq (11):

$$C(i,j) = \begin{cases} C_1(i,j) & TV_1(i,j) > TV_2(i,j) \\ C_2(i,j) & \text{Otherwise} \end{cases} \quad (8)$$

SML-based fusion on high-frequency coefficients: SML plays a crucial role in medical image fusion. However, fused rules based on higher SML values often lead to image distortion or a lack of detail. By merging both images with the approximation and details after NSCT, a new medical image fused is improved by the SML-based proposing process. The image detail is preserved in the algorithm, and pixels are deleted. After implementing local energy-based fusion, perform SML on the complex components that contain information that mainly exists in the high direction of frequency sub-bands. It can reflect salient features and is used as an activity level measure to select the high-frequency coefficient as shown in Eq (9):

$$SML(i,j) = \sum_{k-M}^{M} \sum_{l-M}^{M} [[ML(i+k, j+l)]^2] \quad (9)$$

Where $ML(I,j) = |2HC(I,j)-HC(i-1,j)-HC(i+1,j)| + |2HC(i,j)-HC(i,j-1)-HC(I,j+1)|$

HC (i, j) refers to the high-frequency pixel coefficient (i, j). The SML doesn't represent precisely the outstanding features if the parameter M is too high. The algorithm provides the best results if M is set to 1. The pixel coefficient with maximum SML value (i,j) is used as the fused coefficient for the high-frequency subbands obtained. The fusion rule is as shown in Eq (10):

$$HC(i,j) = \begin{cases} HC_1(i,j) & \text{if } SML_1(i,j) > SML_2(i,j) \\ HC_2(i,j) & \text{Otherwise} \end{cases} \quad (10)$$

HC(i, j) denotes the coefficient at pixel(i, j).

Image Reconstruction: Finally, the fused images are obtained by performing the inverse NSCT transform, as follows:

$$F = NSCT^{-1}(C,HC) \quad (11)$$

A summary of algorithm is given below:

Algorithm 1: Directional Total Variation Based Linear Spectral Clustering

Input:

I ← Input image
D ← Depth map

Output:

N ← Clustered or fused image

Step 1: Feature Mapping

For each pixel p in I:

$\gamma(p) \leftarrow [l_p, a_p, b_p, x_p, y_p, z_p]$

Step 2: Seed Initialization

Sample K seeds uniformly across the image:

Define horizontal interval v_x and vertical interval v_y

Step 3: Gradient Selection for Seeds

For each seed:

Select the seed with the lowest gradient within 3×3 neighborhood

Step 4: Initialization

For each seed K:

Initialize mean_K and search center c_K

For each pixel m in image:

Label(m) ← 0
D(m) ← ∞

Step 5: Clustering via Feature Distance

For each cluster center c_K:

For each pixel m in $v_x \times v_y$ neighborhood of c_K:

D ← EuclideanDistance($\gamma(p)$, $mean_K$)
If D < D(m):
D(m)← D
Label(m)← K

Step 6: Update Clusters

Repeat until means converge:

Update $mean_K$ and center c_K based on current labels

Step 7: Spatial Smoothing

Integrate clusters by smoothing labels with adjacent pixels

Step 8: Directional Total Variation (DTV) Computation

For each coefficient J^m:

To$J_{B(J^m)} \leftarrow \sum\{c,d\} [\alpha(c,d)\cdot|\nabla_y J\hat{}m(c,d)| + \alpha(c,d)\cdot|\nabla_z J^{m(c,d)}|]$

Step 9: DTV Minimization Optimization

$v\hat{}(m+1) \leftarrow argmin\_V \{ \mu/2\cdot||V - G||^2 +$

$$\frac{\lambda}{2}\cdot\left\|f_y^m - \nabla_y^{CbDWF}V - e_y^m\right\|^2 +$$
$$\lambda/2\cdot||f\_z\hat{}m - \nabla\_z\hat{}CbDWF\ V - e\_z\hat{}m||^2\ \}$$

$$f_y^{m+1} \leftarrow \operatorname{argmin}_{fy}\left\{|f_y| + \frac{\lambda}{2}\cdot\left\|f_y^m - \nabla_y^{CbDWF}V^{m+1} - e_y^m\right\|^2\right\}$$
$$f_z^{m+1} \leftarrow \operatorname{argmin}_{fz}\left\{|f_z| + \frac{\lambda}{2}\cdot\left\|f_z^m - \nabla_z^{CbDWF}V^{m+1} - e_z^m\right\|^2\right\}$$

Step 10: Low-Frequency Fusion Using DTV Rule

For each pixel (i, j):
If $TV_{1(i,j)} > TV_{2(i,j)}$:
$C(i,j) \leftarrow C_{1(i,j)}$
Else:
$C(i,j) \leftarrow C_{2(i,j)}$

Step 11: High-Frequency Fusion Using SML

For each pixel (i, j):
$ML(i,j) \leftarrow |2HC(i,j) - HC(i - 1,j) - HC(i + 1,j)|$
$+ |2HC(i,j) - HC(i,j - 1) - HC(i,j + 1)|$
$SML(i,j) \leftarrow \Sigma_{k=-M\ to\ M}\Sigma_{l=-M\ to\ M}[ML(i + k, j + l)]^2$

If $SML_{1(i,j)} > SML_{2(i,j)}$:
$HC(i,j) \leftarrow HC_{1(i,j)}$
Else:
$HC(i,j) \leftarrow HC_{2(i,j)}$

Step 12: Image Reconstruction

$F \leftarrow NSCT^{-1}(C, HC)$

The Directional Total Variation Based Linear Spectral Clustering (DTV-LSC) algorithm is an advanced image fusion and segmentation method that combines spectral, spatial, and depth information for enhanced image analysis. It begins by mapping each image pixel to a 12-dimensional feature vector that includes color components, spatial coordinates, and depth values. Uniform seeds are sampled across the image to initialize cluster centers, which are then refined by selecting points with the lowest local gradients. Clustering is performed iteratively, where each pixel is assigned to the nearest cluster center based on Euclidean distance in the feature space, and cluster means are updated until convergence.

To improve detail preservation, the algorithm incorporates Directional Total Variation (DTV), which evaluates edge content in horizontal and vertical directions. This DTV information guides the fusion of low-frequency components by selecting those with stronger directional structure. High-frequency details are handled using the Sum Modified Laplacian (SML), which highlights regions with significant local variation, ensuring that important

details are retained during fusion. Finally, the fused image is reconstructed using the inverse Non-Subsampled Contourlet Transform (NSCT), resulting in a high-quality output that maintains structural integrity and detail. This method is particularly useful in applications such as medical imaging and remote sensing, where preserving both global structure and fine details is critical.

3.1 Significance of proposed method

The proposed work for medical image has been analyzed where the importance of CNN based denoising can be clearly mentioned in Table 1. Table 1 show that the results without denoising are not upto the mark. However it can be clearly analyzed that proposed method including CNN based denoising performs well in many parameters.

Table1: Analysis for proposed work with or without CNN

Method	Proposed method without CNN	Proposed method with CNN
MI	4.5432	4.9151
SD	69.4542	72.4345
QAB/F	0.7033	0.7399
SF	27.6532	28.3434
Mean	53.2356	54.2255
Entropy	9.7864	12.5341
FS	1.6432	1.9343
AG	7.3211	9.6532

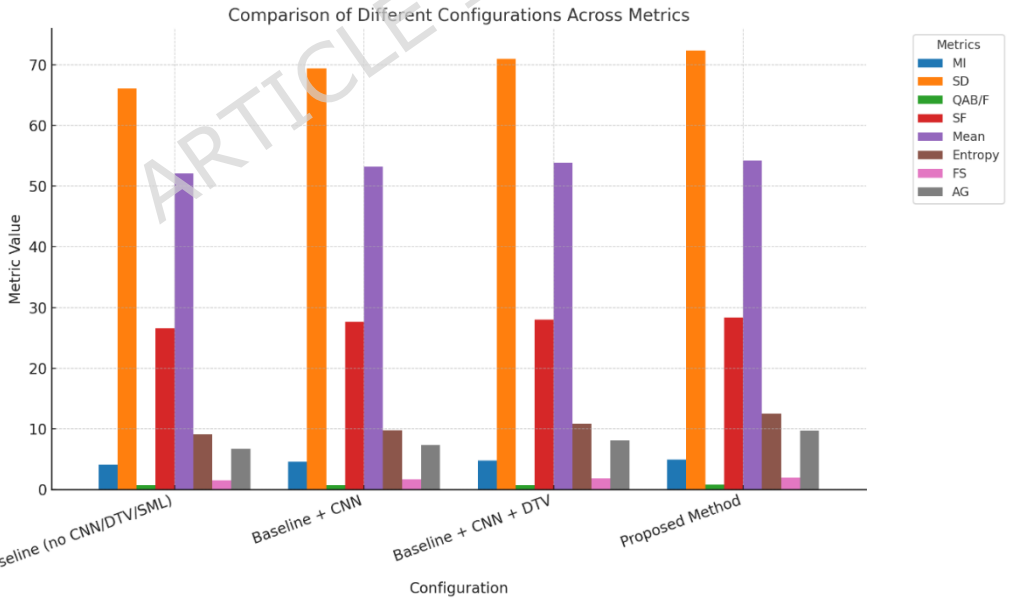


Figure 3: Graphical analysis for ablation study

Table 1 presents a comparative analysis of the proposed method with and without the integration of a Convolutional Neural Network (CNN). The results indicate that incorporating CNN significantly improves performance across all evaluated metrics. Specifically, Mutual Information (MI) increases from 4.5432

to 4.9151, suggesting better fusion of complementary information. The Standard Deviation (SD), which reflects contrast and variation in pixel intensities, improves from 69.4542 to 72.4345, while the QAB/F metric, indicating edge information preservation, rises from 0.7033 to 0.7399. Spatial Frequency (SF), which measures textural detail, shows a slight increase from 27.6532 to 28.3434. Additionally, the average pixel intensity (Mean) increases from 53.2356 to 54.2255, and Entropy, a measure of image information richness, shows a notable gain from 9.7864 to 12.5341. Feature Similarity (FS) also improves from 1.6432 to 1.9343, while Average Gradient (AG), indicative of edge strength and sharpness, increases from 7.3211 to 9.6532. These results confirm that the proposed CNN-enhanced approach offers superior fusion quality, improved structural preservation, and enhanced detail clarity over the non-CNN variant. The figure 3 represents the comparative graph analysis of the proposed CNN-enhanced approach.

The results shown in figure 3 collectively demonstrate the progressive enhancement in medical image fusion performance as CNN, DTV, and SML modules are incorporated into the baseline framework. The baseline configuration shows the lowest values across all metrics, indicating limited ability to preserve complementary details. With the inclusion of CNN, noticeable improvements are observed in mutual information, standard deviation, and entropy, reflecting better information retention and noise suppression. The addition of DTV further enhances structural consistency and edge preservation, while the complete proposed method (CNN + DTV + SML) consistently achieves the highest scores across all metrics. The graphical trends clearly highlight a steady upward progression without regressions, with entropy and average gradient showing the most significant gains, confirming richer information content and sharper anatomical details. Overall, both the table and the graph confirm that the proposed method significantly outperforms the baseline, producing fused medical images with higher diagnostic fidelity, better structural preservation, and enhanced interpretability for clinical applications.

4 Results and Discussion

The results and experimental analysis have been performed in the MATLAB 2021 environment. To analyze the results, visual analyses have been done on the basis of heterogeneous regions, homogeneous regions, sharpness, smoothness, texture, edges and contrast. The visual results are also examined by some medical experts such as Researchers in medical imaging and medical experts (Both type of medical experts who have perfect vision as well as who have low vison). However, human eyes are not so capable to analyze the visuals perfectly. Therefore some standard performance metrics are also utilized to examine the results. All the results have been tested and compared with some recent and state-of-art methods such as [20], [21], [23], [25], [27], [28] and [29]. The results are evaluated and tested on the publically available datasets (Available at http//www.med.harvard.edu/AANLIB/) where medical image pairs

are utilized for the medical image fusion. There are many different modality medical image pairs datasets [35], [36], [37] such as MRI-PECT, CT-MRI and many more. The results are evaluated and tested over the 181 pair of medical images. To construct the training datasets, we collect 181paired images of each group such as from CT-MRI, PET-MRI, and SPECT-MRI of size 256x256 for fusion tasks. To ensure data integrity and prevent leakage, we adopted a patient-level data split, ensuring no overlap of patient images across training, validation, and testing sets. The dataset was divided into 126 for training, 27 for validation, and 27 for testing image pairs. This approach aligns with established practices in recent studies, promoting reproducibility and reliable model generalization. All images were registered, normalized, and resized to a uniform resolution of 256×256 pixels. Each image was then divided into overlapping patches of 64×64 pixels with a stride of 32 pixels. To improve model robustness, random augmentations such as flips and rotations were applied, along with a 10% chance of random occlusion using black or white masks. Finally, patches were upscaled to 128×128 pixels via nearest-neighbor interpolation and normalized to the [0, 1] intensity range.

In this procedure, to take the quantitative standard applied in the objective analysis respectively. It is known that distinguish image quality sharpness of the standard evaluate the visual quality of images from the distinguish facet, but none of them evaluate the quality of image directly. In this research article, we observe both the visual representation and the quantitative assessment of the fused images. For computation of the introduction of the fusion approach, we have contemplated three distinct fusion performance sharpness of the standard are described below.

a) Entropy = $-\sum_{i=0}^{l-1} p_i \log_2(p_i)$
(12)

b) Mutual information (mI) : $mI = mI^{AF} + mI^{BF}$,
(13)
In which

$$mI^{AF} = \sum_{f=0}^{l}\sum_{\alpha=0}^{l} P^{AF}(\alpha,f)\log_2\left(\frac{P^{AF}(\alpha,f)}{P^{A}(\alpha)P^{F}(f)}\right) ,$$

And

$$mI^{BF} = \sum_{f=0}^{l}\sum_{b=0}^{l} P^{BF}(\beta,f)\log_2\left(\frac{P^{BF}(\beta,f)}{P^{B}(\beta)P^{F}(f)}\right) ,$$

c) Edge Based Similarity Measure ($q^{AB/F}$):

$$q^{AB/F} = \frac{\sum_{i=1}^{ml}\sum_{j=1}^{N}[q^{AF}(i,j)wg^{A}(i,j) + q^{BF}(i.j)wg^{B}(i,j)]}{\sum_{i=1}^{ml}\sum_{j=1}^{N}[wg^{A}(i,j) + wg^{B}(i,j)]}$$

(14)

in which

$$q^{AF}(i,j) = q_{\alpha}^{AF}(i,j)q_{r}^{AF}(i,j)$$
$$q^{BF}(i,j) = q_{\alpha}^{BF}(i,j)q_{r}^{BF}(i,j)$$

d) Standard Deviation (std): $\left(\frac{1}{R\times N}\sum_{i=1}^{R}\sum_{j=1}^{N}\left(E(i,j) - \hat{\theta}\right)^2\right)^{1/2}$,
(15)

e) $sf = \sqrt{rf^2 + cf^2}$

(16)

where rf is the row frequency , and cf is the column frequency :

$$rf = \sqrt{\frac{1}{R(N-1)} \sum_{i=0}^{R-1} \sum_{j=0}^{N-2} (f(i,j+1) - f(i,j))^2}$$

$$cf = \sqrt{\frac{1}{R(N-1)} \sum_{i=0}^{R-2} \sum_{j=0}^{N-1} (f(i+1,j) - f(i,j))^2}$$

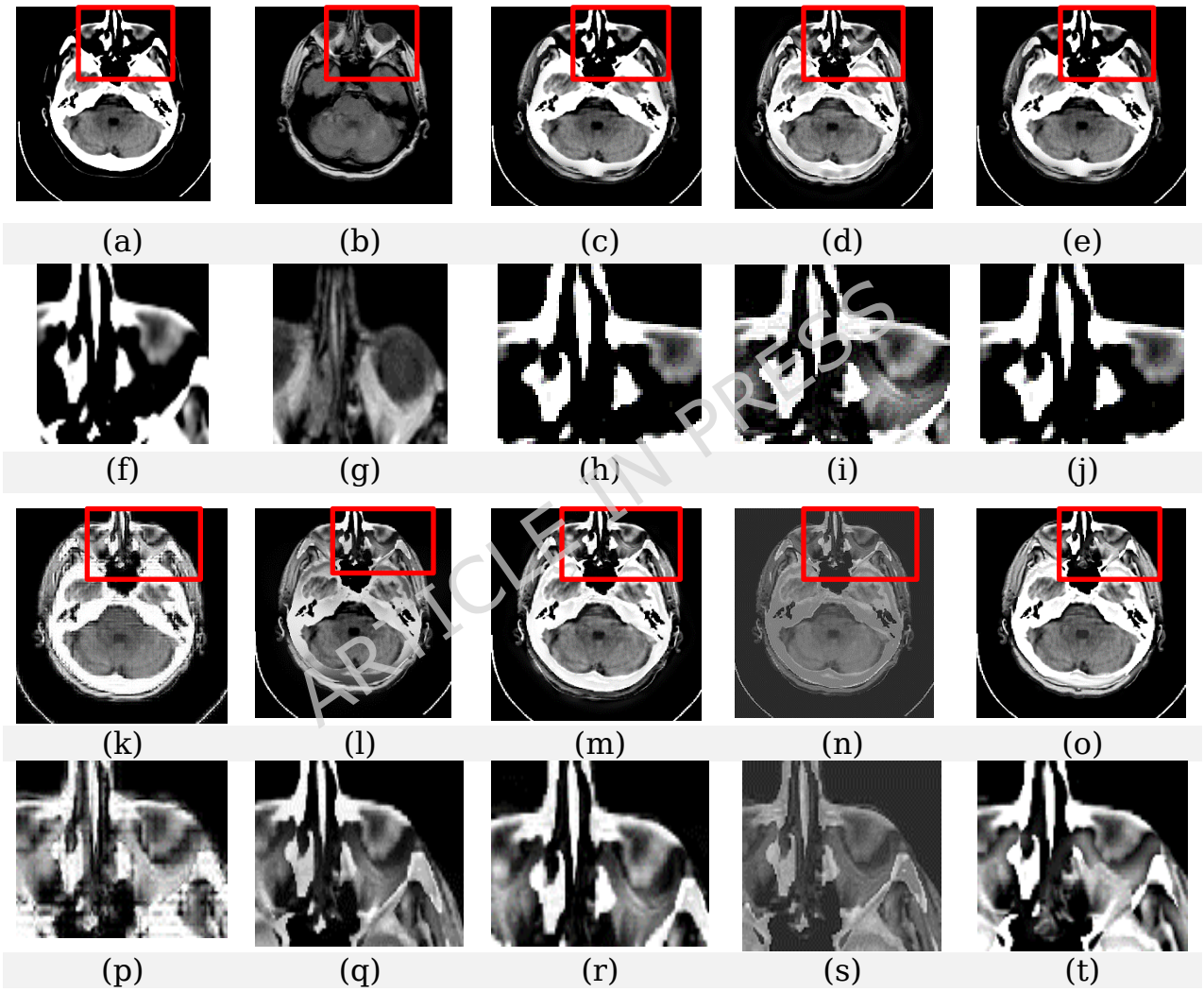


Figure 4: (a) and (b) represent multi-modal medical images. (c) illustrates the outcome using the method of [20], (d) shows the outcome of [21], (e) presents the result of [23], (f) zoomed analysis of first input medical image, (g) zoomed analysis of second input medical image, (h) zoomed analysis of [20], (i) zoomed analysis of [21], (j) zoomed analysis of [23], (k) shows the outcome from [25], (l) shows the outcome from [27], (m) shows the outcome from [28], (n) shows the outcome from [29], (o) shows the outcome of proposed method, (p) zoomed analysis of [25], (q) zoomed analysis of [27], (r) zoomed analysis of [28], (s) zoomed analysis of [29], (t) zoomed analysis of proposed method

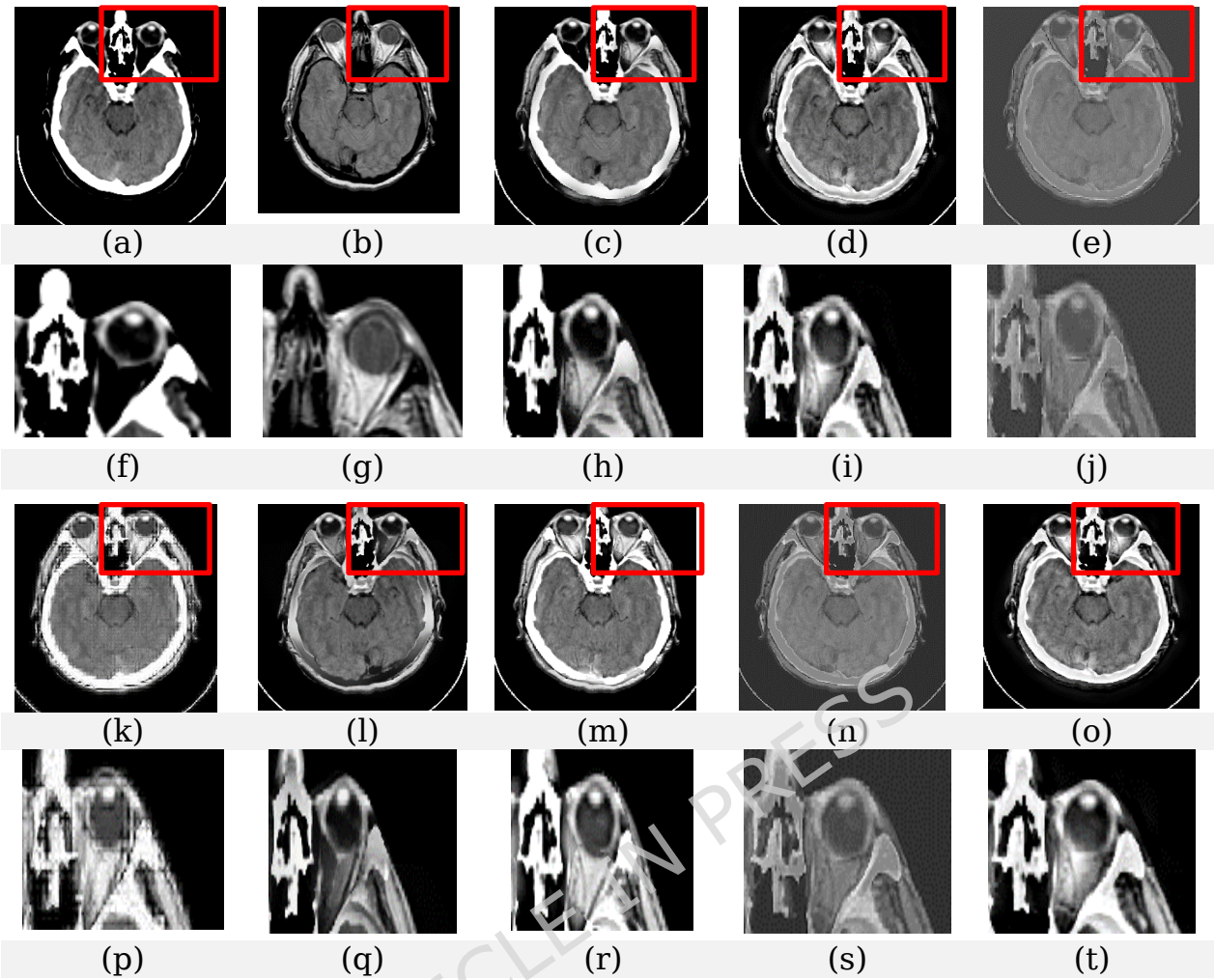


Figure 5: (a) and (b) represent multi-modal medical images. (c) illustrates the outcome using the method of [20], (d) shows the outcome of [21], (e) presents the result of [23], (f) zoomed analysis of first input medical image, (g) zoomed analysis of second input medical image, (h) zoomed analysis of [20], (i) zoomed analysis of [21], (j) zoomed analysis of [23], (k) shows the outcome from [25], (l) shows the outcome from [27], (m) shows the outcome from [28], (n) shows the outcome from [29], (o) shows the outcome of proposed method, (p) zoomed analysis of [25], (q) zoomed analysis of [27], (r) zoomed analysis of [28], (s) zoomed analysis of [29], (t) zoomed analysis of proposed method

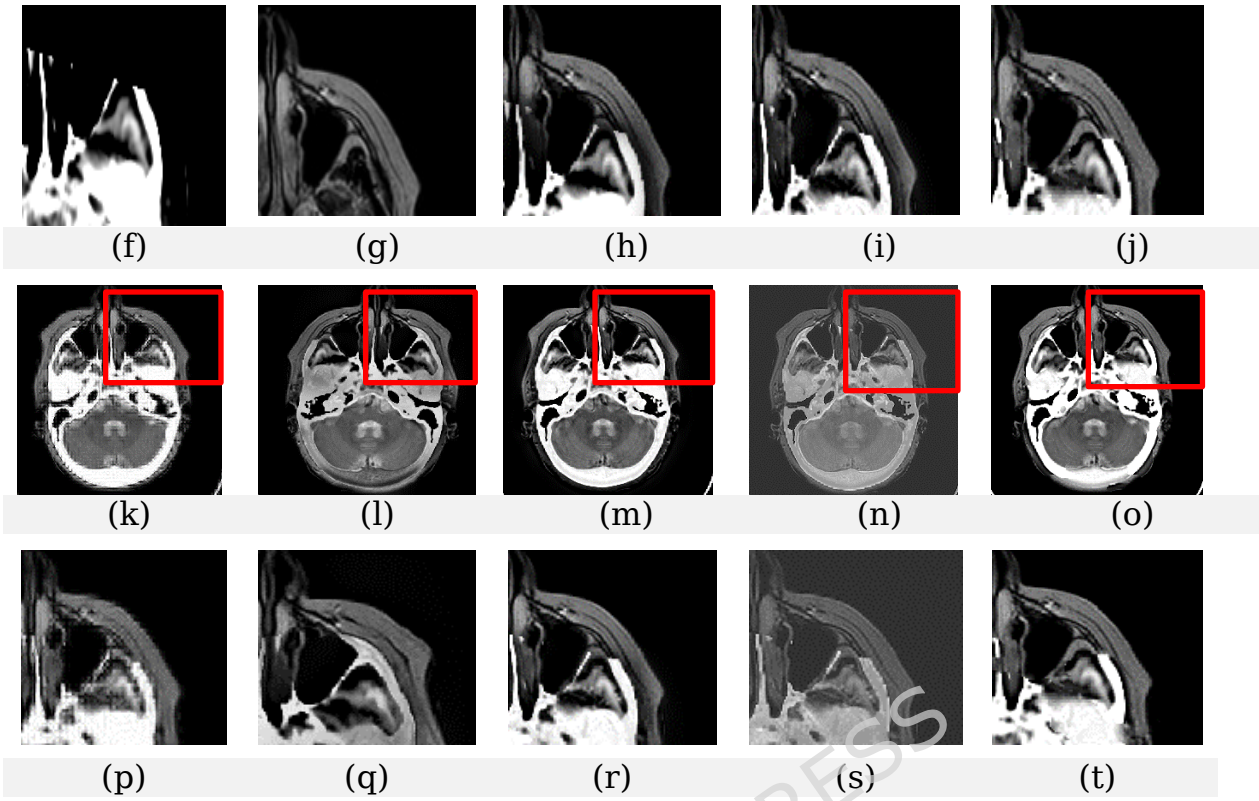


Figure 6: (a) and (b) represent multi-modal medical images. (c) illustrates the outcome using the method of [20], (d) shows the outcome of [21], (e) presents the result of [23], (f) zoomed analysis of first input medical image, (g) zoomed analysis of second input medical image, (h) zoomed analysis of [20], (i) zoomed analysis of [21], (j) zoomed analysis of [23], (k) shows the outcome from [25], (l) shows the outcome from [27], (m) shows the outcome from [28], (n) shows the outcome from [29], (o) shows the outcome of proposed method, (p) zoomed analysis of [25], (q) zoomed analysis of [27], (r) zoomed analysis of [28], (s) zoomed analysis of [29], (t) zoomed analysis of proposed method

Figures 4, 5, and 6 present representative visual comparisons of fused medical images obtained from different modality pairs (e.g., CT–MRI, PET–MRI, SPECT–MRI). In each case, subfigures (a) and (b) show the input multi-modality images, while subfigures (c–t) illustrate the fusion results from several existing methods ([20], [21], [23], [25], [27], [28], [29]) and the proposed method. From a comparative visual analysis, methods [20] and [21] deliver reasonably good structural results but fail to preserve texture in homogeneous regions and show blurring in heterogeneous areas. In contrast, methods [23] and [29] suffer from significant contrast degradation, which limits the visibility of subtle anatomical details, making the outputs less useful for clinical interpretation. Method [25] improves upon this by maintaining some structural integrity, but still falls short in edge definition, particularly in complex anatomical regions. Methods [27] and [28] demonstrate better edge and texture preservation in both homogeneous and heterogeneous areas, with [28] offering a slight advantage in contrast. However, the proposed method consistently achieves superior performance across all cases. It effectively enhances contrast, preserves textures, and maintains sharp edges, resulting in clearer visualization of both soft tissue and high-intensity structures. These improvements are particularly

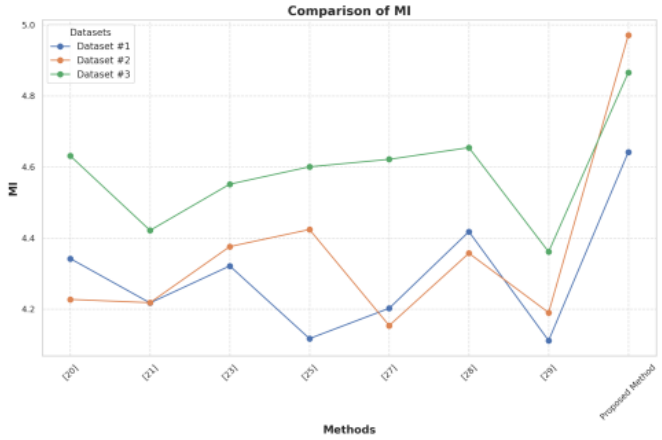
evident in zoomed regions, where fine details and structural transitions are more accurately retained. Given the consistency of these observations across different modality pairs, and to avoid redundancy, we consolidate the analysis and present only two representative figures in the revised manuscript.

Table 2: The outcomes evaluated using performance metrics

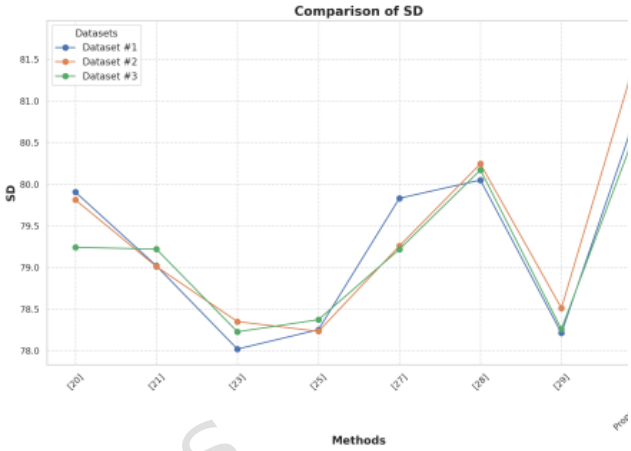
Parameter	Dataset	[20]	[21]	[23]	[25]	[27]	[28]	[29]	Proposed Method
MI	#1	4.3418	4.2171	4.3212	4.1167	4.2019	4.4178	4.1111	4.6418
	#2	4.2270	4.2178	4.3757	4.4238	4.1534	4.3570	4.1898	4.9710
	#3	4.6310	4.4213	4.5515	4.6003	4.6214	4.6543	4.3610	4.8658
SD	#1	79.9053	79.0191	78.0198	78.2526	79.8310	80.0498	78.2122	81.0563
	#2	79.8118	79.0111	78.3498	78.2325	79.2587	80.2448	78.5118	81.7798
	#3	79.2424	79.2195	78.2272	78.3723	79.2187	80.1710	78.2596	80.8467
QAB/F	#1	2.6454	2.6312	2.6151	2.6222	2.6123	2.6152	2.6171	2.6960
	#2	2.5251	2.5140	2.5178	2.5218	2.5187	2.5211	2.5183	2.7288
	#3	2.6311	2.6151	2.6281	2.6271	2.6171	2.6351	2.5919	2.7398
SF	#1	27.8120	27.6511	27.0710	28.3186	27.7142	28.4504	27.1456	29.1822
	#2	27.4123	27.7833	27.0141	27.6113	27.5422	28.7233	27.1113	29.8123
	#3	27.0019	27.1813	27.0111	28.1818	27.0718	28.0019	27.0926	28.7319
Mean	#1	49.3249	50.2346	53.8543	55.1209	56.0238	57.5120	57.5189	58.5350
	#2	44.1433	45.7246	47.4440	50.3356	51.1270	52.8219	53.3409	53.9609
	#3	41.3453	41.2233	42.1753	43.0125	44.1241	45.1240	46.2134	47.7970

From Table 2, it is clearly visible that the proposed method consistently outperforms all other existing methods across all mentioned datasets by analyzing performance metrics (MI, SD, QAB/F, SF, and Mean). Proposed method achieves better outcome values by using performance metrics in most cases. Furthermore, the statistical analysis of the t-test results demonstrates that the Proposed Method consistently delivers superior performance across all evaluated metrics, often with statistically significant improvements over baseline methods. In terms of Mutual Information (MI), the Proposed Method achieved the highest mean value (4.8262) across all datasets, with significant gains over methods [21], [23], [25], and [29] (p-values ranging from 0.0205 to 0.0474), and the largest improvement of +0.6056 observed against [29]. Standard Deviation (SD), an indicator of image contrast and detail preservation, showed the most consistent statistical advantage, with the Proposed Method (mean 81.2276) significantly outperforming six out of seven baselines, including the largest improvements of +3.0287 over [23] (p=0.0059) and +2.8997 over [29] (p=0.00465). For the QAB/F metric, although the Proposed Method (mean 2.7215) produced the highest values across all datasets, none of the pairwise comparisons reached statistical significance, likely due to the small effect sizes and limited sample size (n=3). In Spatial Frequency (SF), which reflects image sharpness and detail, the Proposed Method (mean 29.2421) showed significant improvements in five out of seven comparisons, with the most notable gain of +2.2100 over [23] (p=0.0197). Similarly, for the Mean intensity metric, the Proposed Method (mean 53.4309) significantly outperformed six of seven baselines, recording a substantial

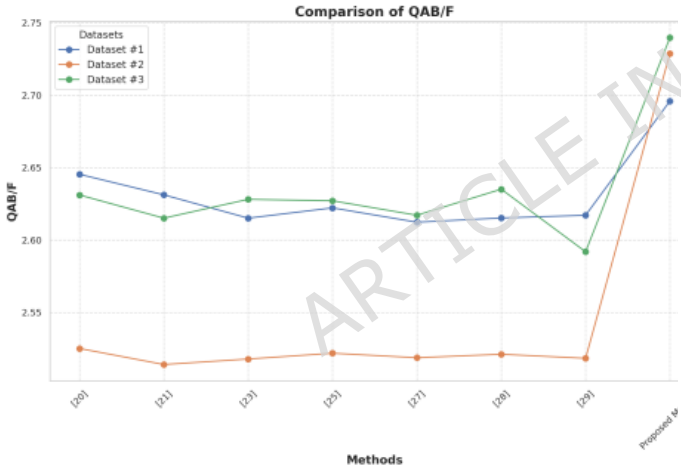
increase of +8.4931 over [20] (p=0.0204). Overall, the Proposed Method not only achieved the highest numerical performance for all metrics across all datasets but also demonstrated statistically significant superiority in SD, Mean, and SF metrics against most baselines, confirming that the observed improvements are both consistent and meaningful in terms of image fusion quality.



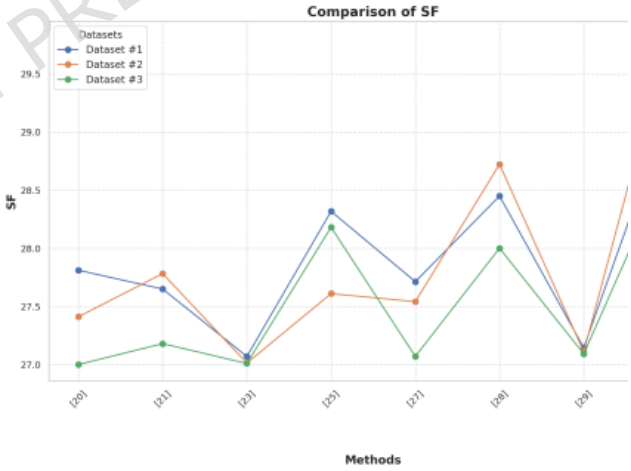
(a)



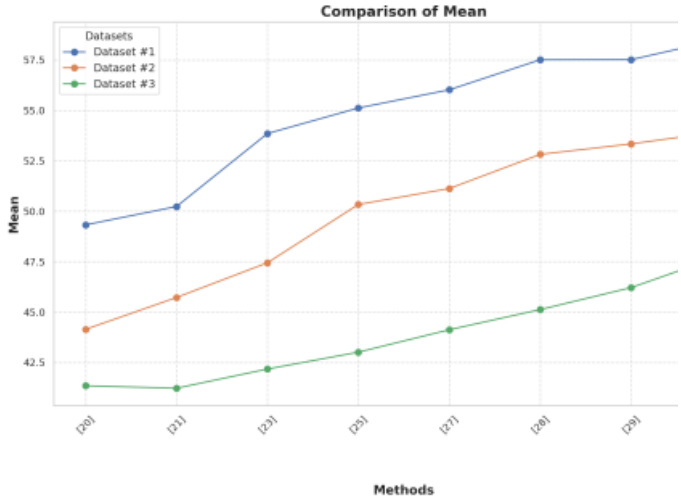
(b)



(c)



(d)



(e)

Figure 7: Graphical analysis of the outcomes evaluated using performance metrics; (a): Comparison of MI; (b): Comparison of SD; (c): Comparison of QAB/F ; (d): Comparison of SF; (e): Comparison of Mean

In figure 7, the outcomes of all the performance metrics which are evaluated here with different input images, are shown in graphical representation for better analysis. Here in figure 7(a-e), the peak node can be easily identified. In all these graphs, the highest peak node is taken by the proposed method which indicates that in all these performance metrics the result of proposed methods are better in compare to all these existing methods and stat-of-arts.

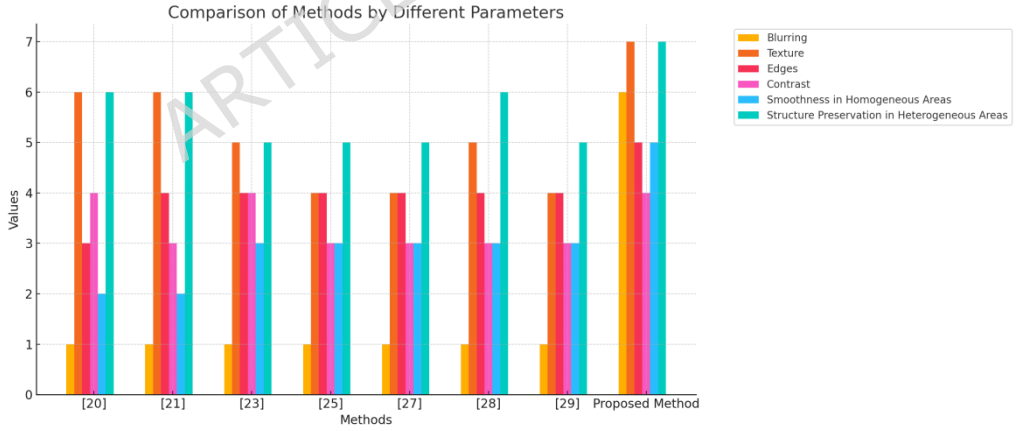


Figure 8: Graphical analysis of the outcomes evaluated by visual analysis via experts

Additionally, the results are visually evaluated by the low visibility experts in the fields of medical science, medical doctors, and researches/scientists in medical imaging fields. Total of 115 medical experts reviewed the visual outcomes of the proposed methods and existing methods. The results are evaluated on the basis of visual parameters such as blurring, texture, contrast, edges, smoothness in homogeneous and sharpness in heterogeneous regions. The results are evaluated by scaling the value from 1-7 where 1 is the poorest

result and 7 is the excellent result. The overall average result is estimated of all scale values of respective methods which is shown as graphical representation in figure 8. From figure 8, it can be clearly analyzed that the proposed outcomes gives better outcomes.

The computational complexity of multimodal medical image fusion algorithms determines their scalability, execution speed, and feasibility for real-time or large-scale clinical deployment. In this subsection, we derive and compare the theoretical time complexity of the proposed method with that of several representative baseline approaches ([20], [21], [23], [25], [27], [28], [29]) in a unified notation. Let: H, W — height and width of the image (in pixels). N = H × W — total number of pixels per image. L — number of multiscale decomposition levels. r — spatial radius (or half-kernel size) of local filters. Kernel size is k × k, where k ≈ 2r + 1. p — patch size (in pixels per side) for patch-based operations. f — number of filters (feature channels) in a convolutional neural network (CNN). d — depth (number of layers) in a neural network. C_conv — cost per convolutional layer of size p × p with f filters: $O(N \cdot p^2 \cdot f)$. FFT-based transforms are assumed to have cost O(N log N). Small fixed-size spatial kernels (r constant) are treated as O(N).

In [20], NSCT + Local Std. Dev. + PCNN: NSCT decomposition & reconstruction: 3 × O(L·N log N). Local area standard deviation (LF fusion): $O(N \cdot r^2)$. Pulse Coupled Neural Network (HF fusion): $O(N \cdot p^2)$ per iteration. Overall: O(L·N log N) (NSCT dominates for large N). In [21], O-DTCWT + NSST + Fuzzy Logic + ODNN. O-DTCWT decomposition/reconstruction: 2 × $O(L_1 \cdot N \log N)$. Fuzzy logic rule (HF fusion): O(N). Maximum rule (LF fusion): O(N). NSST decomposition/reconstruction: 2 × $O(L_2 \cdot N \log N)$. Optimized Deep Neural Network (ODNN) for HF fusion: $O(N \cdot p^2 \cdot f \cdot d)$. Overall: O(L·N log N) + $O(N \cdot p^2 \cdot f \cdot d)$. In [23] SMRFnet (VGG + Multi-scale Residual Net). VGG feature extraction: $O(N \cdot p^2 \cdot f\_VGG \cdot d\_VGG)$. Feature addition & fusion: O(N). Multi-scale residual network decoding: $O(N \cdot p^2 \cdot f\_res \cdot d\_res)$. Overall: $O(N \cdot p^2 (f\_VGG \cdot d\_VGG + f\_res \cdot d\_res))$. In [25], PSA + Multiscale Structure Patch Decomposition. Penalty function-based filtering: $O(N \cdot r^2)$. Pixel-level structure-aware filtering: $O(N \cdot r^2)$. Multiscale patch decomposition: $O(L \cdot N \cdot p^2)$. Weighted fusion: O(N). Overall: $O(N \cdot r^2) + O(L \cdot N \cdot p^2)$. In [27], MSD + Visual Saliency + Weight Maps. Image enhancement: O(N). MSD decomposition/reconstruction: $O(L \cdot N \cdot r^2)$ or O(L·N log N) with FFT. Visual saliency & weight maps: $O(N \cdot r^2)$. Overall: $O(L \cdot N \cdot r^2)$. In [28], STV + Max-cloud Fusion + Multichannel Neural P System. Spectral Total Variation decomposition: O(t_STV·N). Max-cloud fusion: O(N). Multichannel coupled neural P system: $O(N \cdot p^2 \cdot f \cdot d)$. Overall: O(t_STV·N) + $O(N \cdot p^2 \cdot f \cdot d)$. In [29], NSCT + sCNN + FOTGV. NSCT decomposition/reconstruction: 3 × O(L·N log N). Siamese CNN: $O(N \cdot p^2 \cdot f \cdot d)$. Fractional Order TGV denoising: O(t_TGV·N). Overall: O(L·N log N) + $O(N \cdot p^2 \cdot f \cdot d)$. In proposed CNN + NSCT + DTV + SML. CNN preprocessing: $O(N \cdot p^2 \cdot f \cdot d)$. NSCT decomposition/reconstruction: 3 × O(L·N log N). Directional Total Variation: O(t_DTV·N). Modified SML: O(N). Overall: O(L·N log N) + $O(N \cdot p^2 \cdot f \cdot d)$ + O(t_DTV·N).

5 CONCLUSION

This paper presents a novel approach to image fusion using an unsampled contourlet transform for multi-modal medical images. Two distinct criteria are used to preserve supplementary information in the merged image, resulting in improved precision for fusion. Direction Total Variation dependent fusion is used to combine the low-frequency bands, whereas the fusion process for high-frequency bands is accomplished via SML. Visual examination by low vision medical experts, scientists and medical doctors and measurements of output metrics demonstrate that the suggested algorithm has the ability to enhance picture details and enhance the quality of the visible outcomes. Based on the examination of qualitative and quantitative findings, it can be inferred that the suggested technique is crucial for guaranteeing that multi-modality pictures provide more dependable analytical outcomes in terms of visual results and performance metrics. The suggested approach may be used in the medical sector, namely for the purpose of preparing medical equipment to produce medical pictures. One way to use a multi-modal picture fusion approach is by including an additional feature. It is possible to get and evaluate composite pictures or data from several medical systems pertaining to certain organs. Fusion has the capability to extract a greater amount of information and fused images with supplementary data. Multi-modal image fusion enables the creation of image that can be easily identified using both physiological and anatomical data, also helpful to low vision medical experts, scientists and medical doctors for better analysis. This proposed work is only applicable for multimodality medical images such as CT-MRI. As a limitation of this work, it is not effectively work for ultrasound images. However this work also can be extended in future for healthcare industry by analyzing the fused image for any kind of disease predication.

Declarations

Ethics approval and consent to participate: Not Applicable

Clinical Trial: Not Applicable

Consent for publication: Not Applicable

Availability of data and material: The dataset analyzed during the current study is available from the corresponding author (i.e. Manoj Diwakar) on reasonable request.

Competing interests: The authors declare that they have no conflict of interest.

Funding: The authors extend their appreciation to the King Salman Center For Disability Research for funding this work through Research Group no. KSRG-2024-141.

Acknowledgements: The authors extend their appreciation to the King Salman Center For Disability Research for funding this work through Research Group no. KSRG-2024-141.

Author Contributions Statement: M.J.K. and M.D. : writing original draft; P.Sr., P.Si., N.K.P., M.M.A. and J.A. : writing, review, and editing.

REFERENCES

1. Singh R, Khare A (2014) Redundant discrete wavelet transform based medical image fusion. In: Advances in signal processing and intelligent recognition systems. Springer, Cham, pp 505– 515. https://doi.org/10.1007/978-3-319-04960-1_44
2. Bhatnagar G, Wu QJ, Liu Z (2015) A new contrast based multi-modal medical image fusion framework. Neurocomputing 157: 143–152. https://doi.org/10.1016/j.neucom.2015.01.025
3. Bindu CH, Prasad KS (2018) Automatic region segmentation and variance based multi-modal medical image fusion. In: Cognitive science and health bioinformatics. Springer, Singapore, pp 57–63.
4. Pohl C, Nazirun NN, Tamin SS Multimodal medical image fusion in cardiovascular applications. In Medical Imaging Technology, Springer, Singapore. 2015;91–109. https://doi.org/10.1007/978-981-287-540-2_4
5. Wu D, Yang A, Zhu L, Zhang C (2014) Survey of multi-sensor image fusion. In: In International Conference on Life System Modeling and Simulation and International Conference on Intelligent Computing for Sustainable Energy and Environment. Springer, Berlin, pp 358–367. https://doi.org/10.1007/978-3-662-45283-7_37
6. Kaur N, Bahl M, Kaur H (2014) Introduce review on: Image Fusion Using Wavelet and Curvelet Transform (IJCSIT). Int J Comp Sci Inform Technol 5(2):2467–2470
7. A.L. daCunha, J. Zhou, M.N. Do, The nonsubsampled contourlet transform: theory, design and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.
8. O. Rockinger, T. Fechner, Pixel-level image fusion: the case of image sequences, Proc. SPIE 3374 (1998) 378–388.
9. V.S. Petrovic, C.S. Xydeas, Gradient-based multiresolution image fusion, IEEE Trans. Image Process. 13 (2) (2004) 228–237.
10. Li, Wei, Qinyong Lin, Keqiang Wang, and Ken Cai. "Improving medical image fusion method using fuzzy entropy and nonsubsampling contourlet transform." International Journal of Imaging Systems and Technology 31, no. 1 (2021): 204-214.
11. Li, Xiaosong, Fuqiang Zhou, Haishu Tan, Wanning Zhang, and Congyang Zhao. "Multi-modal medical image fusion based on joint bilateral filter and local gradient energy." Information Sciences 569 (2021): 302-325.
12. Wang, Shiying, and Yan Shen. "Multi-modal image fusion based on saliency guided in NSCT domain." IET Image Processing 14, no. 13 (2020): 3188-3201.

13. Zhu, Zhiqin, Mingyao Zheng, Guanqiu Qi, Di Wang, and Yan Xiang. "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain." IEEE Access 7 (2019): 20811-20824.
14. Li, Qiaoqiao, Weilan Wang, Guoyue Chen, and Dongdong Zhao. "Medical image fusion using segment graph filter and sparse representation." Computers in Biology and Medicine 131 (2021): 104239.
15. Q. Guihong, Z. Dali, Y. Pingfan, Medical image fusion by wavelet transform modulus maxima, Opt. Express 9 (2001) 184–190.
16. L. Yang, B.L. Guo, W. Ni, Multimodality medical image fusion based on multi-scale geometric analysis of contourlet transform, Neurocomputing 72 (2008) 203–211.
17. Xingbin Liu, Webo Mei, Huiqian:'Multi-modality medical image fusion based on image fusion decomposition framework and non-subsampled shearlet transform',2018.
18. James, A.P.,Dasarathy, B.V.:'Medical image fussion:a survey of the state-of-the-art', Inf. Fusion,2014.
19. X. Qu, J. Yan, H. Xiao, and Z. Zhu, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Automatica Sinica*, vol. 34, no. 12, pp. 1508–1514, 2008.
20. Li, Xinhua, and Jing Zhao. "A novel multi-modal medical image fusion algorithm." Journal of Ambient Intelligence and Humanized Computing 12, no. 2 (2021): 1995-2002.
21. Babu, B. S., & Narayana, M. V. (2023). Two stage multi-modal medical image fusion with marine predator algorithm-based cascaded optimal DTCWT and NSST with deep learning. Biomedical Signal Processing and Control, 85, 104921.
22. Zhang, J., Yu, K., Wen, Z., Qi, X., & Paul, A. K. (2021). 3D reconstruction for motion blurred images using deep learning-based intelligent systems. CMC-computers Materials & Continua, 66(2), 2087-2104.
23. Fu, J., Yang, J., Wang, Y., Yang, D., Yang, M., Ren, Y., & Wei, D. (2025). SMRFnet: Saliency multi-scale residual fusion network for grayscale and pseudo color medical image fusion. Biomedical Signal Processing and Control, 100, 107050.
24. Liu, Q., Kang, B., Yu, K., Qi, X., Li, J., Wang, S., & Li, H. A. (2020). Contour-maintaining-based image adaption for an efficient ambulance service in intelligent transportation systems. IEEE Access, 8, 12644-12654.
25. Wei, L., Zhu, R., Li, X., Zhao, L., Hu, X., & Zhang, X. (2024). Pixel-level structure awareness for enhancing multi-modal medical image fusion. Biomedical Signal Processing and Control, 97, 106694.
26. Zhu, H., Gowen, A., Feng, H., Yu, K., & Xu, J. L. (2020). Deep Spectral-Spatial Features of Near Infrared Hyperspectral Images for Pixel-Wise Classification of Food Products. Sensors, 20(18), 5322.
27. Kaur, H., Vig, R., Kumar, N., Sharma, A., Dogra, A., & Goyal, B. (2024). Multimodal Medical Image Fusion Utilizing Two-scale Image Decomposition via Saliency Detection. Current Medical Imaging, 20(1), e15734056260083.
28. Wang, G., Li, W., Gao, X., Xiao, B., & Du, J. (2022). Multimodal medical image fusion based on multichannel coupled neural P systems and max-cloud models in spectral total variation domain. Neurocomputing, 480, 61-75.
29. Goyal, S., Singh, V., Rani, A., & Yadav, N. (2022). Multimodal image fusion and denoising in NSCT domain using CNN and FOTGV. Biomedical Signal Processing and Control, 71, 103214.
30. Duan, Junwei, Long Chen, and CL Philip Chen. "Multifocus image fusion with enhanced linear spectral clustering and fast depth map estimation." Neurocomputing 318 (2018): 43-54.

31. Vijendran, A.S. and Ramasamy, K., 2023. Optimal segmentation and fusion of multi-modal brain images using clustering based deep learning algorithm. Measurement: Sensors, 27, p.100691.
32. Singh, P. and Bose, S.S., 2021. Ambiguous D-means fusion clustering algorithm based on ambiguous set theory: Special application in clustering of CT scan images of COVID-19. Knowledge-Based Systems, 231, p.107432.
33. Arora, P., Mehta, R., & Ahuja, R. (2024). An integration of meta-heuristic approach utilizing kernel principal component analysis for multimodal medical image registration. Cluster Computing, 1-24.
34. Zhang, H., Zuo, X., Zhou, H., Lu, T., & Ma, J. (2024, March). A Robust Mutual-Reinforcing Framework for 3D Multi-Modal Medical Image Fusion Based on Visual-Semantic Consistency. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 7, pp. 7087-7095).
35. Multimodal medical image dataset. Available at: https://github.com/bsun0802/Zero-Learning-Fast-Medical-Image-Fusion/tree/master/images/MRI-PET (Access date: 26-12-2024)
36. Multimodal medical image dataset. Available at: https://github.com/ashna111/multimodal-image-fusion-to-detect-brain-tumors/tree/master/dataset. (Access date: 26-12-2024)
37. Multimodal medical image datasets. Available at: https://www.researchgate.net/figure/Multimodal-medical-image-datasets_fig4_351121182 (Access date: 26-12-2024)
38. Liang, N. (2024). Medical image fusion with deep neural networks. *Scientific Reports*, *14*(1), 7972.
39. Luo, F., Wu, D., Pino, L. R., & Ding, W. (2025). A novel multimodel medical image fusion framework with edge enhancement and cross-scale transformer. *Scientific Reports*, *15*(1), 11657.
40. Dinh, P. H. (2025). MIF-BTF-MRN: medical image fusion based on the bilateral texture filter and transfer learning with the ResNet-101 network. *Biomedical Signal Processing and Control*, *100*, 106976.
41. Do, O. C., Luong, C. M., Dinh, P. H., & Tran, G. S. (2024). An efficient approach to medical image fusion based on optimization and transfer learning with VGG19. *Biomedical Signal Processing and Control*, *87*, 105370.
42. Dinh, P. H., & Giang, N. L. (2024). Medical image fusion based on transfer learning techniques and coupled neural P systems. *Neural Computing and Applications*, *36*(8), 4325-4347.
43. Dinh, P. H., Vu, V. H., & Giang, N. L. (2024). A new approach to medical image fusion based on the improved extended difference-of-gaussians combined with the coati optimization algorithm. *Biomedical Signal Processing and Control*, *93*, 106175.
44. Dinh, P. H. (2023). A novel approach using the local energy function and its variations for medical image fusion. *The Imaging Science Journal*, *71*(7), 660-676.
45. Dinh, P. H. (2023). Combining spectral total variation with dynamic threshold neural P systems for medical image fusion. *Biomedical Signal Processing and Control*, *80*, 104343.

46. Dinh, P. H. (2025). Enhancing Medical Image Fusion Through Advanced Decomposition and Optimization Methods. *Digital Signal Processing*, 105315.
47. Dinh, P. H. (2021). A novel approach based on grasshopper optimization algorithm for medical image fusion. *Expert Systems with Applications*, *171*, 114576.