

Application of Convolutional Neural Networks and Rolling Guidance Filter in Image Fusion for Detecting Brain Tumors

S. Karthikeyan^{1*}, P. Velmurugadass²

¹Kalasalingam Academy of Research and Education, Tamilnadu, India

* Corresponding Author Email: mr.kaarthik@gmail.com - ORCID: 0009-0002-9033-4065

²Kalasalingam Academy of Research and Education, Tamilnadu, India

Email: dass.spv@gmail.com - ORCID: 0000-0002-4036-0532

Article Info:

DOI: 10.22399/ijcesen.621

Received : 14 October 2024

Accepted : 12 December 2024

Keywords :

Guided Filter,
Image Fusion,
Convolutional Neural Networks,
Rolling Guidance,
Quantitative Evaluation.

Abstract:

Medical image fusion is the technique of integrating images from several medical imaging modalities without causing any distortion or information loss. By preserving every feature in the fused image, it increases the value of medical imaging for the diagnosis and treatment of medical conditions. A novel fusion mechanism for multimodal image data sets is proposed in this paper. Each of the source image is smoothed using cross guided filter in the initial step. Guided filter output is further smoothed to remove fine structures using rolling guidance filter. Then details (high frequency) of each source image are extracted by subtracting the rolling guidance filter output from corresponding source image. These details are fed to convolutional neural networks to obtain decision maps. Finally, the source images are fused based on decision map using maximum rule of combination. We assessed the performance of our suggested methodology using several pairs of medical imaging datasets that are accessible to the general public. According to the quantitative evaluation, the recommended fusion strategy for multimodal image fusion improves the average IE by 12.4%, MI by 41.8%, SF by 21.4%, SD by 22.81%, MSSIM by 31.1%, and Q_(AB/F) by 39% when compared to existing methods, which makes it appropriate for use in the medical field for accurate diagnosis.

1. Introduction

A variety of sensors can now capture a wide range of images owing to the quick development of sensor technologies. Multimodal biomedical imaging commonly involves the use of Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Single Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), Green Fluorescent Protein (GFP) & Phase Contrast (PC) images. The Cross-section, sagittal plane, coronal plane and any inclined plane of the human body can all be displayed by MRI, which helps to show lesions in three dimensions and pinpoint their locations. CT scans show how organs and tissues absorb X-rays. The brighter zone in a CT scan indicates a higher-density portion, such as bone tissue having high calcium content. Consequently, a CT scan can identify a human body's high-density tissue region. Radiation tomography includes PET and SPECT scans. According to their imaging concept, the various concentrations of tracer

molecules that enter the human circulatory system in various tissues correspond to varying intensities of tissue activity. Thus, the ability of SPECT and PET to identify bodily tissue with aberrant metabolism helps in the identification and localization of illnesses. In molecular biology, the two most used imaging techniques are GFP and PC. In the latter, the structural properties of cells are revealed, and it is useful to note any minute modifications to the mitochondria, cytoplasm, or nucleus. The distribution of proteins is seen in the former. The PC image has an exceptional grayscale spatial resolution, while the GFP image shows the protein distributions. The PC-GFP fusion aims to identify the intense protein region and preserve the structural elements of the phase contrast image. The fusion image facilitates biomedical analysis with its improved details and precise protein localization. Still, one sensor can only record one modal image. Compared to the single modal image, the fusion image has better visual quality and is more detailed. Because of this, image fusion methods are widely

employed in many different disciplines, including object identification [1-9], computer-aided diagnosis and detection [1-3], and image retrieval and classification [4-6].

While many fusion methods have been developed, the majority of them rely on the multiscale and multi-resolution transformation [10–12] (also known as MST-MRT). Algorithms for fusion based on MST-MRT can extract multiscale features at many resolutions. Furthermore, different layers can employ different fusion rules to enhance the fusion outcomes. Numerous MST-based fusion techniques have been developed, including Non-Subsampled Shearlet transform (NSST) [13-15], Non-Subsampled contourlet transform (NSCT) [16–18], wavelets [19–21], and pyramids [22-24]. Wavelet-based fusion approaches extract multi-scale characteristics using wavelet transform. Furthermore, wavelet coefficients can be fused using flexible methods. Wavelet inverse transformation is used to produce the fusion images. Despite having a versatile decomposition strategy that employs numerous wavelet basis functions, wavelet-based fusion algorithms lose a large number of coefficients during the process of fusion, resulting in a fuzzy fused image. Pyramid-based fusion algorithms share some of the same benefits and drawbacks as wavelet-based fusion techniques. The loss of detail during the pyramid deconstruction and fusion process cannot be prevented. With the benefits of multi-scale and multi-directionality, NSST and NSCT represent a new type of rapid transformation. To a certain degree, the edges and details of the fused image can be preserved using fusion algorithms based on NSST and NSCT. Different fusion procedures, however, can quickly cause the loss of high-frequency coefficients, which leaves details obscure and edges fuzzy. Besides, a popular fusion algorithm is sparse representation (SR) [25–27]. Coefficient optimization and dictionary establishment are the two key components of SR. Due to its rapid speed and broad range of applications, SR is frequently utilized in image denoising, information fusion, and restoration [28,29]. SR-based fusion algorithms nevertheless possess a lot of flaws even with good fusion results. For instance, choosing the dictionary size can be challenging. An excessively large dictionary will result in a slow fusion speed and a lengthy optimization process. A too-small size dictionary will result in incomplete information and an unsatisfactory fusion output. Furthermore, selecting the optimization algorithm is challenging. The effects of various optimization algorithms on fusion outcomes vary. Because of their quick speed and effective filtering, filter-based algorithms have been increasingly popular in the imaging sector recently.

A proposal for guided image filter [30] was made in 2013. While guided image filtering has been a significant advancement in the imaging field, its ability to preserve edges and remove noise is limited because it solely takes spatial weights into account. In 2014, the rolling guiding filter [31] was proposed as a solution to these issues. It introduces several guidance processes and takes into account both spatial and range weights. As a result, the rolling guidance filter has the ability to simultaneously eliminate microscopic structure and preserve large-scale structure. Deep convolutional neural networks [32,33] are currently extensively employed in imaging-related disciplines. Convolution, pooling, and activation layers are the three primary layers that make up CNN. After being trained, CNN can fit a wide variety of data. Furthermore, CNN can adapt to a wide range of difficult tasks because it is simple to construct various network configurations. CNN is therefore applicable to a wide range of tasks, including super-resolution of images [34,35], image classification [36], image segmentation [37], and image fusion [38].

2. Related work

2.1 Guided Image Filter

A guided image filter (GIF) is an explicit image filtering technique that computes the output for each pixel by examining the statistics of its surrounding neighborhood. This method enables the filter to adapt to local image characteristics, effectively smoothing the image while preserving key features such as edges [30]. The output F_i of a guided image filter at pixel 'i' is computed using a local linear model based on the guidance input P focused at pixel 'k' in a square dimensioned window w_k . It involves computation of local mean as given from equation 1 to equation 4.

$$\mu_P(k) = \frac{1}{|w_k|} \sum_{i \in w_k} P(i) \quad (1)$$

$$\mu_I(k) = \frac{1}{|w_k|} \sum_{i \in w_k} I(i) \quad (2)$$

$$\mu_{IP}(k) = \frac{1}{|w_k|} \sum_{i \in w_k} I(i)P(i) \quad (3)$$

$$\mu_{P^2}(k) = \frac{1}{|w_k|} \sum_{i \in w_k} (P(i))^2 \quad (4)$$

Using the local mean values, the linear coefficients of guided filter are evaluated by equation 5 and equation 6.

$$m_k = \frac{\mu_{IP}(k) - \mu_P(k)\mu_I(k)}{(\mu_{P^2}(k) - (\mu_P(k))^2) + \epsilon} \quad (5)$$

$$n_k = \mu_i(k) - m_k \mu_p(k) \quad (6)$$

The output of filter F due to pixel ' i ' is computed as follows:

$$F_i = m_k P_i + n_k, \forall i \in w_k \quad (7)$$

Where ε is the regularization control parameter. For multiple windows w_k centered at k each having a common pixel i , calculate the mean estimates of all F_i to resolve this overlap, as expressed in equation 8, which represents the output of the filter.

$$F_i = \overline{m}_i P_i + \overline{n}_i \quad (8)$$

In this work, the guided filter is utilized to decompose the source images. The CT source image guides the MRI input, while the MRI source image guides the CT input within the guided filtering process.

2.2 Rolling Guidance filter

Ever since it was first introduced, the rolling guiding filter has been widely used in imaging. It fully regulates the smoothing of details under a scale measure while maintaining edge integrity, depending on the rolling guidance. The rolling guiding filter is faster at achieving convergence than other edge-preserving filters because it uses iteration in the filtering process. Furthermore, it has the ability to maintain the extensive structure removal and edge recovery without artifacts by using automated structures [31].

Process of Small Structure Removal

The tiny structure is eliminated in this method by applying a Gaussian filter. Texture, small target information and noise are typically included in the small structure. Assume that I is the input image, G is the output image, and that (i, j) is the image pixel coordinate vector and σ_s is the standard deviation, then equation 9 can be used to define Gaussian filter.

$$G(i) = \frac{1}{K_i} \sum_{j \in N(i)} \exp\left(-\frac{\|i-j\|^2}{2\sigma_s^2}\right) I(j) \quad (9)$$

Where $N(i)$ represents the neighborhood of pixel i and $K_i = \sum_{j \in N(i)} \exp\left(-\frac{\|i-j\|^2}{2\sigma_s^2}\right)$ is employed for normalization. The filter has the ability to remove small structures when the scale is smaller than σ_s in the scale space.

Process of Edge Recovery

This method consists of joint filtering stages as well as an iterative process. The reason guided filtering

was chosen as the joint filtering method is because it offers good edge preservation performance in addition to great computational efficiency. J_1 is initially configured to be the Gaussian filtering output G . J_{t+1} is the filter output of the t^{th} iterative process, which is generated from guided filtering using the guiding image as J_t and the input as image I . Here is an expression for the edge recovery process.

$$J_{t+1}(i) = \frac{1}{K_i} \sum_{j \in N(i)} \exp\left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J_t(i) - J_t(j)\|^2}{2\sigma_r^2}\right) I(j) \quad (10)$$

Where $K_i = \sum_{j \in N(i)} \exp\left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J_t(i) - J_t(j)\|^2}{2\sigma_r^2}\right)$ is a normalization parameter and σ_r is a control parameter used to control the range weights. In our work, rolling guidance filter is employed on the output of cross guided filter to remove all tiny structures that still persist.

2.3 Convolutional Neural Network (CNN)

As seen in Figure 1, a CNN is used in this study to create weight maps for the source images based on high-frequency image information. The network consists of two different CNN branches that represent weights and structures, with each receiving high-frequency (having large variation in intensity) sub-bands from each of the images. This form of architecture, known as a Siamese network, allows for simultaneous training of both pictures, allowing it to distinguish between each source image's high-frequency sub-band. Figure 1 depicts a schematic of the Siamese network's two branches [32,33].

Each branch has three convolutional layers with 64, 128 and 256 filters, respectively, and one max-pooling layer. Both the max-pooling layer and the convolutional layers use a 2×2 filter with a stretch of 2 and a 3×3 filter with a stretch of 1. To minimize memory usage and computation time, we have eliminated two fully connected layers from our approach. This allows for the processing of images of any dimension, rather than just fixed-size images. The concatenated 512 feature maps are fed into a 2D vector softmax layer, which produces a probability distribution for the two classes. We utilize stochastic gradient descent (SGD) for optimization, applying the softmax loss function. Additionally, the batch size is set to 128, with a weight decay of 0.00001 and a momentum of 0.9. The CNN fusion process for high-frequency images consists of three steps: pattern recognition, preliminary segmentation, and consistency

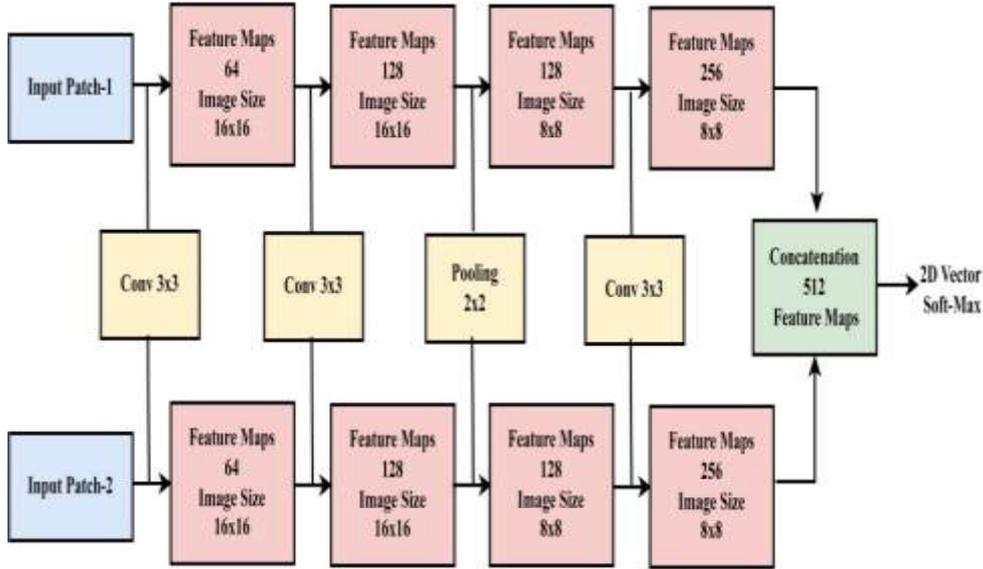


Figure 1. Feature extraction using the Siamese network built on CNN.

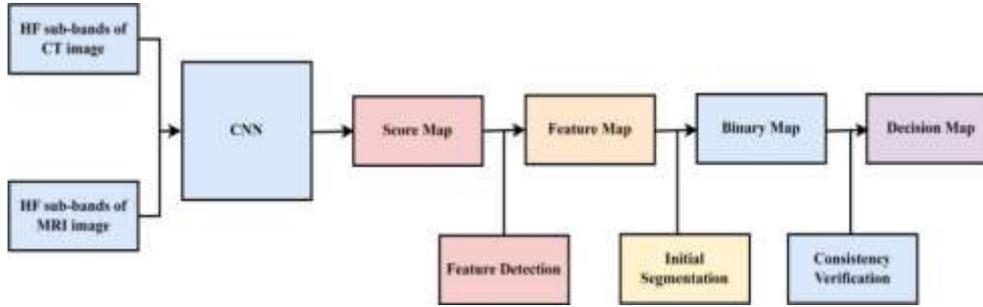


Figure 2. Decision map generation with CNN using high frequency bands.

checking, as illustrated in Figure 2. Each of these phases is discussed in the following sections.

(a) **Pattern Recognition:** To generate a score map, high-frequency sub-bands from each image are sent through two separate network branches. The score map uses coefficients with weights ranging from 0 to 1 to represent the 16×16 equivalent blocks produced from the high-pass sub-images. In order to create a feature map $M(p, q)$ of similar size, the overlapping regions in the score map are averaged.

(b) **Preliminary Segmentation:** The binary map $B(p, q)$ must be derived once the feature map $M(p, q)$ has been obtained. The binary maps are created using a choose-max selection approach with a threshold of 0.5 to extract features that are more helpful. Equation 11 is utilized to determine this concept.

$$B(p, q) = \begin{cases} 1 & \text{if } M(p, q) \geq 0.5 \\ 0 & \text{if } M(p, q) < 0.5 \end{cases} \quad (11)$$

The following step is used to process the binary map $B(p, q)$, by validating consistency.

(c) **Consistency testing:** Pixels that are incorrectly classified and have values that are substantially different from those of their neighbors are known as discontinuity points in the binary map.

To overcome this, our approach removes singularities from the focus map using an 8×8 window consistency check. Furthermore, the fused image may have artifacts introduced by the decision map. In order to reduce the artifact issue, we employ a guided filtering method that maintains edges. The normalization parameter " ϵ " and window size " r " are set to 0.1 and 5, respectively.

3. Proposed mechanism

Figure 3 depicts a flow diagram for our suggested fusion process. The detailed technique is as follows: Step 1: Take the source images intended for fusion and designate them as I_1 and I_2 .

Step 2: Apply cross guided filter on I_1 and I_2 using equation 1 to equation 5 and compute filter outputs. Filter output for first source image,

$$B_1 = \text{guidedfilter}(I_2, I_1, r, \epsilon)$$

Filter output for second source image,

$$B_2 = \text{guidedfilter}(I_1, I_2, r, \epsilon)$$

Where r is the radius that controls the size of neighborhood during the smoothing operation, which is taken as $r_k = 5$ and ε is the regularization control parameter that controls preservation of edges, which is taken as 10^{-6} .

Step 3: Apply rolling guidance filter on output of cross guided filter using equation 6 and equation 7 to remove tiny structures.

$$Base_1 = RGF(B_1, \sigma_s, \sigma_r)$$

$$Base_2 = RGF(B_2, \sigma_s, \sigma_r)$$

Step 4: Generate the high frequency (detail) layers of source images by subtracting the rolling guidance filter output from respective source image.

$$detail_1 = I_1 - Base_1$$

$$detail_2 = I_2 - Base_2$$

Step 5: Feed the detail layers of source images to the convolutional neural network and obtain decision map $D(p, q)$ from the feature weights.

Step 4: Combine the source images based on decision map using maximum combination rule described in equation 3 12 to generate the fused image.

$$F(p, q) = D(p, q) * I_1(p, q) + (1 - D(p, q)) * I_2(p, q) \quad (12)$$

Where $D(p, q)$ represents the decision map obtained from high frequency sub bands using CNN, $I_1(p, q)$ and $I_2(p, q)$ are the source images while $F(p, q)$ is the fused image.

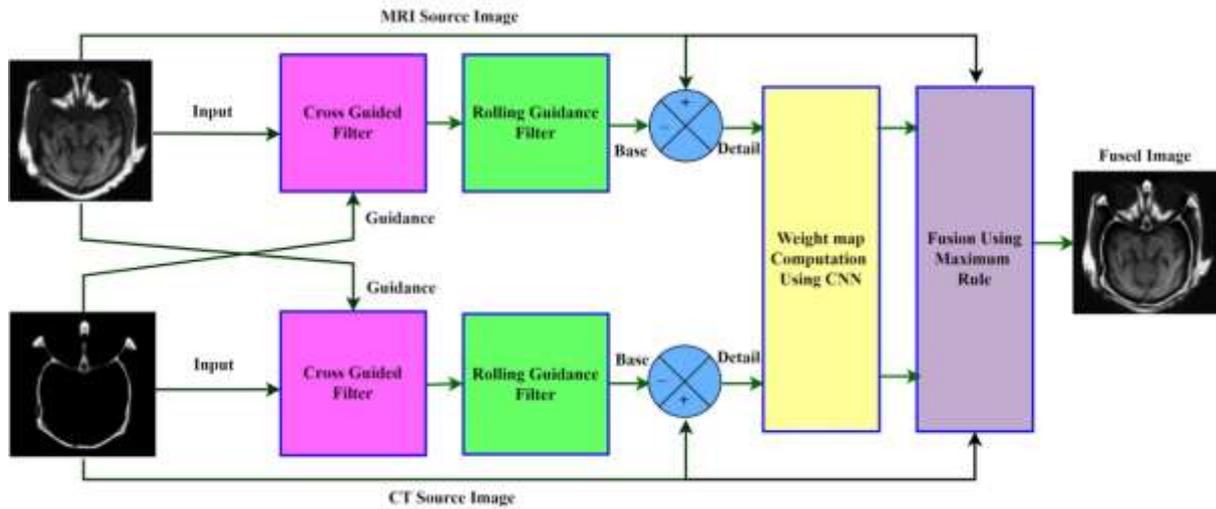


Figure 3. Process flow of proposed mechanism of fusion.

4. Qualitative Analysis

To demonstrate the effectiveness of the proposed fusion technique, five pairs of brain MRI and CT scans—labeled as "Dataset-A," "Data set-B," "Data set-C," "Data set-D," and "Data set-E"—were selected. Data set-A features the brain of a healthy individual; Data set-B depicts the brain of a patient who experienced a fatal stroke, Data set-C showcases a brain with a neoplastic tumor, Dataset-D contains sagittal plane pictures of the cranial structure, while Dataset-E shows the brain of an individual with cerebellar metastases. Additionally, MR-T2 and SPECT datasets for metastatic bronchogenic illness, as well as MR-T2 and PET datasets for Alzheimer's disease, are included. Each dataset consists of 256 distinct shades of gray and has a resolution of 256 x 256 pixels. All datasets were obtained from the publicly available benchmark brain atlas at Harvard Medical School, accessible at

<http://www.med.harvard.edu/aanlib/home.html> [39]. The results of existing methods are collected from [40-43]. Figure 4 illustrates the image fusion results for Dataset-A using various methods. Figures 4(a) and (b) present the CT and MRI images, respectively, while Figures 4(c)–(h) show the fusion results from guided filter techniques, the proposed method, sparse representation (SR), convolutional neural networks (CNN), non-subsampled shearlet transformation (NSST), and discrete wavelet transformation (DWT). The soft tissues components from the MRI and the bone structures from the CT scan are primarily preserved in the fused data. The techniques do, however, differ only slightly in terms of contrast preservation and detail. The discrepancies between the compared procedures are indicated by a yellow rectangle. The highlighted area exhibits slightly reduced intensity in the fusion results, as seen in Figures 4(c) and (d). On the other hand, the fusion pictures produced by CNN and NSST, which are displayed in Figures 4(e) and (f),

may lose certain MRI details but tend to collect large data from the CT image. The guided filter preserves the characteristics of the original photos while exhibiting visual clarity comparable to our suggested method; however, as illustrated in Figures 4(g) and (h), its contrast is inferior to that of our approach.

Figure 5 shows Data set-B, the second batch of medical imagery. Using DWT causes a loss of dense tissue data, such as bone features, as seen in Figure

5(d), which results in a bad visual effect. Figure 5(c) likewise shows low contrast. The outcomes of the remaining three approaches that are now in use did not differ significantly from one another. In contrast to the results of guided filter method, the proposed technique in this case has trouble collecting the column and row-wise fluctuations in intensity from the CT image, resulting in a lower preservation of CT information.

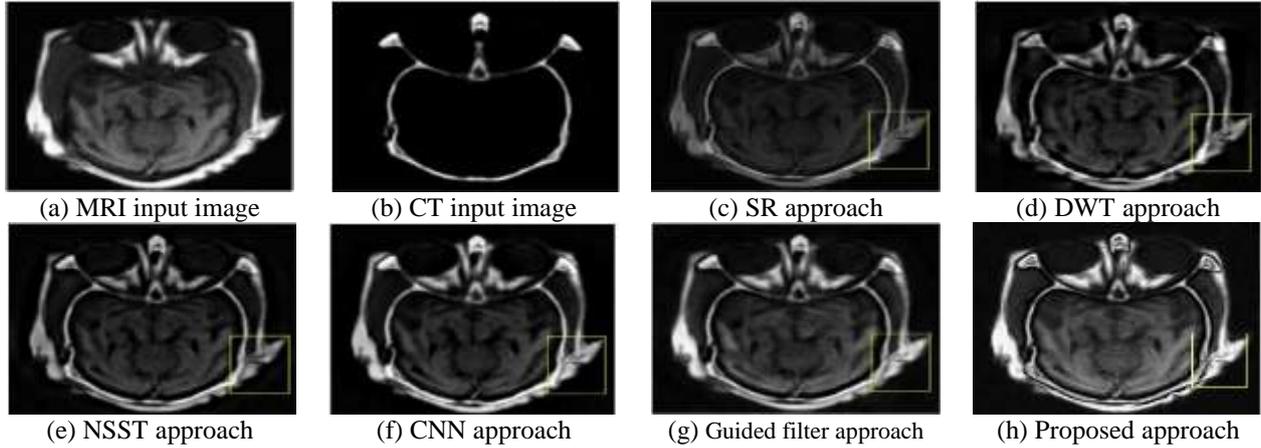


Figure 4. Fusion results of CT-MRI of healthy brain.

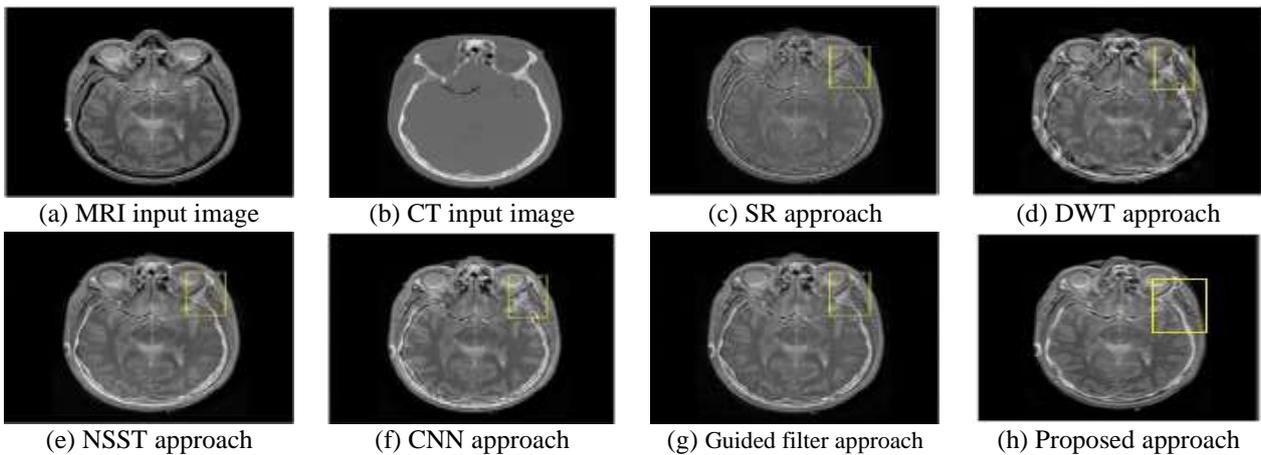


Figure 5. Fusion results of CT-MRI of fatal stroke.

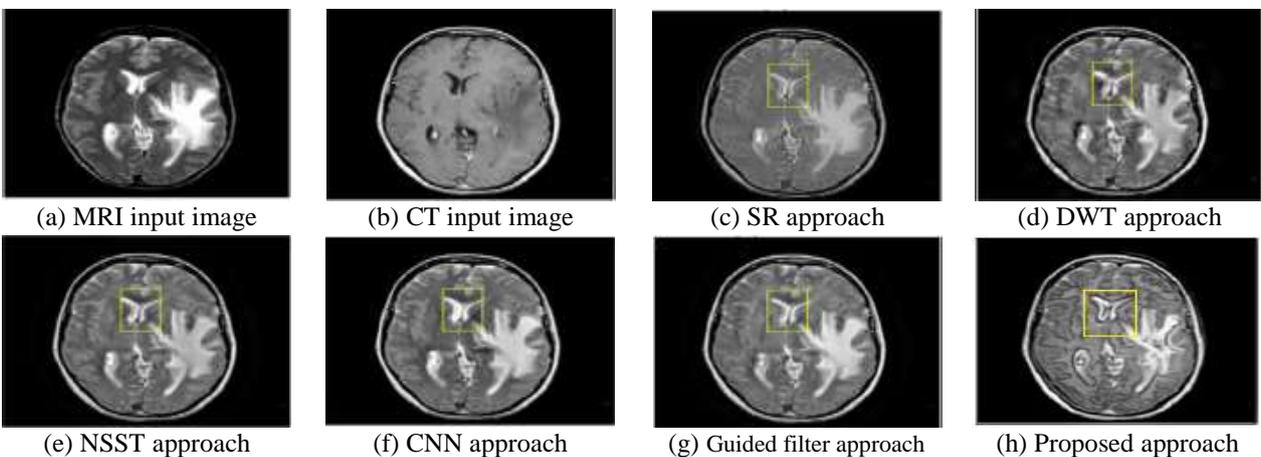


Figure 6. Fusion results of CT-MRI of neoplastic tumor.

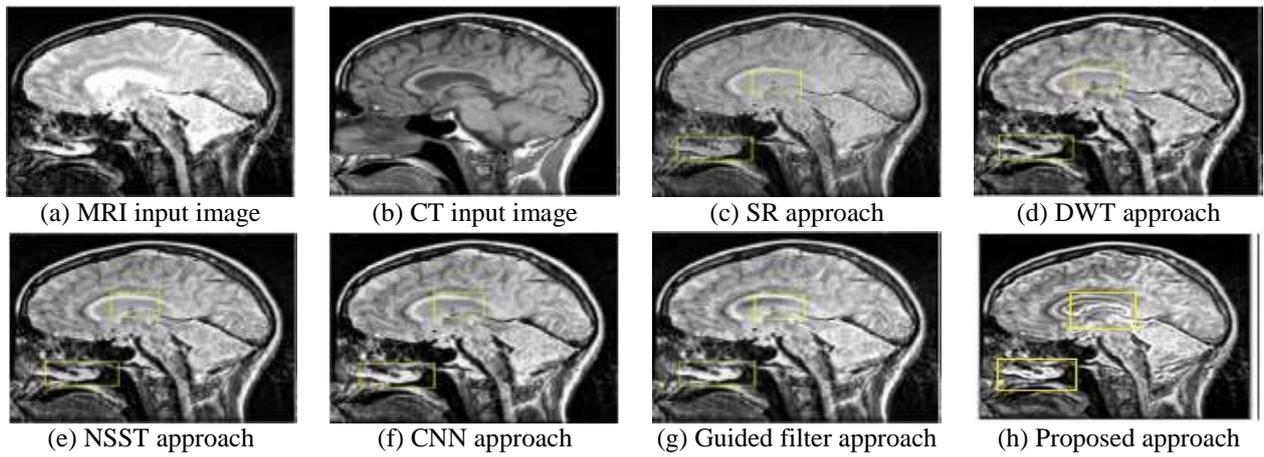


Figure 7. Fusion results of CT-MRI of brain skull.

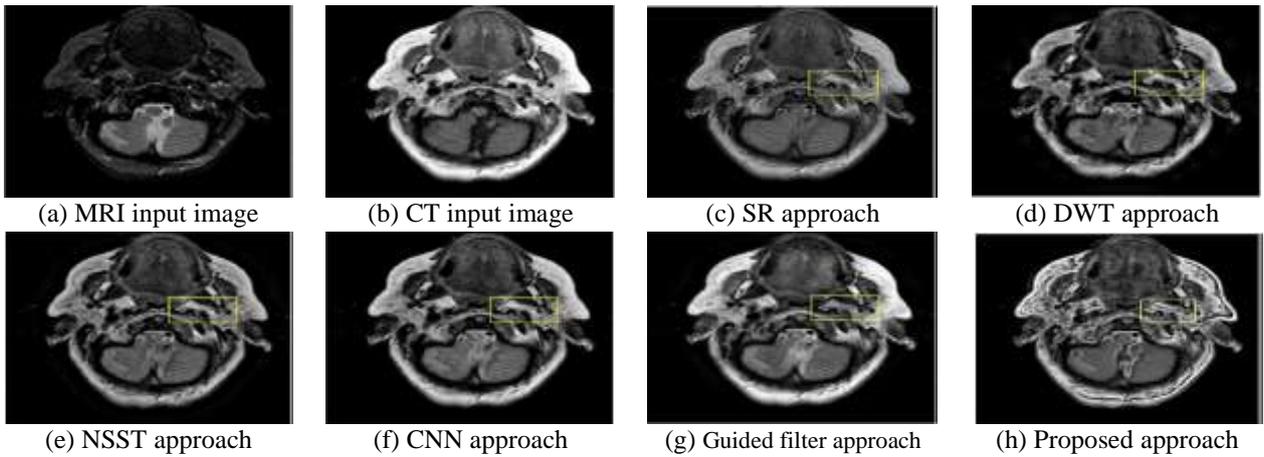


Figure 8. Fusion results of CT-MRI of cerebella metastasis.

The fused image generated by the proposed algorithm exhibits excellent contrast and fully preserves the soft tissue information, as illustrated in Fig. 6(h). This highlights the advantages of the proposed technique compared to existing methods in the third medical image set, Data set-C. In contrast, the SR, DWT, and NSST techniques fail to retain the details of the source images, resulting in insufficient information about the bone structure, as seen in figure 7 and figure 8 for Data set-D and Dataset-E respectively. The results from our suggested method demonstrate greater detail, sharper edges, and enhanced contrast.

5. Quantitative Analysis

Measurement of fusion performance requires both a qualitative and quantitative evaluation criteria. In this study, the effectiveness of various fusion techniques is assessed using quantitative assessment measures such as mutual information (MI), mean structural similarity (MSSIM), image entropy (IE), spatial frequency (SF), margin information retention ($Q_{AB/F}$) and standard deviation (SD)[40-43].

1. In MI, two events are related to one another. These two independent random variables, Y and

Z , have the following mutual information described by equation 13.

$$MI(Y, Z) = \sum_{z \in Z} \sum_{y \in Y} p(y, z) \log_2 \frac{p(y, z)}{p(y)p(z)} \quad (13)$$

where $p(y, z)$ denotes the probability distribution of Y and Z combined, and $p(y)$ and $p(z)$ denote the respective marginal probability distributions of Y and Z . The total mutual information metric could be formulated using equation 14 to ascertain the disparity in fusion quality after the quantity of MI between the fused image F and each of each of the input images (A, B) has been measured.

$$MI_F^{AB} = MI(A, F) + MI(B, F) \quad (14)$$

A higher MI score indicates that more information has been retrieved from the original images.

2. The standard deviation (SD) is a metric used to quantify the level of divergence in an average collection of image data. equation 15 is used to calculate the standard deviation for the fused image. Where $F(m, n)$ is the fused picture pixel value at the location (m, n) and μ is referred to as the mean. This statistic measures the clarity

of the combined output image; the higher the *SD* value, the higher the quality of the image.

$$SD = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (F(m, n) - \mu)^2} \quad (15)$$

$$\mu = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N F(m, n) \quad (16)$$

- The total of the column frequency and row frequency is known as the spatial frequency (SF). It is calculated using equation 17. Here, the image size is shown by $M \times N$, while the pixel value is indicated by $I(i, j)$. The resolution of the fused image increases with the score of this parameter.

$$SF(m, n) = \sqrt{|RF(m, n)|^2 + |CF(m, n)|^2} \quad (17)$$

$$RF(m, n) = \sqrt{\frac{1}{M \times N} \sum_{i=2}^M \sum_{j=2}^N [I(m, n) - I(m, n - 1)]^2} \quad (18)$$

$$CF(m, n) = \sqrt{\frac{1}{M \times N} \sum_{i=2}^M \sum_{j=2}^N [I(m, n) - I(m - 1, n)]^2} \quad (19)$$

- Image entropy (IE) is a measure of how much information is contained within a fused image. If the distribution of grey levels in a picture I is $P = \{P_0, P_1, \dots, P_{L-1}\}$, where P_k is the probability that the k^{th} grey level will appear in the image, and L is the number of grey levels (256 for a grey scale image), then IE might be calculated as follows:

$$IE = - \sum_{i=0}^{L-1} p(i) \log_2 p(i) \quad (20)$$

The amount of information in the fused image increases with the IE score.

- The following formula can be used to determine the Mean Structural Similarity Index Measure (MSSIM), which is a helpful indicator of picture similarity.

$$MSSIM = \frac{SSIM(A,F) + SSIM(B,F)}{2} \quad (21)$$

$$SSIM = \frac{(2\mu_a\mu_b + C1)(2\sigma_{ab} + C2)}{(\mu_a^2 + \mu_b^2 + C1)(\sigma_a^2 + \sigma_b^2 + C2)} \quad (22)$$

Where μ_a is the average value of a and μ_b is the average value of b , σ_{ab} is the covariance of a and b , σ_a^2 is the variance of a , and σ_b^2 is the variance of b . To avoid instability that results from a division with zero, two constants, $C1$ and $C2$, are used. The range of SSIM values is 0 to 1, where 1 denotes exceptional quality and 0 denotes poor quality. As the MSSIM score rises, the distortion effect in the fused image diminishes.

- $Q_{AB/F}$ represents the degree to which the edge information from the input images transitions into the fused image. The evaluation is as follows:

$$Q_{AB/F} = \frac{\sum_{i=1}^M \sum_{j=1}^N (Q_{AF}(i,j)W_A(i,j) + Q_{BF}(i,j)W_B(i,j))}{\sum_{i=1}^M \sum_{j=1}^N (W_A(i,j) + W_B(i,j))} \quad (23)$$

When the $Q_{AB/F}$ value rises and gets close to one, it means that there hasn't been much edge information lost in the fused image. Table 1 shows the quantitative analysis using image evaluation parameters, with the most significant results indicated in bold. The proposed image fusion method significantly outperformed others in terms of MI, IE, $Q_{AB/F}$, and MSSIM, while the remaining metrics show minimal comparability. This indicates that the method effectively preserves edge features and salient information. Figure 9 shows the average metric values for the fusion outcomes across 30 slices of each dataset, highlighting the four important assessment metrics: MI, IE, $Q_{AB/F}$, and MSSIM. The average metric values for the fusion results across 30 slices of each dataset are visually represented in figure 9, illustrating the four key assessment metrics: MI, IE, $Q_{AB/F}$, and MSSIM.

Table 1. Statistical metrics of proposed method for multimodal datasets.

Dataset type	Method	Standard Deviation (SD)	Mutual Information (MI)	Spatial Frequency (SF)	Image Entropy (H)	Edge Strength	MSSIM
Dataset-A	SR	30.82	2.57	11.68	5.8	0.5756	0.5122
	DWT	44.71	1.92	17.13	6.17	0.6073	0.5246
	NSST	44.16	2.05	17.05	6.2	0.6816	0.5366
	CNN	52.89	2.43	17.4	6.07	0.7184	0.5518
	Guided filter	52.89	2.31	16.97	6.52	0.721	0.5634
	Proposed	54.48	4.02	18.1	6.67	0.8776	0.9908
Dataset-B	SR	51.4	3.42	17.76	4.94	0.5178	0.8248
	DWT	55.73	3.19	22.01	5.19	0.5051	0.7915

	NSST	54.56	3.34	20.95	5.12	0.5887	0.816
	CNN	59.92	3.34	21.93	4.89	0.5888	0.8146
	Guided filter	55.68	3.79	20.25	5.2	0.6028	0.8207
	Proposed	60.56	5.07	21.71	5.84	0.8166	0.9898
Dataset-C	SR	61.5	3.18	20.19	4.52	0.5157	0.764
	DWT	66.53	3.12	25.11	4.86	0.5473	0.7489
	NSST	65.89	3.2	24.52	4.88	0.5971	0.7733
	CNN	69.6	3.38	25.99	4.39	0.6042	0.7775
	Guided filter	69.63	3.34	24.39	5.05	0.6119	0.7762
	Proposed	73.2	4.72	26.73	5.26	0.8051	0.9869
Dataset-D	SR	69.84	3.33	28.98	7.56	0.4964	0.6532
	DWT	76.8	3.08	35.94	7.41	0.4699	0.6263
	NSST	79.49	3.23	34.6	7.44	0.5349	0.6628
	CNN	79.84	3.26	37.03	7.31	0.5171	0.6462
	Guided filter	75.36	3.52	34.3	7.76	0.551	0.6602
	Proposed	79.06	4.64	30.38	7.65	0.8552	0.985
Dataset-E	SR	51.71	3.19	17.58	5.24	0.4823	0.7427
	DWT	55.72	2.8	22.28	5.36	0.4573	0.7098
	NSST	53.79	2.94	21.47	5.44	0.5226	0.7311
	CNN	61.11	3.18	23.06	4.83	0.5214	0.7448
	Guided filter	66.98	3.23	21.56	5.78	0.533	0.7342
	Proposed	76.05	5.83	25.5	5.94	0.8883	0.9884

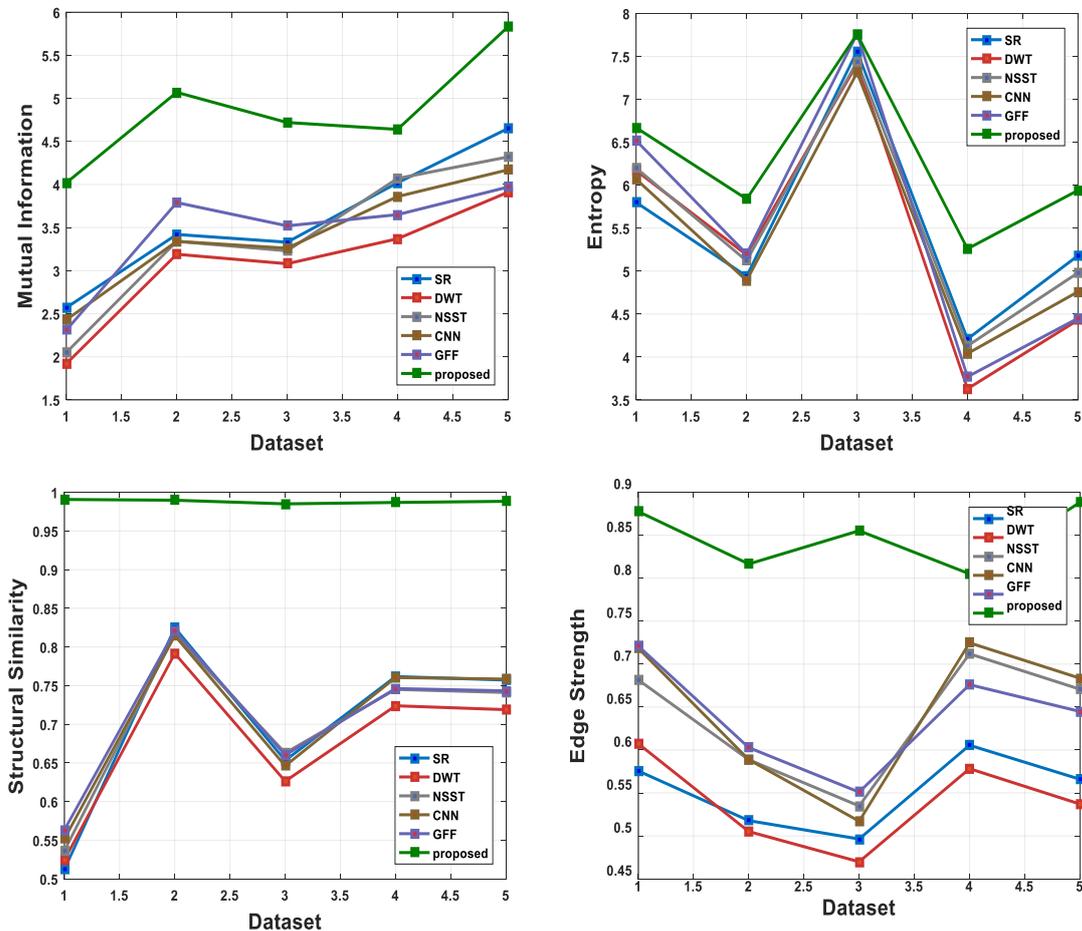


Figure 9. Average metric values for 30 slices of five different datasets.

The quantitative results indicate that the fusion outcomes of the Sparse Representation (SR) and DWT methods are unsatisfactory due to low intensity and insufficient detail regarding bone structures. While the CNN NSST, and guided filter techniques provide adequate visual quality, they fail to fully retain the edge and texture details within the highlighted yellow region in MR-CT images. Additionally, in the MR-PET and MR-SPECT fusion, the Guided Filter technique does not effectively preserve color information. In contrast, the proposed method successfully maintains salient features, delivering maximum information about bone structures and soft tissues, which results in more vivid and detailed fused images.

Three of the six metrics taken for assessment—image entropy (IE), standard deviation (SD) and spatial frequency (SF)—are commonly utilized to assess the quality of fused images and reflect the intrinsic properties of individual images. IE indicates the entropy of the fused image, while SF provides insight into the clarity of the image. SD, on the other hand, describes the contrast within the fused image. A higher SD disperses the gray level dispersion more broadly, while a higher contrast makes the fused image easier to see. The values of these metrics tend to increase because some existing techniques incorporate redundant features. To offer a more comprehensive objective analysis, this study introduces three additional metrics: mutual information (MI), Mean structural similarity index measure (MSSIM), and quality of edge information ($Q_{AB/F}$). MI evaluates the amount of data retrieved from the source images and assesses the similarity of intensity distributions between related image pairs; as more data is extracted from the original images, the MI value increases alongside the clarity and activity of the combined image. MSSIM quantifies the level of distortion in the fused image. On the other hand, $Q_{AB/F}$ assesses how much edge information is retained in the fused image from the source images. A higher $Q_{AB/F}$ value is particularly important for clinical image fusion as it enhances the accurate pathological examination of edges by integrating more edge information, including details about bone structure and texture. Convolutional Neural Networks has been used in different fields and reported in the literature [44-50].

6. Conclusion

A method of image fusion for better diagnosis of brain related tumors is proposed in this work using guided filter and convolutional neural networks. The performance of proposed algorithm is tested on diverse datasets of brain related scans of CT and

MRI. Based on the experimental results, we conclude that the proposed approach for multimodal image fusion offers significant advantages over existing methods, particularly in terms of improved image clarity, retention of important soft tissue and dense structure characteristics, and minimal edge distortion. This study also underscores the importance of using additional metrics such as MI, MSSIM, and $Q_{AB/F}$ for comprehensive evaluation of fused image quality. According to the quantitative evaluation, the recommended fusion strategy for multimodal image fusion improves the average IE by 12.4%, MI by 41.8%, SF by 21.4%, SD by 22.81%, MSSIM by 31.1%, and $Q_{AB/F}$ by 39% when compared to existing methods, which makes it appropriate for use in the medical field for accurate diagnosis. Thus, the proposed approach has the potential to significantly enhance diagnostic capabilities in medical practice.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- [1] Lu, W., Li, Z., & Chu, J. (2017). A novel computer-aided diagnosis system for breast MRI based on feature selection and ensemble learning. *Computers in Biology and Medicine*. 83;157–165. <https://doi.org/10.1016/j.compbiomed.2017.03.002>.
- [2] Lin, R., & Chuang, C. (2010). A hybrid diagnosis model for determining the types of the liver disease. *Computers in Biology and Medicine*. 40(7);665–670. <https://doi.org/10.1016/j.compbiomed.2010.06.002>.
- [3] Jabarulla, M. Y., & Lee, H.-N. (2017). Computer aided diagnostic system for ultrasound liver images: A systematic review. *Optik*. 140;1114–1126. <https://doi.org/10.1016/j.ijleo.2017.05.013>.

- [4] Kavitha, K., & Arivazhagan, S. (2014). Fuzzy inspired image classification algorithm for hyperspectral data using three-dimensional log-Gabor features. *Optik*. 125(20);6236–6241. <https://doi.org/10.1016/j.ijleo.2014.08.029>.
- [5] Lin, D., Sun, L., Toh, K.-A., Zhang, J. B., & Lin, Z. (2018). Biomedical image classification based on a cascade of an SVM with a reject option and subspace analysis. *Computers in Biology and Medicine*. 96;128–140. <https://doi.org/10.1016/j.compbiomed.2018.03.005>.
- [6] Andreini, P., Bonechi, S., Bianchini, M., Garzelli, A., & Mecocci, A. (2016). Automatic image classification for the urinoculture screening. *Computers in Biology and Medicine*. 70;12–22. <https://doi.org/10.1016/j.compbiomed.2015.12.025>.
- [7] Liao, C.-C., Xiao, F., Wong, J.-M., & Chiang, I.-J. (2010). Automatic recognition of midline shift on brain CT images. *Computers in Biology and Medicine*. 40(3);331–339. <https://doi.org/10.1016/j.compbiomed.2010.01.004>.
- [8] He, R., Cao, J., Song, L., Sun, Z., & Tan, T. (2020). Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 42(5);1025–1037. <https://doi.org/10.1109/tpami.2019.2961900>.
- [9] Köping, L., Shirahama, K., & Grzegorzec, M. (2018). A general framework for sensor-based human activity recognition. *Computers in Biology and Medicine*. 95;248–260. <https://doi.org/10.1016/j.compbiomed.2017.12.025>.
- [10] Liu, Y., Liu, S., & Wang, Z. (2015). A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion*. 24;147–164. <https://doi.org/10.1016/j.inffus.2014.09.004>.
- [11] Zhao, J., Chen, Y., Feng, H., Xu, Z., & Li, Q. (2014). Infrared image enhancement through saliency feature analysis based on multi-scale decomposition. *Infrared Physics & Technology*. 62;86–93. <https://doi.org/10.1016/j.infrared.2013.11.008>.
- [12] Ben Hamza, A., He, Y., Krim, H., & Willsky, A. (2005). A multiscale approach to pixel-level image fusion. *Integrated Computer-Aided Engineering*. 12(2);135–146. <https://doi.org/10.3233/ica-2005-12201>.
- [13] Singh, S., Gupta, D., Anand, R. S., & Kumar, V. (2015). Nonsubsampled shearlet based CT and MR medical image fusion using biologically inspired spiking neural network. *Biomedical Signal Processing and Control*. 18;91–101. <https://doi.org/10.1016/j.bspc.2014.11.009>.
- [14] Gao, G., Xu, L., & Feng, D. (2013). Multi-focus image fusion based on non-subsampled shearlet transform. *IET Image Processing*. 7(6);633–639. <https://doi.org/10.1049/iet-ipr.2012.0558>.
- [15] Luo, X., Zhang, Z., Zhang, B., & Wu, X. (2017). Image Fusion With Contextual Statistical Similarity and Nonsubsampled Shearlet Transform. *IEEE Sensors Journal*. 17(6);1760–1771. <https://doi.org/10.1109/jsen.2016.2646741>.
- [16] Zhang, Q., & Guo, B. (2009). Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Processing*. 89(7);1334–1346. <https://doi.org/10.1016/j.sigpro.2009.01.012>.
- [17] Bhatnagar, G., Wu, Q. M. J., & Liu, Z. (2013). Directive Contrast Based Multimodal Medical Image Fusion in NSCT Domain. *IEEE Transactions on Multimedia*. 15(5);1014–1024. <https://doi.org/10.1109/tmm.2013.2244870>.
- [18] Yang, Y., Que, Y., Huang, S., & Lin, P. (2016). Multimodal Sensor Medical Image Fusion Based on Type-2 Fuzzy Logic in NSCT Domain. *IEEE Sensors Journal*. 16(10);3735–3745. <https://doi.org/10.1109/jsen.2016.2533864>.
- [19] Hill, P., Al-Mualla, M. E., & Bull, D. (2017). Perceptual Image Fusion Using Wavelets. *IEEE Transactions on Image Processing*. 26(3);1076–1088. <https://doi.org/10.1109/tip.2016.2633863>.
- [20] Li, S., & Yang, B. (2008). Multifocus image fusion by combining curvelet and wavelet transform. *Pattern Recognition Letters*. 29(9);1295–1301. <https://doi.org/10.1016/j.patrec.2008.02.002>.
- [21] Qu, G., Zhang, D., & Yan, P. (2001). Medical image fusion by wavelet transform modulus maxima. *Optics Express*. 9(4);184–190. <https://doi.org/10.1364/OE.9.000184>.
- [22] Zhang, Q., Wang, L., Li, H., & Ma, Z. (2011). Similarity-based multimodality image fusion with shiftable complex directional pyramid. *Pattern Recognition Letters*. 32(13);1544–1553. <https://doi.org/10.1016/j.patrec.2011.06.002>.
- [23] Du, J., Li, W., Xiao, B., & Nawaz, Q. (2016). Union Laplacian pyramid with multiple features for medical image fusion. *Neurocomputing*. 194;326–339. <https://doi.org/10.1016/j.neucom.2016.02.047>.
- [24] Du, J., Li, W., & Xiao, B. (2017). Anatomical-functional image fusion by information of interest in local Laplacian filtering domain. *IEEE Transactions on Image Processing*. 26(12);5855–5866. <https://doi.org/10.1109/TIP.2017.2745202>.
- [25] Zhang, Q., & Levine, M. D. (2016). Robust Multi-Focus Image Fusion Using Multi-Task Sparse Representation and Spatial Context. *IEEE Transactions on Image Processing*. 25(5);2045–2058. <https://doi.org/10.1109/tip.2016.2524212>.
- [26] Liu, Y., Chen, X., Ward, R., & Wang, Z. (2016). Image fusion with convolutional sparse representation. *IEEE Signal Processing Letters*. 23(12);1882–1886. <https://doi.org/10.1109/LSP.2016.2618776>.
- [27] Li, S., Yin, H., & Fang, L. (2012). Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Transactions on Biomedical Engineering*. 59(12);3450–3459. <https://doi.org/10.1109/TBME.2012.2217493>.
- [28] Yu, N., Qiu, T., Bi, F., & Wang, A. (2011). Image Features Extraction and Fusion Based on Joint Sparse Representation. *IEEE Journal of Selected Topics in Signal Processing*. 5(5);1074–1082. <https://doi.org/10.1109/jstsp.2011.2112332>.
- [29] Yang, B., & Li, S. (2010). Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*.

- 59(4);884–892.
<https://doi.org/10.1109/TIM.2009.2026612>.
- [30] He, K., Sun, J., & Tang, X. (2013). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 35(6);1397–1409. <https://doi.org/10.1109/TPAMI.2012.213>.
- [31] Zhang, Q., Shen, X., Xu, L., & Jia, J. (2014). Rolling guidance filter. In European Conference on Computer Vision (ECCV). Springer. 815–830. https://doi.org/10.1007/978-3-319-10578-9_53.
- [32] Passalis, N., & Tefas, A. (2019). Training lightweight deep convolutional neural networks using bag-of-features pooling. *IEEE Transactions on Neural Networks and Learning Systems*. 30(6);1705–1715. <https://doi.org/10.1109/TNNLS.2018.2872995>.
- [33] Cao, C., Lan, C., Zhang, Y., Zeng, W., Lu, H., & Zhang, Y. (2019). Skeleton-based action recognition with gated convolutional neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*. 29(11);3247–3257. <https://doi.org/10.1109/TCSVT.2018.2879913>.
- [34] Jebadurai, J., & Dinesh Peter, J. (2018). Super-resolution of retinal images using multi-kernel SVR for IoT healthcare applications. *Future Generation Computer Systems*. 83;338–346. <https://doi.org/10.1016/j.future.2018.01.058>.
- [35] Mei, S., Jiang, R., Li, X., & Du, Q. (2020). Spatial and Spectral Joint Super-Resolution Using Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*. 58(7);4590–4603. <https://doi.org/10.1109/tgrs.2020.2964288>.
- [36] Bai, X., Xue, R., Wang, L., & Zhou, F. (2019). Sequence SAR Image Classification Based on Bidirectional Convolution-Recurrent Network. *IEEE Transactions on Geoscience and Remote Sensing*. 57(11);9223–9235. <https://doi.org/10.1109/tgrs.2019.2925636>.
- [37] Moeskops, P., Viergever, M. A., Mendrik, A. M., de Vries, L. S., Benders, M. J. N. L., & Isgum, I. (2016). Automatic Segmentation of MR Brain Images With a Convolutional Neural Network. *IEEE Transactions on Medical Imaging*. 35(5);1252–1261. <https://doi.org/10.1109/tmi.2016.2548501>.
- [38] Liu, X., Deng, C., Chanussot, J., Hong, D., & Zhao, B. (2019). StfNet: A Two-Stream Convolutional Neural Network for Spatiotemporal Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*. 57(9);6552–6564. <https://doi.org/10.1109/tgrs.2019.2907310>.
- [39] Johnson, K. A., & Becker, J. A. (n.d.). *The whole brain atlas*. Retrieved from <http://www.med.harvard.edu/aanlib/home.html>
- [40] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. 13(4);600–612. <https://doi.org/10.1109/TIP.2003.819861>.
- [41] Petrovic, V., & Xydeas, C. (2004). Evaluation of image fusion performance with visible differences. In European Conference on Computer Vision (ECCV). Springer. 380–391. https://doi.org/10.1007/978-3-540-24672-5_30.
- [42] Xydeas, C., & Petrovic, V. (2000). Objective image fusion performance measure. *Electronics Letters*. 36(4);308–309. <https://doi.org/10.1049/el:20000267>.
- [43] Venkata Srikanth, M., Suneel Kumar, A., Nagasirisha, B., & Lakshmi, T. (2024). Brain MRI and CT image fusion using multiscale local extrema and image statistics. *ECTI Transactions on Electrical Engineering, Electronics, and Communications*. 22(1). <https://doi.org/10.37936/ecti-ec.2024221.249146>.
- [44] S. Amuthan, & N.C. Senthil Kumar. (2025). Emerging Trends in Deep Learning for Early Alzheimer’s Disease Diagnosis and Classification: A Comprehensive Review. *International Journal of Computational and Experimental Science and Engineering*, 11(1). <https://doi.org/10.22399/ijcesen.739>
- [45] Radhi, M., & Tahseen, I. (2024). An Enhancement for Wireless Body Area Network Using Adaptive Algorithms. *International Journal of Computational and Experimental Science and Engineering*, 10(3). <https://doi.org/10.22399/ijcesen.409>
- [46] Bandla Raghuramaiah, & Suresh Chittineni. (2025). BCDNet: An Enhanced Convolutional Neural Network in Breast Cancer Detection Using Mammogram Images. *International Journal of Computational and Experimental Science and Engineering*, 11(1). <https://doi.org/10.22399/ijcesen.811>
- [47] T. Deepa, & Ch. D. V Subba Rao. (2025). Brain Glial Cell Tumor Classification through Ensemble Deep Learning with APCGAN Augmentation. *International Journal of Computational and Experimental Science and Engineering*, 11(1). <https://doi.org/10.22399/ijcesen.803>
- [48] J Jeysudha, K. Deiwakumari, C.A. Arun, R. Pushpavalli, Ponnuragan Panneer Selvam, & S.D. Govardhan. (2024). Hybrid Computational Intelligence Models for Robust Pattern Recognition and Data Analysis. *International Journal of Computational and Experimental Science and Engineering*, 10(4). <https://doi.org/10.22399/ijcesen.624>
- [49] L. Smitha, Maddala Vijayalakshmi, Sunitha Tappari, N. Srinivas, G. Kalpana, & Shaik Abdul Nabi. (2024). Plant Disease Detection Using CNN with The Optimization Called Beluga Whale Optimization Mechanism. *International Journal of Computational and Experimental Science and Engineering*, 10(4). <https://doi.org/10.22399/ijcesen.697>
- [50] Boddupally JANAIHAH, & Suresh PABBOJU. (2024). HARGAN: Generative Adversarial Network Based Deep Learning Framework for Efficient Recognition of Human Actions from Surveillance Videos. *International Journal of Computational and Experimental Science and Engineering*, 10(4). <https://doi.org/10.22399/ijcesen.587>