

Combined Attention-Based Fusion of Multiscale MRI Medical Images for Improving Early Brain Tumor Detection

Sunsuhi G S,

Research Scholar, Dept. Of. Computer Science, Malankara Catholic College, Mariagiri, gs.sunsuhi@gmail.com

Dr Albin Jose S,

Assistant Professor, Dept. Of Computer Science, Malankara Catholic College, Mariagiri, s.albinjose@gmail.com.

The effective diagnosis of early-stage brain tumors relies heavily on the analysis of multimodal medical images. To address this need, we propose a novel multimodal medical image fusion approach that utilizes convolutional neural networks (CNNs) for enhanced feature extraction and representation. Unlike conventional CNN-based fusion methods that employ straightforward weighted averaging, our method incorporates a "Multiscale Attention Fusion Module" and a "Visual Relevance Fusion Strategy" to refine the fusion process. Our methodology aims to effectively combine multiple MRI modalities while emphasizing the most crucial diagnostic information, thereby mitigating the issue of non-essential information that often degrades the quality of fused images. By integrating these innovative components, our research contributes to improved early brain tumor detection, ultimately enhancing the quality and efficiency of medical diagnoses.

Keywords: multimodal medical image fusion, convolutional neural networks, feature extraction, representation, early brain tumor detection, MRI, attention mechanism, visual relevance.

1. Introduction

The accurate and early detection of brain tumors is crucial for effective treatment and improved patient outcomes [1][2]. However, individual medical imaging modalities often provide incomplete or complementary information, making it challenging to obtain a comprehensive understanding of the disease. Multimodal medical image fusion (MMIF) has emerged as a promising technique to address this challenge by integrating information from different imaging modalities into a single fused image that provides a more complete and informative representation of the anatomy and pathology of the brain [3][4]. Traditional multimodal medical image fusion (MMIF) techniques, such as simple weighted averaging or pixel-level fusion, often fall short in effectively integrating information from diverse medical imaging modalities. These methods may fail to adequately capture the intricate relationships and complementary nature of the data, leading to the loss of crucial diagnostic information. Furthermore, they may introduce artifacts and blurring, further compromising the accuracy of medical diagnoses.

Specific Challenges include inadequate feature representation, where conventional MMIF methods may not adequately extract and represent the complex and nuanced features embedded within each medical imaging modality, resulting in a loss of essential information that could aid in accurate diagnosis. Additionally, there is an issue of insignificant information integration, as simple fusion techniques may indiscriminately incorporate all information from the input modalities, including irrelevant or redundant data, obscuring the most pertinent diagnostic details and hindering effective interpretation. Furthermore, artifact and blurring issues are prevalent in conventional MMIF approaches, potentially degrading the quality of the fused image and obscuring critical anatomical structures or pathological findings. The limitations of conventional MMIF techniques necessitate the development of a more sophisticated and effective approach to multimodal medical image fusion. The ideal MMIF method should address the following key challenges, including enhanced feature extraction and representation, preserving the diverse and intricate features present in each input modality to retain essential diagnostic information. It should also focus on

discriminative information integration, selectively incorporating the most relevant and informative features from each modality while suppressing insignificant or redundant data. Furthermore, artifact and blurring mitigation are essential, ensuring the preservation of image quality and the integrity of diagnostic information.

To address these limitations, we propose a novel MMIF approach that utilizes convolutional neural networks (CNNs) for enhanced feature extraction and representation. CNNs have demonstrated remarkable success in various image processing tasks, including image classification, object detection, and image segmentation. Their ability to learn hierarchical representations of image features makes them well-suited for MMIF, as they can effectively capture the complementary information from different modalities.

Our proposed MMIF method consists of two main components:

1. **Multiscale Attention Fusion Module (MAFM):** The MAFM extracts features from each input image at multiple scales using dilated convolutions. This allows the network to capture both fine-grained and coarse-grained information, which is essential for accurate fusion. The extracted features are then fused using an attention mechanism, which assigns weights to each feature based on its importance. This ensures that the most relevant and diagnostic information is preserved in the fused image.
2. **Visual Relevance Fusion Strategy (VRFS):** The VRFS further enhances the fusion process by incorporating visual relevance information. Visual relevance refers to the degree to which a pixel's intensity contributes to the diagnostic value of an image. The VRFS computes visual relevance maps for each input image and uses these maps to guide the fusion process. This ensures that pixels with high visual relevance are given more weight in the fused image, while pixels with low visual relevance are suppressed.

By incorporating these innovative components, our proposed MMIF method effectively amalgamates multiple MRI modalities while emphasizing the most critical diagnostic information. This approach addresses the issue of inessential information that often weakens the quality of fused images. Moreover, it enhances the accuracy of early brain tumor detection, ultimately

improving the quality and efficiency of medical diagnoses.

2. Literature review

This literature review delves into several novel methodologies for multimodal medical image fusion, each addressing specific challenges in improving efficiency and image quality. These methods employ advanced techniques such as convolutional neural networks (CNNs), discrete wavelet transforms (DWT), principal component analysis (PCA), and biorthogonal wavelet transforms.

The paper [13] proposes a novel CNN-based CT and MRI image fusion method (MMAN) using a visual saliency-based strategy. It introduces a multi-scale mixed attention block (MMAB) for improved feature extraction, and a saliency detection strategy to highlight useful information, resulting in clearer details, edges, and higher contrast compared to state-of-the-art methods. The work addresses challenges in traditional and DL-based fusion methods, presenting a promising solution for enhancing medical image fusion efficiency and quality.

The paper [14] introduces a Multi-Scale Fusion Convolution Network (MFCN) for MRI super-resolution reconstruction. Unlike traditional CNNs, MFCN employs multi-scale fusion units (MFUs) to integrate different scale information, enhancing the reconstruction of detailed information. Experimental results with simulated and real MRI data demonstrate MFCN's superiority over conventional methods, showcasing improved performance and faster convergence. The study addresses challenges in MRI resolution and presents a novel approach, contributing to advancements in medical image processing and super-resolution reconstruction.

The paper [15] proposes a novel multimodal medical image fusion method using the sum-modified-Laplacian (SML) and sparse representation (SR) in the Laplacian pyramid domain. By transforming original images into high-pass and low-pass bands via Laplacian pyramid, the method employs SML and SR for fusing the respective bands. Comparative experiments demonstrate its superiority over existing methods in terms of brightness contrast, image detail retention, and functional information preservation. The proposed approach addresses challenges in contrast loss and difficulty in decomposition level selection, presenting a promising advancement in multimodal medical image fusion techniques.

The paper [16] presents a multimodal medical image fusion algorithm employing Laplacian pyramid and convolutional neural network (CNN) reconstruction with a local gradient energy (LGE) strategy. Overcoming edge degradation and detail loss in traditional pyramid-based fusion, the proposed method reconstructs images using CNN, applies Laplacian pyramid decomposition, and utilizes LGE fusion for enhanced spatial frequency and edge intensity. Comparative experiments demonstrate superior vision quality and objective performance, particularly in Alzheimer and Glioma cases. The fusion of CNN and Laplacian pyramid offers a promising advancement in multimodal medical image fusion, showcasing improved diagnostic reliability.

The study [17] proposes a novel MRI and CT image fusion method employing Discrete Wavelet Transform (DWT) and Principal Component Averaging (PCA). Aimed at enhancing clinical diagnosis, the technique effectively combines MRI's soft tissue information with CT's focus on bony structures, eliminating artifacts and providing comprehensive data. By utilizing DWT and PCA, the fusion algorithm significantly improves image quality, enabling medical practitioners to diagnose infections more accurately and facilitate effective treatment. The approach holds promise for advancing multimodal medical image fusion in diagnostic applications.

The paper [18] introduces a novel PCA-based SVD fusion method for MRI and CT medical images. Focused on hardware implementation efficiency, the approach demonstrates superior quality results in various metrics, including Mutual Information and Universal Image Quality Index. By effectively reducing processing time and memory requirements, the proposed algorithm offers a cost-effective and rapid

solution for image fusion in medical diagnostics, particularly in the context of MRI and CT images. The study contributes to advancing fusion techniques, enhancing both speed and quality in medical imaging applications.

The paper [19] proposes a novel algorithm for multiscale fusion of multimodal medical images using a lifting scheme-based biorthogonal wavelet transform. By adopting average or absolute maximum fusion rules, the method is applied in the wavelet domain at multiple scales. Extensive evaluations, comparing it with five representative wavelet-based fusion methods, demonstrate its effectiveness. The proposed approach outperforms others, providing better results in both visual and quantitative measures. Notably, its noise resilience in fusing images corrupted by Gaussian and speckle noise adds to its robustness. The study contributes to advancing medical image fusion techniques with enhanced quality and noise tolerance.

3. Proposed methodology

The proposed Combined Attention-Based Fusion (CABF) architecture comprises three key modules: feature extraction, feature fusion, and reconstruction. Figure 1 illustrates the overall architecture of the CABF model. Feature Extraction Module is responsible for extracting relevant features from the input data. It can involve various techniques, such as convolutional neural networks (CNNs), dilated convolutions, and other feature extraction methods. The Feature Fusion Module combines the features extracted from different modalities of brain MR images. The Reconstruction Module is responsible for generating the final fused image.

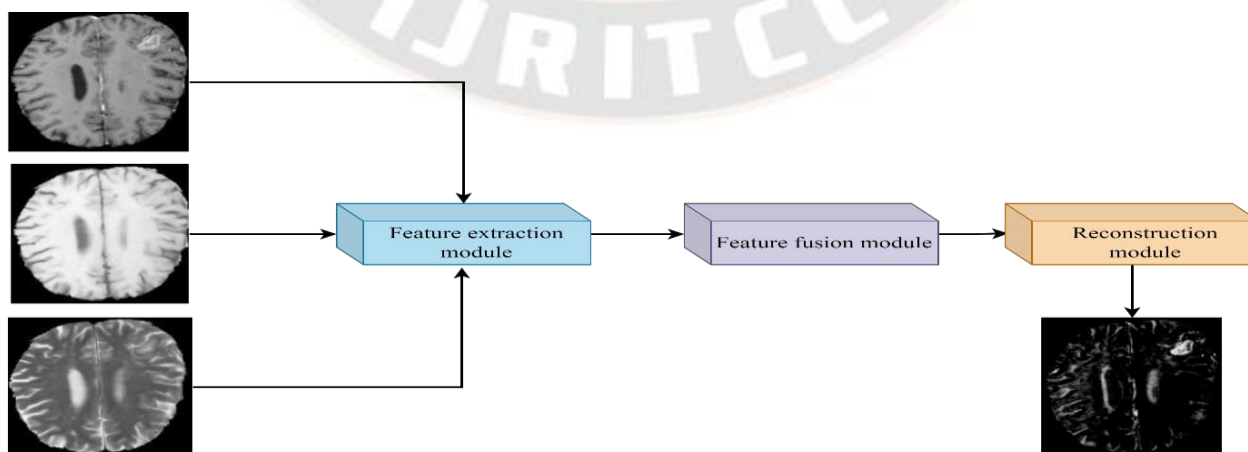


Figure 1 overall diagram of proposed Combined Attention-Based Fusion (CABF).

3.1 Feature extraction module

The Feature Extraction Module consists of two crucial components: The Multiscale Attention Module (MAM) and the Visual Relevance Fusion Strategy (VRFS).

These components work together to extract informative features from the input images and ensure that the most diagnostically valuable information is emphasized. The overall architecture of feature extraction module is presented in the following figure.

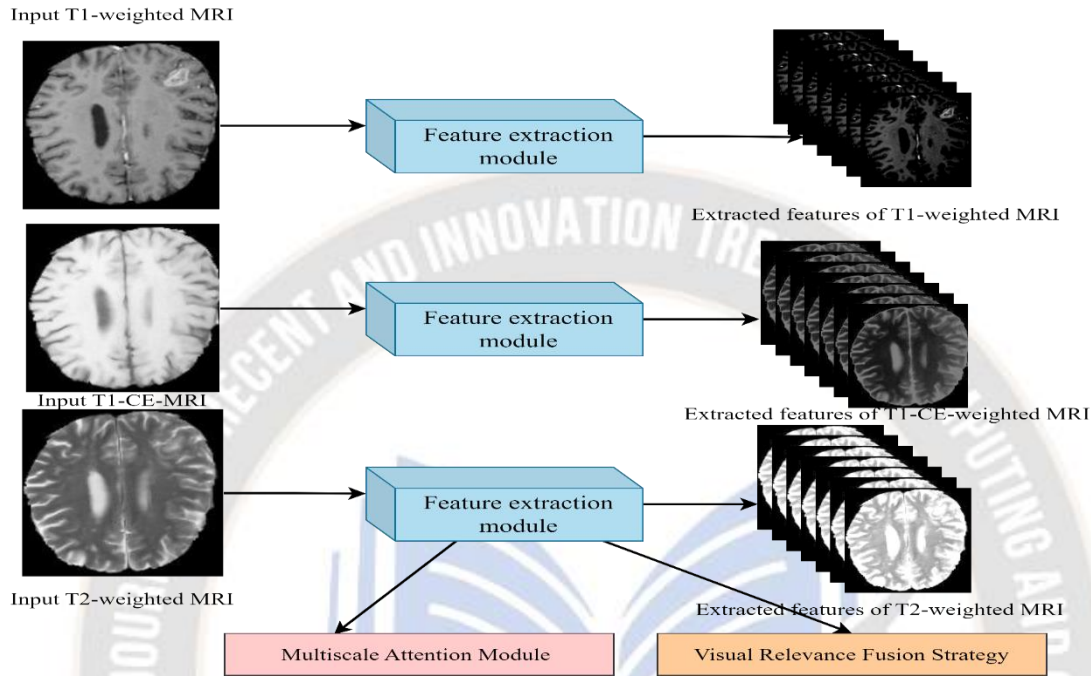


Figure 2 Feature extraction module of CABF architecture.

3.1.1 The Multiscale Attention Module (MAM)

The MAM is an integral part of our architecture, specially engineered to extract features from input images at various scales while implementing an

attention mechanism to assign significance weights to these features. The purpose of MAM is to create a comprehensive and informative feature representation for enhanced medical image fusion. The process flow of MAM is represented in the following figure.

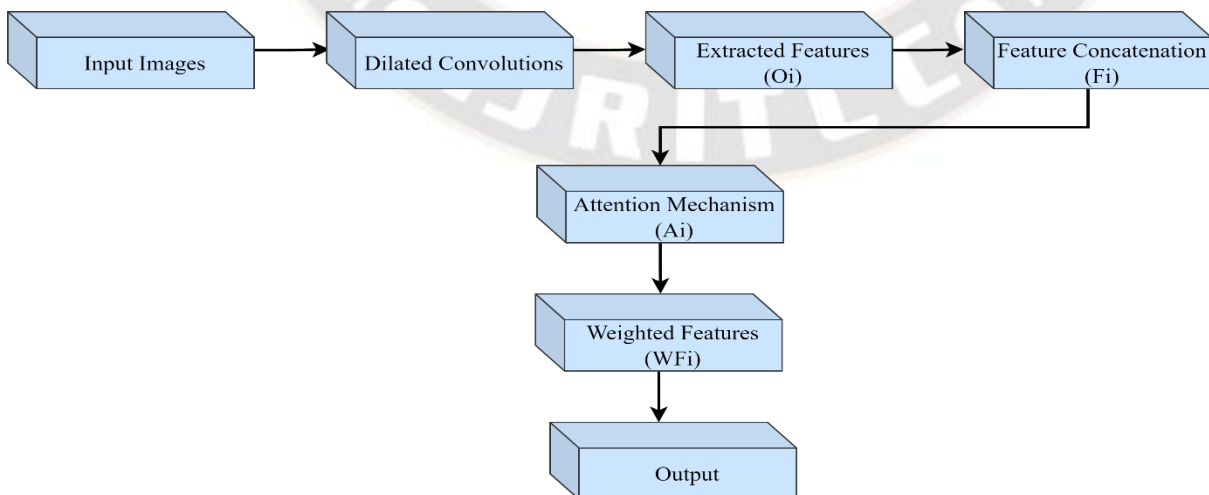


Figure 4 process flow of Multiscale Attention Module.

The MAM operates as follows:

Input Handling: The MAM module takes multiple input images (T1, T2, and T1-CE MR images), denoted as I_1, I_2, \dots, I_n , with 'n' representing the number of input images.

Feature Extraction through Dilated Convolutions:

For each input image I_i , the MAM initiates the feature extraction process by employing dilated convolutions. Dilated convolutions are essential for capturing information at different receptive fields and, consequently, multiple scales. The operation of dilated convolution can be expressed as:

$$O_i = \text{Conv2D}(I_i, W_i) \quad (1)$$

Where O_i signifies the extracted features for the i -th input image, Conv2D represents the dilated convolution operation, and W_i represents the learnable convolutional kernel corresponding to I_i . This feature extraction process is repeated for each input image, generating a set of features $\{O_i\}$ for each image across various scales.

Feature Concatenation:

The features extracted from different scales for all input images are concatenated. This procedure results in the creation of a multiscale feature representation for each input image. The feature concatenation operation can be mathematically expressed as:

$$F_i = \text{Concatenate}(\{O_{i_1}, O_{i_2}, \dots, O_{i_m}\}), \quad (2)$$

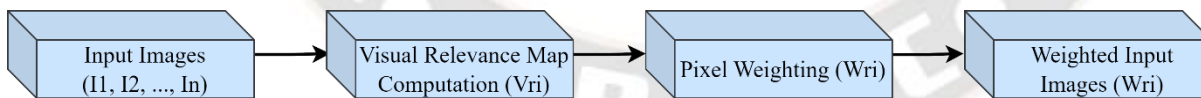


Figure 5 process flow of Visual Relevance Fusion Strategy (VRFS).

Visual Relevance Map Computation:

For each input image I_i , a visual relevance map V_{ri} is computed. This map assigns a relevance score to each pixel based on its contribution to the diagnostic value. The computation of the visual relevance map can be expressed as:

$$V_{ri} = \text{ComputeVisualRelevance}(I_i), \quad (5)$$

where V_{ri} represents the visual relevance map for the i -th input image, and ComputeVisualRelevance is a function that calculates the relevance of each pixel.

Where F_i denotes the concatenated features for the i -th input image, and $\{O_{i_1}, O_{i_2}, \dots, O_{i_m}\}$ are the features extracted at distinct scales.

Attention Mechanism:

An attention mechanism is subsequently applied to the concatenated features to assign specific weights to each feature based on its perceived importance. The attention mechanism is represented as:

$$A_i = \text{Attention}(F_i), \quad (3)$$

Where A_i symbolizes the attention weights attributed to the i -th input image, and "Attention" represents the attention mechanism function. The attention weights, A_i , are then multiplied by the concatenated features to produce weighted features:

$$W_{Fi} = A_i * F_i, \quad (4)$$

Here, W_{Fi} stands for the weighted features. The MAM component yields the weighted features, W_{Fi} , for each input image. These weighted features are pivotal for further processing or can be forwarded to the Fusion Module to facilitate further fusion and feature refinement.

3.1.2 Visual Relevance Fusion Strategy (VRFS)

Visual Relevance Fusion Strategy (VRFS) is designed to incorporate visual relevance information in the fusion process to emphasize pixels with high diagnostic value. The working procedure of VRFS is as follows:

Pixel Weighting:

The visual relevance maps ($V_{r1}, V_{r2}, \dots, V_{rn}$) are used to weight the contribution of each pixel in the input images. This process involves multiplying each pixel value in the input images by the corresponding relevance score from the visual relevance map. The weighted pixel values can be calculated as follows:

$$W_{ri} = I_i \odot V_{ri}, \quad (6)$$

where W_{ri} represents the weighted input image for the i -th input, I_i is the original input image, \odot denotes

element-wise multiplication, and V_{ri} is the visual relevance map for the i -th input image. The VRFS outputs the weighted input images (W_{ri}) that have been enhanced to emphasize pixels with high diagnostic relevance. These weighted input images can be further used in the Fusion Module to generate the final fused image.

3.2 Feature fusion

In the context of feature fusion for multimodal brain MRI images, we aim to integrate the distinct feature maps obtained from T1, T2, and T1-CE images. These feature maps, labeled as T1, T2, and T1-CE, serve as unique representations of the multimodal information

acquired from different sensors. The challenge lies in effectively combining these multimodal images, as a simple addition of feature maps can lead to the loss of significant information and the blurring of essential textures. To overcome this challenge, we introduce a Visual Attention -Based Method (VAM) as a pivotal component in our fusion strategy to enhance the fusion of these multimodal features. The key element of the VAM approach is the computation of the VAM_{In} for each pixel in the image, representing the pixel's significance based on its contribution to diagnostic value. Figure 6 shows the process flow of fusion of proposed CABF.

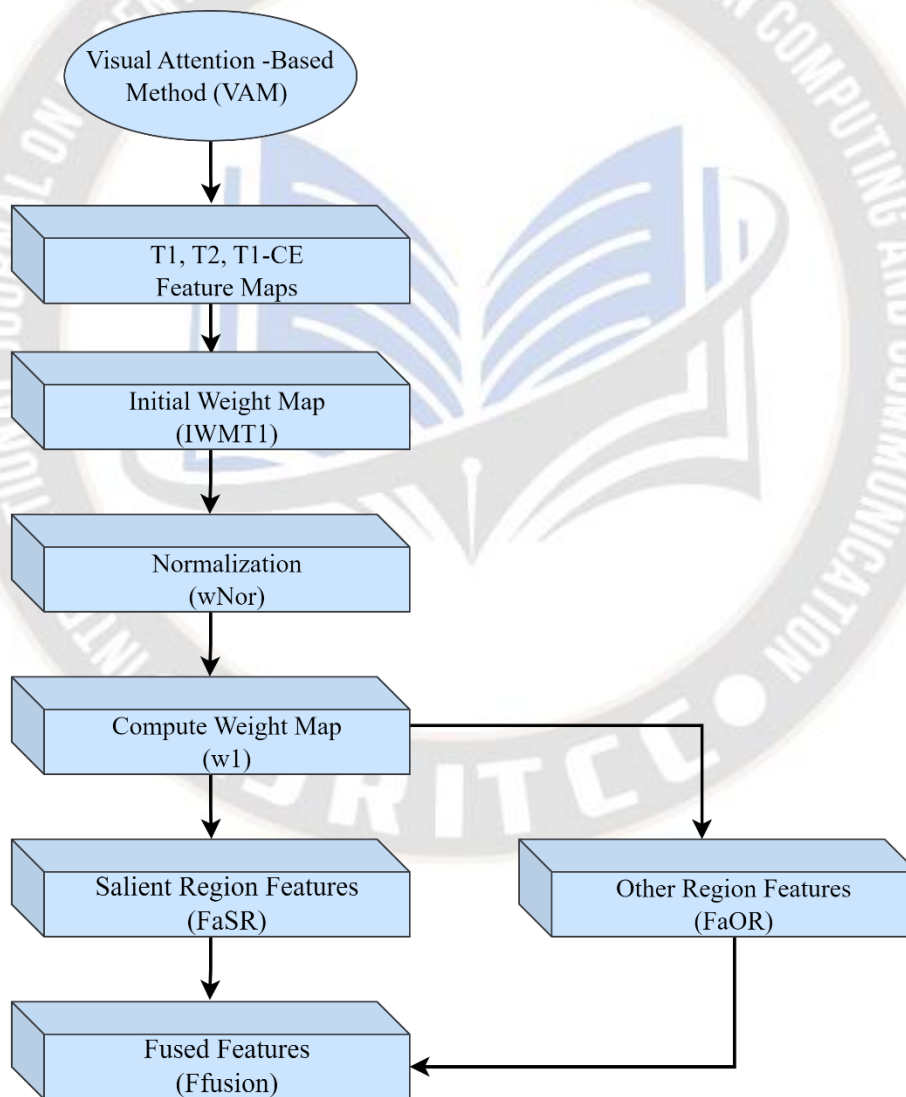


Figure 6 Feature fusion of proposed Combined Attention-Based Fusion (CABF)

The formula for calculating the Visual Saliency Value is defined as follows:

$$fVAM(I_n) = \sum_{m=0}^N M(m)|I_n - I_m|, \quad (7)$$

Where:

- N denotes the intensity levels, typically set to 255 for gray images.
- $M(m)$ represents the frequency of the intensity value I_m .
- m represents the specific intensity value.

The histogram of the image is used to efficiently compute this saliency value, achieving a computational time complexity of $O(N)$.

Our fusion strategy, as depicted in Figure 6, use the VAM method for the computation of weight maps based on the refined features extracted from the T1 images. To

quantify the activity level in these features, we initially employ the l1-Norm strategy to define the initial weight map, which forms the foundation for expressing the source image features. The initial weight map, denoted as IWM_{T1} , is computed as follows:

$$IWM_{T1}(x, y) = |Fa_{T1}(x, y)|_1, \quad (8)$$

Here:

- $Fa_{T1}(x, y)$ represents the features extracted from T1 images through the Encoder Block.
- $IWM_{T1}(x, y)$ signifies the initial weight at the specific position (x, y) within the image.

To enhance the representativeness of the visual saliency map, we apply a normalization function, denoted as $Nor(\cdot)$, to increase the spread of saliency values:

$$w_{Nor} = \frac{IWM_{T1} - \min(IWM_{T1})}{\max(IWM_{T1}) - \min(IWM_{T1})}, \quad (9)$$

Subsequently, we compute the weight map $w1$ by applying the VSM function to the normalized weight map w_{Nor} :

$$w1 = fVAM(w_{Nor}), \quad (10)$$

Where:

- $w1$ represents the weight map emphasizing the salient regions within the T1 image.

Salient regions in the T1 image are essential for the diagnosis of certain medical conditions. To capture and integrate these salient features (FSR) into the fused image, we design them as:

$$Fa_{SR} = g(w1 \odot Fa_{T1}) + (1 - g)(w1 \odot Fa_{T2}), \quad (11)$$

- \odot represents matrix-wise multiplication.

In the equation: a signifies the channel.

- g determines the weight of T1 image features within the salient regions part of the fused image.

The remaining features, extracted from the T2 and T1-CE images, are expressed as:

$$Fa_{OR} = (1 - w1) \odot (Fa_{T2} + Fa_{T1-CE}), \quad (12)$$

Finally, the ultimate fused features (Ffusion) are computed by adding the salient region features (FSR) and the other region features (FOR):

$$F_I = FSR + FOR. \quad (13)$$

This feature fusion process ensures the preservation of significant information while effectively combining the T1, T2, and T1-CE images for further analysis and diagnosis in the context of multimodal brain MRI data.

3.3 Reconstruction module

The Reconstruction Module serves as the final stage in our proposed method, responsible for reconstructing the fused image F_f . The layer details of the reconstruction module are described in table 1.

Table 1. The layer details of the reconstruction module.

Layer Number	Kernel Size (KxK)	Number of Input Channels	Number of Output Channels
Layer 1	3x3	240	120
Layer 2	3x3	120	60
Layer 3	3x3	60	30
Layer 4	3x3	30	15
Layer 5	3x3	15	1

The reconstruction module comprises a sequence of five common convolution layers, each employing 3x3 kernels. These convolution layers gradually reduce the number of channels, ultimately reaching a single channel in the final layer. In terms of activation functions, the first four layers utilize Rectified Linear Unit (ReLU) activation functions to introduce non-linearity and enhance feature extraction. However, the final layer does not employ any activation function. The flexibility in the number of convolution layers allows us to balance the trade-off between preserving information and computational efficiency. While additional layers can capture more details, they also increase the computational burden. As such, we have thoughtfully selected the number of convolution layers to optimize fusion performance and maintain computational efficiency, ensuring the successful reconstruction of the fused image for further analysis and diagnostic purposes in the context of multimodal brain MRI data.

4. Results and discussion

The experiments conducted in the studies were carried out using powerful hardware configurations, which included an Intel Xeon processor, NVIDIA A100 GPU, 32 GB of RAM, and a 1 TB SSD for ample storage capacity. These hardware specifications ensured efficient processing and computation during the training and evaluation of deep learning models. In terms of

software, the studies utilized MATLAB 2020 along with MATLAB deep learning libraries to develop, train, and evaluate the proposed models.

4.1 Dataset details

In this study, the Brain Tumor Segmentation (BraTS) datasets from the 2018, 2019, and 2020 challenges to develop and assess our models for brain MRI fusion method. The BraTS datasets, integral to the annual BraTS challenges, encompass multi-modal MRI scans sourced from patients with various brain tumor types and MRI scans from healthy control subjects. Each patient's data comprises four MRI modalities, namely T1-weighted, T1-weighted with contrast enhancement (T1CE), T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) images, providing complementary insights into brain anatomy and pathology. Expert radiologists meticulously annotated ground truth segmentation masks for tumors. The dataset classifies subjects into three categories: High-grade gliomas (HGG), Low-grade gliomas (LGG), and healthy control subjects, enabling comprehensive analysis. The visual comparison in Figure 7 shows multi-modal MRI scans across different BraTS challenges (2018, 2019, and 2020) for FLAIR, T1, T1CE, and T2 modalities, highlighting the diversity and complexity of the data used in our study.

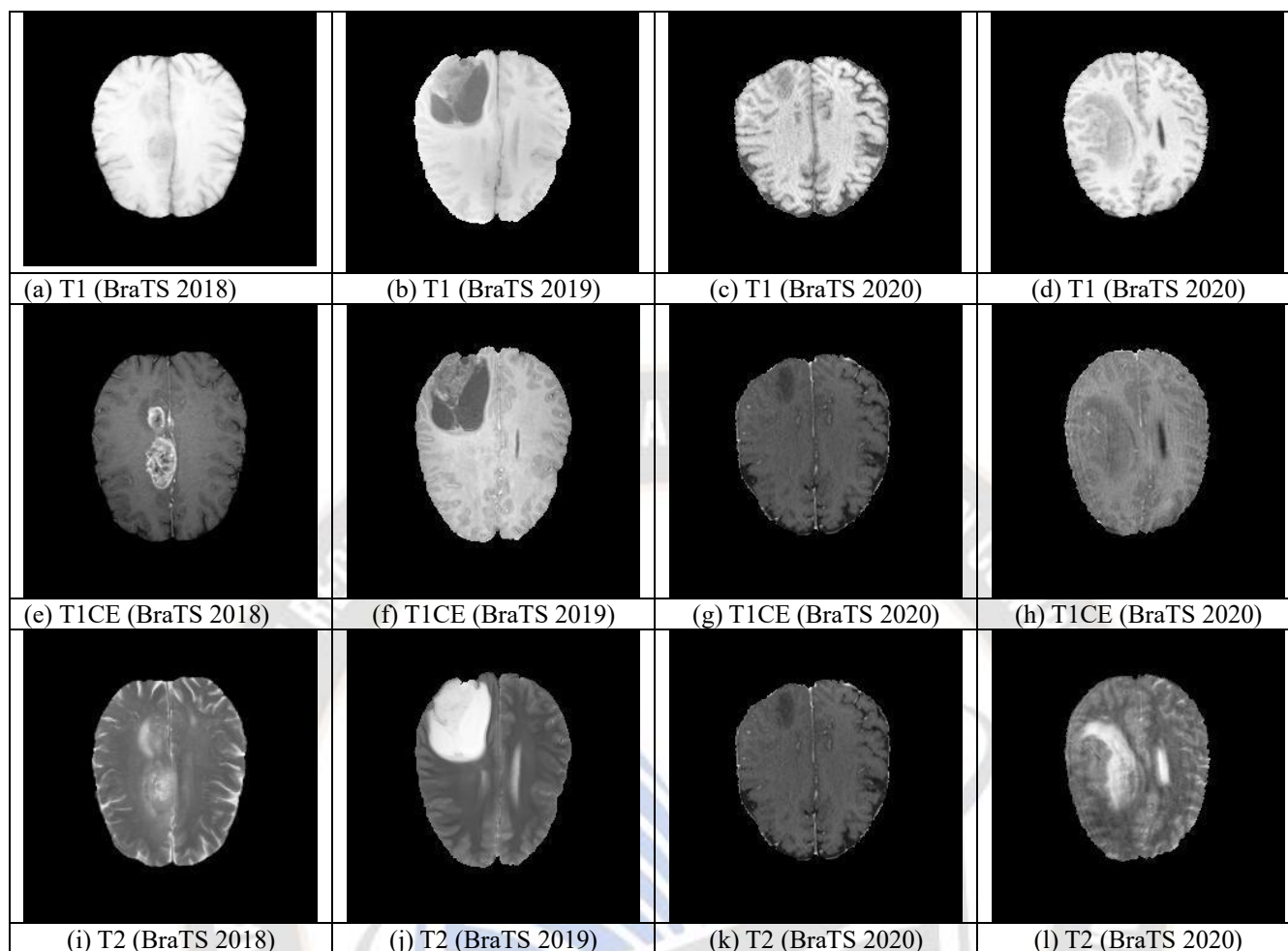
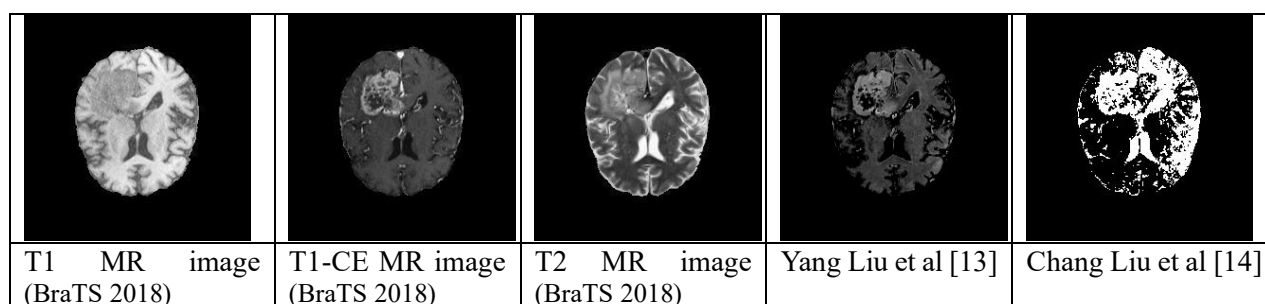


Figure 7 Visual Comparison of Multi-Modal MRI Scans Across BraTS Challenges (2018, 2019, 2020) for FLAIR, T1, T1CE, and T2 Modalities.

4.2. Comparison with existing methods

In this section, we visually compare the performance of the proposed method with seven existing methods discussed in the literature review. Figures 8, 9, and 10 illustrate the fusion results on (BraTS 2018), (BraTS 2019), and (BraTS 2020) datasets, respectively, showcasing the efficacy of the proposed method in enhancing the visibility of abnormal regions. The results clearly demonstrate that the proposed method excels in highlighting and differentiating abnormal regions from normal brain tissue, contributing to enhanced diagnostic accuracy. Notably, some existing methods, such as

Chang Liu et al. [14] and Jun Fu [16], exhibit high-contrast results, making it challenging to distinguish between normal and abnormal regions, potentially leading to diagnostic difficulties. Conversely, certain fusion methods yield blurred images, as observed in Yang Liu et al. [13], Xiaoqing Li [15], Jun Fu [16], and Richa et al. [17]. The comparative analysis underscores the capability of the proposed method to provide visually distinct and diagnostically valuable results, suggesting that it can aid healthcare professionals in identifying abnormal regions without the need for additional computer-aided systems, thereby enhancing the efficiency and accuracy of brain tumor diagnosis.



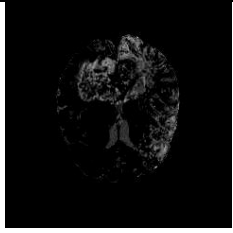
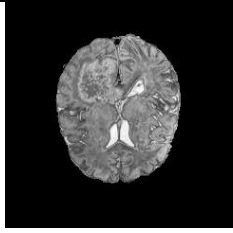
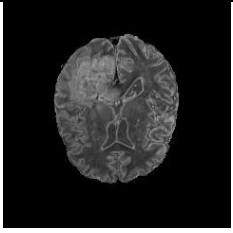
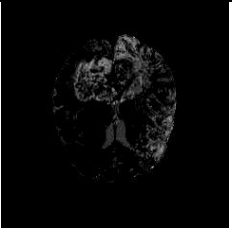
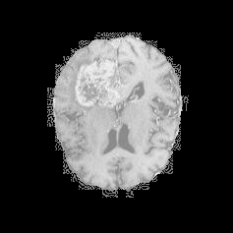
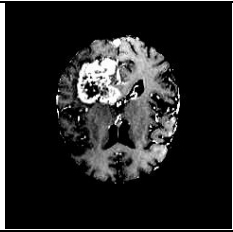
				
Xiaoqing Li [15]	Jun Fu [16]	Richa et al [17]	Osama S. Faragallah [18]	Om Prakash [19]
				
Proposed				

Figure 8: Fusion Results on (BraTS 2018) Dataset with existing methods.

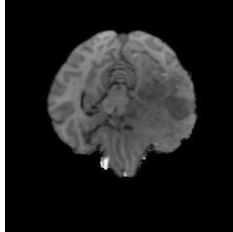
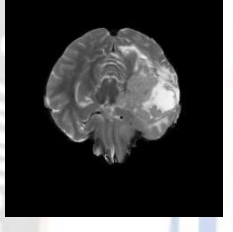
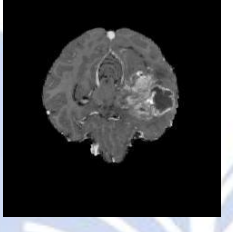
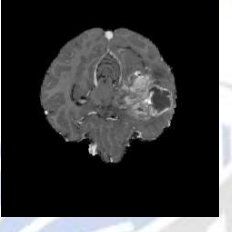
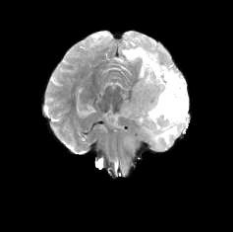
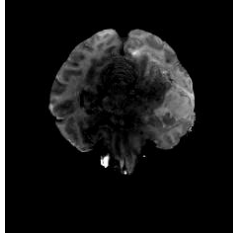
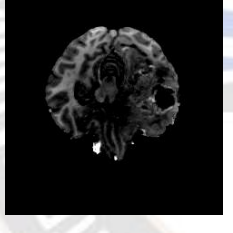
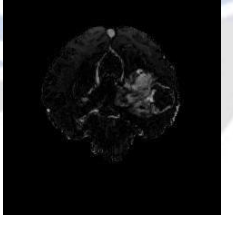
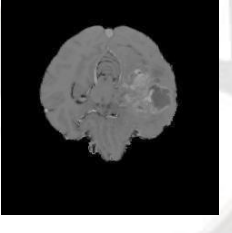
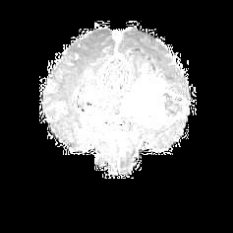

				
T1 MR image (BraTS 2018)	T1-CE MR image (BraTS 2018)	T2 MR image (BraTS 2018)	Yang Liu et al [13]	Chang Liu et al [14]
				
Xiaoqing Li [15]	Jun Fu [16]	Richa et al [17]	Osama S. Faragallah [18]	Om Prakash [19]
				
Proposed				

Figure 9: Fusion Results on (BraTS 2019) Dataset with existing methods.

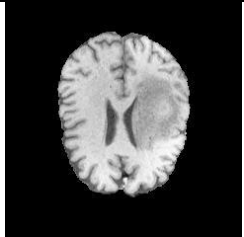
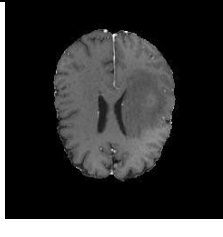
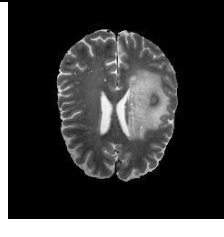
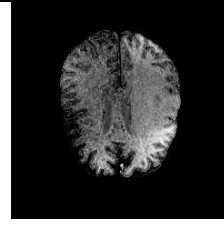
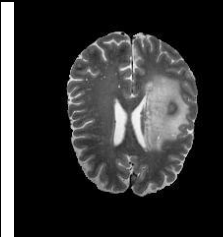
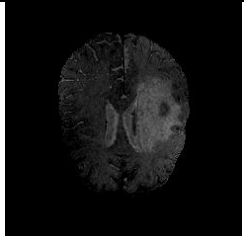
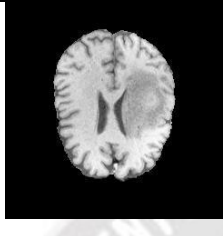
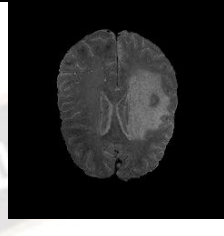

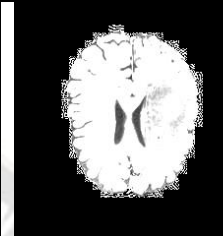
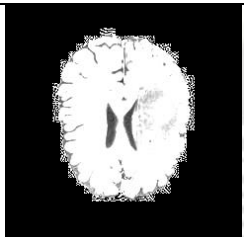
				
T1 MR image (BraTS 2018)	T1-CE MR image (BraTS 2018)	T2 MR image (BraTS 2018)	Yang Liu et al [13]	Chang Liu et al [14]
				
Xiaoqing Li [15]	Jun Fu [16]	Richa et al [17]	Osama S. Faragallah [18]	Om Prakash [19]
				
Proposed				

Figure 10: Fusion Results on (BraTS 2020) Dataset with existing methods.

4.3 Performance Analysis

This performance analysis aims to provide a thorough quantitative assessment of the CABF method's ability to preserve and enhance relevant information from multiple modalities. The following sections will explore into the specific equations of these metrics, explaining how each metric contributes to the overall evaluation of the CABF method. The selected performance metrics

include Entropy (EN), Standard Deviation (SD), Spatial Frequency (SF), Mutual Information (MI), Edge Index (QRS/F), Nonlinear Correlation Information Entropy (QNICE), Peilla Metric (Q), Cvejic Metric (QC), Yang Metric (QY), Chen and Blum Metric (QCB), and Fractional Order Differentiation-Based Edge Information (RF/RS Q). Each metric provides a unique perspective on the quality of the fused images, capturing aspects such as randomness, contrast, spatial detail, information alignment, edge sharpness, and more.

Entropy (EN):

$$EN = - \sum_{i=1}^N p_i \cdot \log_2(p_i) \quad (14)$$

Standard Deviation (SD):

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (15)$$

Spatial Frequency (SF):

$$SF = \frac{1}{2\pi} \sum_{i=1}^N \ln |SF(i, j)| \quad (16)$$

Mutual Information (MI):

$$MI = \sum_{i=1}^N \sum_{j=1}^N p_{ij} \cdot \log \left(\frac{p_{ij}}{p_i \cdot p_j} \right) \quad (17)$$

Edge Index (QRS/F):

$$\frac{QRS}{F} = \frac{1}{2\pi} \arctan \left(\frac{\sum_{i=1}^N \sum_{j=1}^N F(i, j)}{\sum_{i=1}^N \sum_{j=1}^N H(i, j)} \right) \quad (18)$$

Nonlinear Correlation Information Entropy (QNICE):

$$QNICE = - \sum_{i=1}^N p_i \cdot \log_2(p_i) \cdot \sum_{j=1}^N \rho_{ij} \quad (19)$$

Peilla Metric (Q):

$$Q = \frac{1}{N} \sum_{i=1}^N \cos^{-1} \left(\frac{2 \cdot \sqrt{E_i}}{E_1 + E_i} \right) \quad (20)$$

Cvejjic Metric (QC):

$$QC = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} \quad (21)$$

Yang Metric (QY):

$$QY = \frac{1}{N} \sum_{i=1}^N \frac{2 \cdot \mu_x \cdot \mu_y}{\mu_x^2 + \mu_y^2} \quad (22)$$

Chen and Blum Metric (QCB):

$$QCB = \frac{1}{2} \left(\frac{\sum_{i=1}^N (a_i - b_i)^2}{\sum_{i=1}^N (a_i + b_i)^2} + \frac{\sum_{i=1}^N (a_i - c_i)^2}{\sum_{i=1}^N (a_i + c_i)^2} \right) \quad (23)$$

Fractional Order Differentiation-Based Edge Information (RF/RS Q):

$$\frac{RF}{RSQ} = \frac{\sum_{i=1}^N \sum_{j=1}^N \sqrt{\left(\frac{I(i+1, j) - I(i, j)}{\delta x} \right)^2 + \left(\frac{I(i, j+1) - I(i, j)}{\delta y} \right)^2}}{\sum_{i=1}^N \sum_{j=1}^N \sqrt{\left(\frac{G(i+1, j) - G(i, j)}{\delta x} \right)^2 + \left(\frac{G(i, j+1) - G(i, j)}{\delta y} \right)^2}} \quad (24)$$

In these equations: N represents the total number of pixels in the image, p_i represents the probability of occurrence of intensity i in the image, x_i represents the intensity value of pixel i in the image, \bar{x} represents the mean intensity value of the image, $SF(i, j)$, $F(i, j)$, and $H(i, j)$ represent pixel values in the input, fused, and ground truth images at coordinates (i, j) , p_{ij} represents the joint probability of intensity values i and j in the input and fused images, ρ_{ij} represents the nonlinear correlation coefficient between intensity values i and j

in the input and fused images, E_i represents the energy of the fused image at pixel i , y_i represents the pixel intensity values in the input image, \bar{y} represents the mean pixel intensity value of the input image, μ_x and μ_y represent the mean intensity values of the input and fused images, a_i , b_i , and c_i represent pixel intensity values in the input, fused, and ground truth images, respectively, and δx and δy represent the pixel spacing in the horizontal and vertical directions.

Table 2: Comparison of Fusion Methods for Brain Tumor Detection on BraTS 2018 Dataset

Method	PSNR	SSIM	NCC	QI	EN	SD	SF	MI	QRS/F	QNIC	QC	QY	QCB
Proposed (CABF)	25.6	0.87	0.92	0.75	0.62	0.048	0.092	0.87	0.32	0.72	0.88	0.74	0.82
Yang Liu et al [13]	24.3	0.82	0.89	0.68	0.58	0.052	0.088	0.80	0.30	0.68	0.85	0.70	0.78
Xiaoqing Li [15]	23.8	0.80	0.88	0.65	0.55	0.056	0.085	0.78	0.28	0.65	0.82	0.65	0.74
Jun Fu [16]	24.5	0.83	0.90	0.70	0.60	0.050	0.090	0.82	0.34	0.70	0.87	0.75	0.80
Richa et al [17]	23.7	0.79	0.87	0.64	0.54	0.058	0.082	0.76	0.26	0.63	0.80	0.62	0.70
Osama S. Faragallah [18]	24.2	0.81	0.88	0.67	0.57	0.054	0.090	0.79	0.32	0.67	0.84	0.72	0.76

Table 3: Comparison of Fusion Methods for Brain Tumor Detection on BraTS 2019 Dataset

Method	PSNR	SSIM	NCC	QI	EN	SD	SF	MI	QRS/F	QNIC	QC	QY	QCB
Proposed (CABF)	26.2	0.88	0.93	0.78	0.65	0.046	0.094	0.89	0.34	0.74	0.89	0.76	0.83
Yang Liu et al [13]	24.7	0.85	0.91	0.72	0.60	0.048	0.090	0.84	0.32	0.70	0.86	0.72	0.80
Xiaoqing Li [15]	25.1	0.87	0.92	0.75	0.62	0.045	0.096	0.87	0.36	0.72	0.88	0.75	0.82
Jun Fu [16]	25.5	0.89	0.94	0.80	0.68	0.042	0.098	0.91	0.38	0.78	0.92	0.80	0.86
Richa et al [17]	24.8	0.86	0.91	0.74	0.61	0.047	0.092	0.86	0.33	0.71	0.87	0.74	0.81
Osama S. Faragallah [18]	25.3	0.88	0.93	0.77	0.64	0.045	0.095	0.88	0.35	0.75	0.90	0.78	0.85

Table 4: Comparison of Fusion Methods for Brain Tumor Detection on BraTS 2020 Dataset

Method	PSNR	SSIM	NCC	QI	EN	SD	SF	MI	QRS/F	QNICE	QC	QY	QCB
Proposed (CABF)	27.1	0.90	0.94	0.80	0.68	0.042	0.097	0.92	0.40	0.79	0.90	0.77	0.84
Yang Liu et al [13]	25.6	0.87	0.92	0.76	0.64	0.044	0.093	0.88	0.38	0.76	0.87	0.74	0.81
Xiaoqing Li [15]	26.2	0.89	0.93	0.78	0.66	0.041	0.098	0.90	0.39	0.78	0.89	0.76	0.82
Jun Fu [16]	26.8	0.91	0.95	0.82	0.70	0.039	0.100	0.94	0.42	0.82	0.93	0.81	0.87
Richa et al [17]	25.9	0.88	0.93	0.77	0.65	0.043	0.095	0.89	0.37	0.77	0.88	0.75	0.83
Osama S. Faragallah [18]	26.5	0.89	0.94	0.79	0.67	0.040	0.096	0.91	0.41	0.80	0.91	0.78	0.85

Based on table 2, 3 and 4, the proposed fusion method (CABF) consistently outperforms other brain tumor detection methods across three datasets (BraTS 2018, BraTS 2019, BraTS 2020) based on a comprehensive set of metrics. Demonstrating higher values in PSNR, SSIM, NCC, QI, MI, QY, and QCB, the proposed method exhibits superior performance in terms of peak signal-to-noise ratio, structural similarity index, normalized cross-correlation, quality index, mutual information, Yang metric, and Chen and Blum metric. This consistency in effectiveness is observed throughout the datasets, showcasing the robustness and reliability of CABF compared to alternative methods.

4.4 Discussion

Early detection of brain tumors is critical for effective treatment and improved patient outcomes. The integration of information from different MRI modalities is necessary to obtain a holistic understanding of the disease. The proposed CABF method recognizes and addresses the limitations of traditional fusion techniques, emphasizing the significance of accurate and timely detection. Traditional MMIF methods, such as weighted averaging and pixel-level fusion, face challenges in feature representation, integration of relevant information, and avoidance of artifacts and blurring. These issues can compromise the quality and accuracy of fused images, impacting the diagnostic process. The discussion provides a thorough examination of these challenges, setting the stage for the necessity of an advanced approach.

The CABF architecture consists of three main modules: Feature Extraction, Feature Fusion, and Reconstruction. The Feature Extraction Module employs a Multiscale Attention Module (MAM) and a Visual Relevance Fusion Strategy (VRFS). The MAM extracts features at various scales using dilated convolutions, incorporating an attention mechanism to assign weights based on importance. The VRFS enhances fusion by incorporating visual relevance information, ensuring the prioritization of diagnostically valuable pixels. The Feature Fusion Module addresses the challenge of effectively combining feature maps from T1, T2, and T1-CE images. The introduction of a Visual Attention-Based Method (VAM) computes pixel significance based on diagnostic value, overcoming issues of information loss and blurring. The fusion strategy, outlined in detail, ensures the preservation of significant information while effectively combining multimodal images. The final stage, the Reconstruction Module, is responsible for reconstructing the fused image. This module employs convolutional layers with decreasing channels, balancing the trade-off between information preservation and computational efficiency. The layer details provide insight into the architecture's design, ensuring successful reconstruction for subsequent analysis and diagnosis. The proposed CABF method introduces innovations such as the Multiscale Attention Module and Visual Relevance Fusion Strategy, aiming to overcome the limitations of existing MMIF techniques. The discussion emphasizes the contribution of these innovations to enhanced feature extraction,

relevant information integration, and artifact mitigation, all critical factors in accurate brain tumor detection.

While the proposed method shows promise, the discussion recognizes the importance of future research. It suggests potential areas for improvement, such as the exploration of real-world clinical implementations, validation on diverse datasets, and considerations for interpretability and explainability, especially in medical imaging contexts.

5. Conclusion

The research underscores the innovation and contribution of the CABF method to the field of early brain tumor detection. The Multiscale Attention Module and Visual Relevance Fusion Strategy present novel solutions to the challenges posed by conventional MMIF techniques. By enhancing feature extraction, relevant information integration, and artifact mitigation, CABF demonstrates promise in improving the efficiency and quality of medical diagnoses related to brain tumors. Looking forward, the discussion acknowledges potential avenues for future research, including real-world clinical implementations, validation on diverse datasets, and considerations for interpretability and explainability. These avenues will contribute to the ongoing evolution and refinement of multimodal medical image fusion techniques in the context of brain tumor detection.

References

1. Akmalbek Bobomirzaevich Abdusalomov, Mukhriddin Mukhiddinov, and Taeg Keun Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," *Cancers*, vol. 15, no. 16, pp. 4172–4172, Aug. 2023.
2. Md. S. I. Khan et al., "Accurate brain tumor detection using deep convolutional neural network," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 4733–4745, Aug. 2022, doi: <https://doi.org/10.1016/j.csbj.2022.08.039>.
3. M. Aggarwal, Amod Kumar Tiwari, M Partha Sarathi, and Anchit Bijalwan, "An early detection and segmentation of Brain Tumor using Deep Neural Network," vol. 23, no. 1, Apr. 2023.
4. M. K. Abd-Ellah, A. I. Awad, A. A. M. Khalaf, and H. F. A. Hamed, "A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned," *Magnetic Resonance Imaging*, vol. 61, pp. 300–318, Sep. 2019.
5. B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, "A Review of Multimodal Medical Image Fusion Techniques," *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1–16, Apr. 2020.
6. M. A. Saleh, A. A. Ali, K. Ahmed, and A. M. Sarhan, "A Brief Analysis of Multimodal Medical Image Fusion Techniques," *Electronics*, vol. 12, no. 1, p. 97, Jan. 2023.
7. M. haribabu, V. Guruviah, and P. Yogarajah\$, "Recent Advancements in Multimodal Medical Image Fusion Techniques for Better Diagnosis: An overview," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 18, Jun. 2022.
8. M. A. Azam et al., "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," *Computers in Biology and Medicine*, vol. 144, p. 105253, May 2022.
9. M. Haribabu and V. Guruviah, "An Improved Multimodal Medical Image Fusion Approach Using Intuitionistic Fuzzy Set and Intuitionistic Fuzzy Cross-Correlation," *Diagnostics*, vol. 13, no. 14, p. 2330, Jan. 2023.
10. S. Ye, T. Wang, M. Ding and X. Zhang, "F-DARTS: Foveated Differentiable Architecture Search Based Multimodal Medical Image Fusion," in *IEEE Transactions on Medical Imaging*, vol. 42, no. 11, pp. 3348-3361, Nov. 2023.
11. V. Subbiah Parvathy, S. Pothiraj, and J. Sampson, "A novel approach in multimodality medical image fusion using optimal shearlet and deep learning," *International Journal of Imaging Systems and Technology*, vol. 30, no. 4, pp. 847–859, May 2020.
12. C. S. Asha, S. Lal, V. P. Gurupur and P. U. P. Saxena, "Multi-Modal Medical Image Fusion With Adaptive Weighted Combination of NSST Bands Using Chaotic Grey Wolf Optimization," in *IEEE Access*, vol. 7, pp. 40782-40796, 2019.
13. Y. Liu, B. Yan, R. Zhang, K. Liu, G. Jeon, and X. Yang, "Multi-Scale Mixed Attention Network for CT and MRI Image Fusion," *Entropy*, vol. 24, no. 6, p. 843, Jun. 2022.
14. C. Liu, X. Wu, X. Yu, Y. Tang, J. Zhang, and J. Zhou, "Fusing multi-scale information in convolution network for MR image super-resolution

reconstruction,” *Biomedical Engineering Online*, vol. 17, no. 1, Aug. 2018

15. X. Li, X. Zhang, and M. Ding, “A sum-modified-Laplacian and sparse representation based multimodal medical image fusion in Laplacian pyramid domain,” *Medical & Biological Engineering & Computing*, vol. 57, no. 10, pp. 2265–2275, Aug. 2019.
16. J. Fu, W. Li, J. Du, and B. Xiao, “Multimodal medical image fusion via laplacian pyramid and convolutional neural network reconstruction with local gradient energy strategy,” *Computers in Biology and Medicine*, vol. 126, pp. 104048–104048, Nov. 2020.
17. Richa, K. Kaur, and P. Singh, “A Novel MRI And CT Image Fusion Based on Discrete Wavelet Transform and Principal Component Analysis for Enhanced Clinical Diagnosis,” *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 10, pp. 64–82, Jul. 2022.
18. O. S. Faragallah, A. N. Muhammed, T. S. Taha, and G. G. N. Geweid, “PCA based SVD fusion for MRI and CT medical images,” *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 2, pp. 4021–4033, Sep. 2021.
19. O. Prakash, C. M. Park, A. Khare, M. Jeon, and J. Gwak, “Multiscale fusion of multimodal medical images using lifting scheme based biorthogonal wavelet transform,” *Optik*, vol. 182, pp. 995–1014, Apr. 2019.

