Unsupervised Convolutional Autoencoder Framework for Multimodal Medical Image Fusion in Brain Tumour Diagnosis

Saravanan Vijayan
Department of Electronics and Communication Engineering
SRM Institute of Science and Technology
Kattankulathur, India
sv8162@srmist.edu.in

Malarvizhi Subramani
Department of Electronics and Communication Engineering
SRM Institute of Science and Technology
Kattankulathur, India
malarvig@srmist.edu.in

Abstract: *Medical image fusion improves diagnosis accuracy and reliability by combining images from several modalities. It is gaining prominence for many clinical applications. This paper implements an unsupervised model to fuse gray-scaled Magnetic Resonance Imaging (MRI) with colored Positron Emission Tomography/Single Positron Emission Computed Tomography (PET/SPECT) medical image fusion to locate tumor-affected portions and dead cells clearly. This paper's main goal is in determining how well an autoencoder's encoder component can extract features from MRI and PET/SPECT images of brain tumor problems. The autoencoder's decoder component then uses the features to reconstruct the fused image. The autoencoders are tuned accordingly to get a low Mean Squared Error (MSE) with good structural similarity. It is trained with the dataset of MRI and PET/SPECT images in the whole brain atlas dataset, Harvard University. Our suggested approach has been objectively assessed using four distinct image assessment metrics: Feature Mutual Index (FMI), Structural Similarity Index Measure (SSIM), gradient-based Quality index (Qab/f) and Visual Information Fidelity Factor (VIFF) are compared to four other methods currently in use. In both subjective and objective assessments, our method has outperformed well compared to the existing methods in comparison.*

Keywords: *Unsupervised learning, autoencoders, image fusion, medical images, brain tumours.*

Received March 6, 2025; accepted June 23, 2025
https://doi.org/10.34028/iajit/22/5/11

1. Introduction

Medical images are vital for diagnosis as well as treatment in the medical field. Imaging methods include Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and Single Photon Emission Computed Tomography (SPECT). These images represent diverse organ details. CT scans are used to view the structure of the bone, whereas MRI images show internal or soft organ properties. Although the CT scan is more precise and can give precise information about bones, it excludes parenchyma and Cerebrospinal Fluid (CSF). high-fat tissues (parenchyma); yet, T1 and T2 weighted MRI scans can reveal the CSF; depending on the modality, the latter may seem darker or brighter. Similarly, PET and SPECT pictures provide low-resolution metabolic information of organs and are more accurate in capturing tumors.

Human abnormalities can be detected using the two oldest techniques namely PET and MRI. MRI scanning works well for soft tissues, while PET imaging works well for bone structures. For early abnormality detection PET imaging is recommended. Moreover, MRI imaging cannot show calcium anomalies also cough distorts MRI image output. Combining MRI and PET scans allows for a more accurate diagnosis of brain disease by examining the metabolism in specific regions of the cortex function compatibility [25].

For radiologist it may be required to view two modality images for better or clear diagnosis of a disease. If both images are integrated, the doctor will be able to accurately diagnose the ailment. The information in both images when combined either using image processing and latest advancement techniques in machine learning ML will aid the doctor's community.

Medical images integration or combining is popularly seems as image fusion in literatures It seeks to increase the use of medical images and assist physicians in deducing the information they contain. The fused images will provide additional information than a single medical scans image. It can assist medical professionals in making more thorough, timely, and accurate diagnoses and treatments [15].

Reviewing on the conventional methods for images fusion, two broader categories are spatial and transform domain. Transform domain methods has been commonly referred as Multi-Scale Transform (MST) techniques which includes Laplace Pyramid (LP), Wavelet Transform (WT) and Nonsubsampled Contourlet Transform (NSCT) and Stationary Wavelet Transform (SWT). These methods execute the

following actions: The source images will be separated into coefficients and merged using fusion rules. The final image will then be recreated by executing an inverse transform.

Feature space-based methods such as independent component analysis, sparse representation, etc., have been developed recently under MST technique. An appropriate fusion rule is necessary to combine information acquired from each image, which has an impact on the reconstruction of single-image quality. Primarily used methods are weight distribution and activity level assessment. Creating a weight map that incorporates pixel activity information from several sources is important for image fusion. Conventional transform domain fusion techniques use a decomposed coefficient's absolute value (or the sum of its values over a given time period) to determine its activity. An appropriate maximum or weighted-average fusion rule was employed by Zhou *et al*. [30].

Furthermore, after developments in Deep Learning (DL), a number of popular DL networks were introduced for image fusion, including Convolutional Neural Networks (CNNs), Visual Geometry Group Networks (VGGs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs) [10].

DL approaches provide robust feature extraction and easy implementation to ensure that more detailed information is successfully kept in the final fused image [1]. Because it performs well in extracting both high-level and low-level features, DL feature extraction has been widely employed in medical image fusion [29].

Compared to conventional image fusion methods discussed, recent, the DL machine methods has more obvious benefits.

1. The DL model improves its ability to express visual features by iterative training on larger data sets. However, it is more reliant on huge datasets.
2. Increased flexibility in network architecture. DL-based image fusion approaches can continually alter image quality during training, unlike older methods that require manual adjustments to the rules of the algorithm.
3. Traditional image processing needs an appropriate fusion rule to combine the image coefficients DL-based methods can efficiently combine the inherent signature of images.

In summary, though existing DL-based image fusion techniques have shown promising results, but they still have issues while fusing the information content of both the images with respect to noise, requirement of standard training reference images for supervised learning, and some require the use of irrational image fusion weight maps.

To address the aforementioned issues, we present a comprehensive DL model to fuse medical images which can take a grayscale image and a colour image directly from the Picture Achieve Communication System (PACS) and the final single image will be store in the PACS itself. We employ an unsupervised method to derive the internal features of each image.

The main contribution of this paper is an effective utilization of the unsupervised DL method to learn the important features about the image pixels which can be functional or anatomy details of scan area in a medical image. This is the first autoencoder-structure-based suggestion for brain tumor disease medical image fusion that we are aware of. Latent features of an input image that are smaller in size than the original image have been learned by the suggested autoencoder structure. The suggested unsupervised model's performance is compared with that of other machine learning models and traditional models.

2. Related Works

Image fusion has grown more quickly as a result of recent advancements in DL methods like autoencoders, GANs. The quality of fused images can be effectively influenced by DL networks' exceptional ability to express information and extract features. Image fusion aims in creating an informative fused image by first extracting and then integrating the most significant information from the source input images. Image fusion has advanced significantly with DL, and the fused findings are promising due to neural networks' enhanced feature extraction and reconstruction capabilities [30].

A convolutional network was used in building weight map which integrates information on pixel activity from two input source images as part of Liu *et al*. [15] medical image fusion. CNN-based image fusion was introduced by Liu *et al*. [16] who saw that task as a classification issue and used CNN to create decision maps and classified image regions., The most efficient technique for feature extraction and image reconstruction at that time was CNN. Pulse Coupled Neural Network (PCNN) is global fusion technique which uses signal processing techniques akin to those of the human visual nerve system while preserving precise information [5].

Sub-band fusion rules serve as the foundation for transform-based approaches. While adaptive transforms have a somewhat long execution time, first-generation transforms fail to achieve good directional decomposition. Approaches based on sparse representation rely on a compact dictionary that is challenging to create with strong representational capabilities. These approaches are not appropriate for real-time applications because of their high costs. DL techniques necessitate a large training set and high-performance computers. In this case, it is quite difficult task to design an appropriate network architecture. Consequently, Vajpayee *et al*. [23] suggested fusion technique in the domain of non-subsampled Shearlet

transform that combines high-pass sub-bands using cutting-edge Adaptive Gaussian Pulse Coupled Neural Network (AGPCNN) and low-pass sub-bands using enhanced Robert's operator (edge detection-based system). Li *et al*. [12] suggested novel image fusion technique using Coupled Neural P systems (CNP). CNP systems regulate the fusing of the low-frequency coefficients of NSST by using two CNP systems. A unique technique for fusing Visible (VI) and Infrared (IR) images utilizing Stacked Sparse Auto-Encoders (SSAE) and NSCT was suggested by Luo in order to successfully incorporate the infrared item into fused image [17].

The medical image fusion technique utilizing the NSCT and PCNN techniques was proposed by Ibrahim [8]. Low and high frequency subbands were extracted from the input source images using the NSCT approach. These subbands are integrated by the PCNN, a fusion rule. In order to reproduce the fused image, the inverse NSCT method was applied.

For NSST, Sebastian and King [22] suggested a CNN-based MRI and PET image fusion technique. First, the PET image is converted to the YUV color space. CNN creates a weight map using the Y element of the MRI and PET. The generated weight map is decomposed using NSST into MRI and Y PET components. To fuse the deconstructed bands, similarity-based fusion criteria are used. Inverse NSST is used to restore the fused image.

Panigrahy *et al*. [19] suggested Weighted Parameter Adaptive Dual Channel PCNN (WPADCPCNN) for medical fusion using non-subsampled shearlet transform for combining SPECT and MRI of patients with Alzheimer's disease and aids dementia complex. Fractal dimension is used in estimating the parameters of suggested WPADCPCNN model extracted from the sources. End-to-end image fusion techniques use a DL network from source to fused images. The network's inputs are the source images, while its outputs are the fused images. Zhang *et al*. [27] suggested the Image Fusion Convolutional Neural Network (IFCNN), an end-to-end image fusion.

DL-based technique for merging multispectral and panchromatic images in remote sensing applications was presented by Azarang *et al*. [2]. Training of convolutional autoencoder network is carried out in creating original panchromatic images from spatially degraded ones using this technique, which is categorized as a component replacement method.

Li *et al*. [14] suggested a supervised learning-based CNN-based multimodal medical image fusion method in addressing the real-world problem of medical diagnosis. In order to meet medical diagnosis criteria, Li *et al*. [11] suggested multi-mode medical image fusion with DL, taking into consideration the features of multi-modal medical images, practical implementation, medical diagnostic technology. Wang *et al*. [24] suggested medical image fusion technique utilizing CNN with distinct structural elements along with excellent visual qualities.

Weighted average fusion technique was presented by Bavirisetti *et al*. [3] for combining brain CT and MRI images. A weighted average approach based on guided filter that utilizes spatial consistency was proposed by Li *et al*. [13] to seamlessly merge base and detail layers.

The proposed technique separates an image into two scales: a detail layer for small details and a base layer for large intensity fluctuations. Munawwar and Rao [18] proposed a novel enhanced MMIF technique for medical image fusion. The proposed research used two upgraded DL algorithms to extract and merge relevant and distinguishing characteristics from source images. This method combined the benefits of both feature sets to produce high-quality, contextually and technically rich images. Cheng *et al*. [4] proposed a self-evolutionary training approach using a novel Memory Unit Fusion architecture (MUFusion). In this unit, interim fusion was used for training procedure to supervise the merged image.

Zhang *et al*. [28] presented a self-supervised system for multi-modal medical fusion problems that uses contrastive auto-encoding and convolutional information sharing. Multi-modal medical images share common features, which can lead to information redundancy when extracted in pairs. Xu and Ma [26] presented an unsupervised enhanced medical picture fusion network. Surface-level and deep-level limitations were applied to preserve information effectively. The surface-level constraint relies on saliency and abundance measurements to maintain subjective and intuitive qualities. Deep-level constraints objectively specify unique information based on a pretrained encoder's channels.

3. Unsupervised Learning-Based MRI-PET/SPECT Fusion

The goal of any unsupervised learning model is to find any interesting patterns in the dataset with no labels. We approached the medical image fusion as an unsupervised learning which can provide essential features of two images. For example: suppose the inputs image is of size 10x10 the pixel (100 pixels intensity values) and hidden nodes of size 50, then the network is forced to learn the compressed representation of the input and with the size of 50 and we need to train the neural network model to get the latent feature through which the original image of size 100 can be reconstructed. This approach has been employed for image compression, noise removal etc., in computer vision applications.

Our objective in this work is to use a DL-based unsupervised approach for medical picture fusion called convolutional autoencoder. Silent information of input image has to be learnt as features by the proposed autoencoder model. Figures 1 and 2 picture the

complete framework of our proposal. Source images each of size NxN from the PACS system or in the storage space of respective imaging instrument is considered as input to the processing blocks and the output final image will be a single image of size NxN. Autoencoders A and B based on CNN help in getting the interesting and useful latent vector space of images in grey scale and colour. The use of RGB channels may result in significant color distortion; therefore, separating the luminance component using YUV color space transform techniques has been a proven strategy because it considers human perception and is suitable for fusing functional and anatomical images. The YUV method divides a color image into one luminance component (Y), U channel and V channel chrominance components.

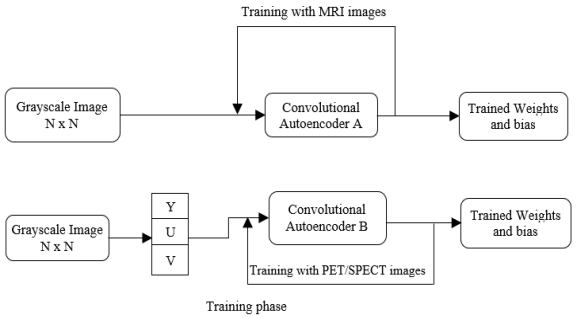


Figure 1. Framework for medical image fusion using autoencoder: training phase.

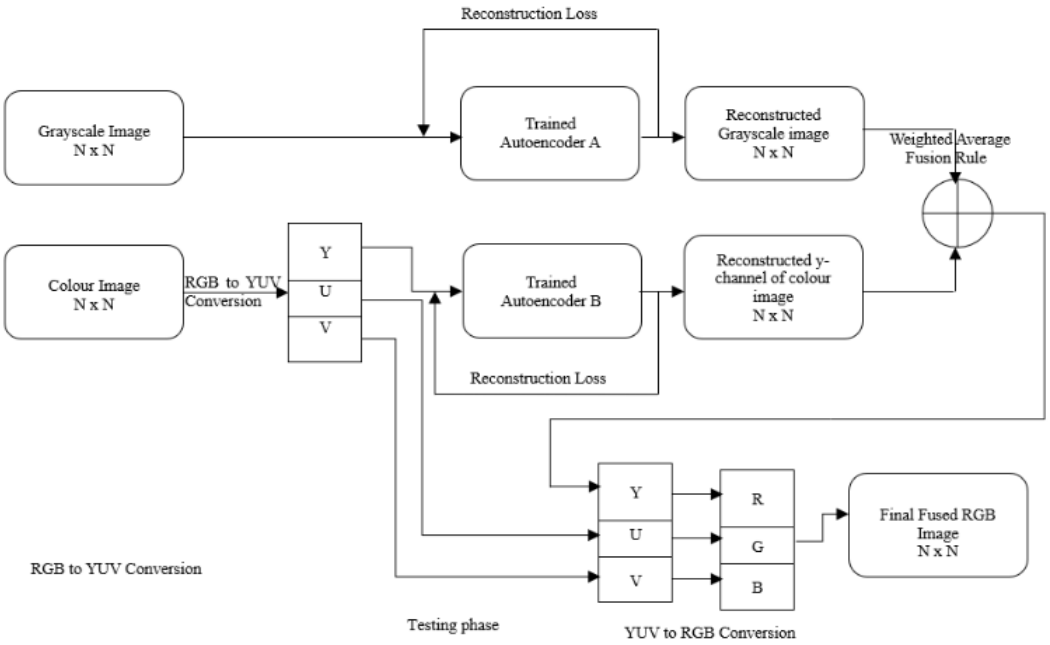


Figure 2. Framework for medical image fusion using autoencoder: testing phase.

Autoencoder includes two CNN networks: an encoder and a decoder. They were trained on images from the standard datasets of MRI and PET/SPET, as explained in the following sections. The encoded features of each image were blended using the weighted average technique. The combined vector space of the Y component of a color and grey image is then mixed with the U and V channels. On the RGB color space, we shall get a single image of size NxN.

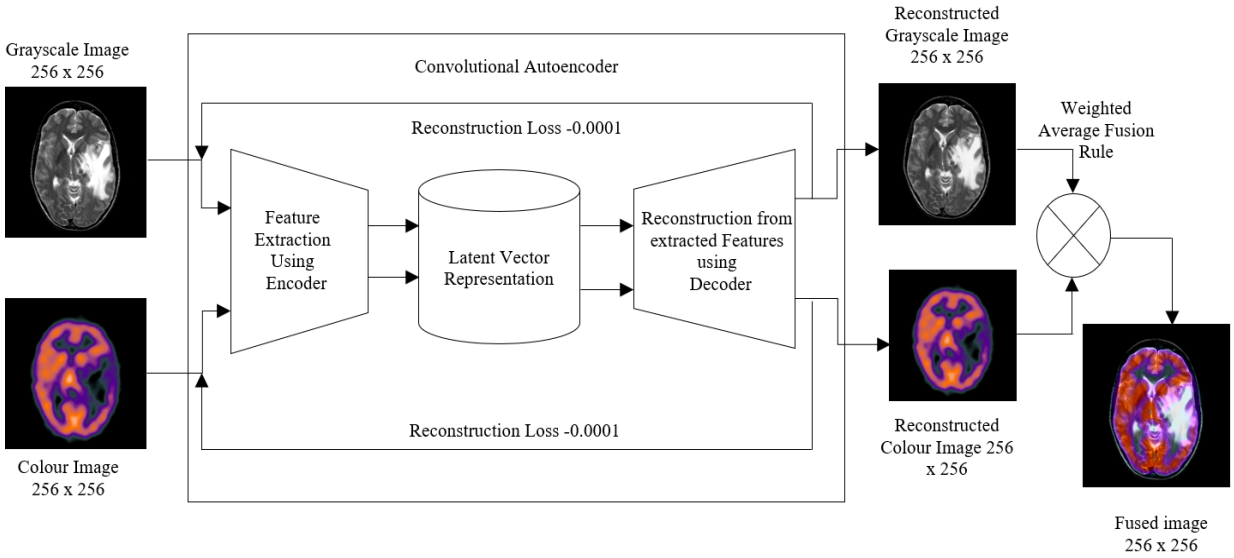


Figure 3. Proposed method.

Figure 3 depicts the outline of our proposed unsupervised autoencoder for MRI-PET/SPECT fusion with encoder and decoder part. Two Encoders were utilized one to extract the features of grayscale MRI images of size 256x256 and the other one to extract the features of y-channel of colored PET/SPECT images of size 256x256. The latent vector representation of MRI and PET/SPECT images with dimensionality reduction to 64x64 are obtained as the outputs of respective encoders.

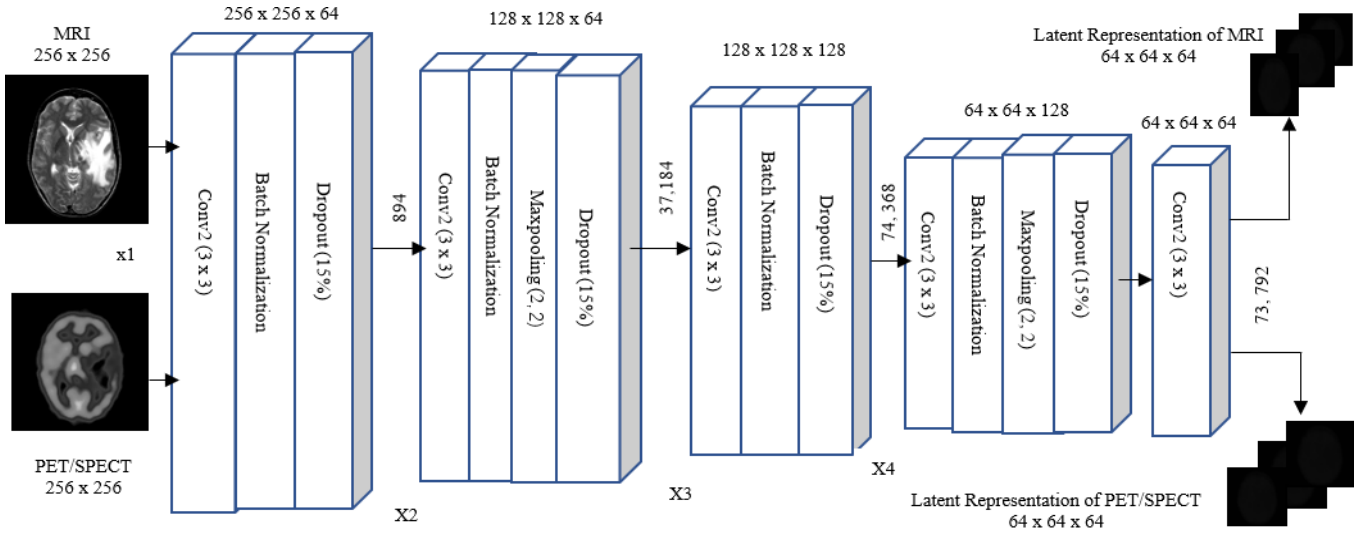


Figure 4. Detailed architecture of encoder and decoder of our proposed method.

Two decoders were used and they are trained with grayscale images and colored images respectively. These latent vectors are given as inputs to the respective decoders and MRI and PET/SPECT images are reconstructed with the original size of 256x256. Then appropriate fusion rule is utilized to obtain the fused image. Figure 4 depicts the detailed architectures of encoder and decoder parts of our proposed fusion method.

3.1. Encoder

It is anticipated that n convolution filters will produce n intermediate features in order to collect the structural properties of the input image data. An intermediate feature maps represent the input image's compressed form. It is common practice to concatenate many convolution layers. The encoder part consists of four convolutional layers of stride (2, 2) along with batch normalization and dropout of 15%.

Convolution layers extract local patterns, like edges, textures, or particular features, from the input image by applying a collection of learnable filters, called kernels. Every convolution process entails:

$$Output = \sum_{i,j} input\ (i,j)\ x\ kernel\ (i,j) \quad (1)$$

Batch normalization helps preventing the overfitting and speed up the training of the network. The activations of every layer are normalized by statistical analysis. Through the introduction of noise or unpredictability into the network, dropout reduces the probability of overfitting and the co-adaptation of neurons. Activation Function provides non-linearity and enable the model to learn more complex features, an activation function (such as ReLU) is frequently used to the convolution output:

$$f(y) = max(0\ , y) \quad (2)$$

A 2x2 Maxpooling are used to reduce the spatial dimension of feature maps thereby retaining important features and reducing computational load. The dimensionality of the input image by progressively extracting prominent features gets reduced while discarding less significant ones. The encoder finally maps the input to a lower-dimensional feature vector (bottleneck layer), z, which represents compressed or encoded information:

$$z = f(Wy + b) \quad (3)$$

where b is the bias, W is the weight matrix, and y is the input.

3.2. Decoder

The decoder part consists of six convolution layers with 2x2 upsampling and skip connections. The upsampling layers increase the spatial dimension of the feature map to its original size, often using methods like nearest-neighbor interpolation, bilinear interpolation, or transposed convolutions. Transpose convolution layers are used similar to convolution but works in reverse, reconstructing the image by spreading out features over a larger space. Mathematically, this is performed by "convolving" the feature map with a set of transposed filters.

LeakyRelu is used as non-linear activation function to enable complex mappings and gradual reconstruction. The final layer of the decoder aims in reconstructing the image by mapping the feature

representation back to the original input dimensions. If the goal is grayscale image reconstruction, a sigmoid activation is typically applied to constrain the output values between 0 and 1:

$$\hat{x} = \sigma(Wz + b) \quad (4)$$

where σ is sigmoid function, b is the bias, z is the latent code and W is weight matrix.

The encoder encodes input data into a compressed feature vector, while the decoder applies inverse operations to reconstruct the original image from this compact representation. A total of over 7 lakhs trainable parameters and over 700 non-trainable parameters are taken into consideration for each source image for feature extraction and final reconstruction of fused image. Tables 1 and 2 show the number of parameters involved in extracting the features and reconstruction of MRI and PET/SPECT

Table 1. Number of parameters involved for feature extraction of given images in encoder part.

Layers	Output shape	Parameters	Connected to
input_layer	(None, 256, 256,1)	0	-
conv2d(Conv_2D)	(None, 256, 256,64)	640	input_layer [0][0]
batch_normalization	(None, 256, 256,64)	256	con2d[0][0]
dropout	(None, 256, 256,64)	0	batch_normalization [0][0]
conv2d_1(Conv_2D)	(None, 256, 256,64)	36,928	dropout [0][0]
max_pooling2d	(None, 128, 128,64)	0	conv2d_1[0][0]
Batch_Normalization_1	(none, 128, 128,64)	256	max_pooling2d[0][0]
dropout_1	(None, 128, 128,64)	0	Batch_Normalization_1[0][0]
conv2d_2(Conv_2D)	(None, 128, 128,128)	73,856	dropout_1[0][0]
Batch_Normalization_2	(None, 128, 128,128)	512	conv2d_2[0][0]
dropout_2	(None, 128, 128,128)	0	Batch_Normalization_2[0][0]
conv2d_3(Conv_2D)	(None, 128, 128,128)	147, 584	dropout_2[0][0]
max_pooling2d_1	(None, 64, 64,128)	0	conv2d_3[0][0]
Batch_Normalization_3	(None, 64, 64,128)	512	max_pooling2d_1[0][0]
dropout_3	(None, 64, 64,128)	0	Batch_Normalization_3[0][0]
conv2d_4(Conv_2D)	(None, 64, 64,64)	73, 792	dropout_3[0][0]

Table 2. Number of parameters involved for reconstruction of images using extracted features of given images in decoder part.

Layers	Output shape	Parameters	Connected to
conv2d_5(Conv_2D)	(None, 64, 64,128)	73, 856	conv2d_4[0][0]
up_sampling2d	(None, 128, 128,128)	0	conv2d_5[0][0]
leaky_re_lu (LeakyRelu)	(None, 128, 128,128)	0	up_sampling2d [0][0]
add (Add)	(None, 128, 128,128)	0	leaky_re_lu [0][0], dropout_2[0][0]
conv2d_6(Conv_2D)	(None, 128, 128,128)	147, 584	add [0][0]
leaky_re_lu _1 (LeakyRelu)	(None, 128, 128,128)	0	conv2d_6[0][0]
add_1 (Add)	(None, 128, 128,128)	0	leaky_re_lu_1[0][0], dropout_2[0][0]
conv2d_7(Conv_2D)	(None, 128, 128,64)	73, 792	add_1[0][0]
up_sampling2d_1	(None, 256, 256,64)	0	conv2d_7[0][0]
leaky_re_lu_2 (LeakyRelu)	(None, 256, 256,64)	0	up_sampling2d__1[0][0]
add_2 (Add)	(None, 256, 256,64)	0	leaky_re_lu_2[0][0], dropout [0][0]
conv2d_8(Conv_2D)	(None, 256, 256,64)	36, 928	add_2[0][0]
leaky_re_lu_3 (LeakyRelu)	(None, 256, 256,64)	0	conv2d_8[0][0]
conv2d_9(Conv_2D)	(None, 256, 256,64)	36, 928	leaky_re_lu_3[0][0]
leaky_re_lu_4 (LeakyRelu)	(None, 256, 256,64)	0	conv2d_9[0][0]
conv2d_10(Conv_2D)	(None, 256, 256,1)	577	leaky_re_lu_4[0][0]

4. Experimental Results

In general, most of the medical fusion methods are based on PCNN and CNN. We find few works based on Unsupervised learning. Moreover, Autoencoders based fusion method is more beneficial compared to conventional PCNN and CNN-based fusion techniques.

Autoencoders are better able to adjust to complicated and diverse datasets since we have used to two diverse input images such as MRI and PET. On the other hand, PCNNs are manually constructed models that are built on biologically inspired concepts and do not learn from data in the same manner. They are more rigid and task-specific. Autoencoders are useful for classification, segmentation, and fusion because they extract more compact and informative latent representations. PCNNs are not optimized to model such complex interaction and they are typically limited to basic features like edges, regions, or intensity contrast.

In image fusion, autoencoders can reconstruct fused images from learned joint features (especially in encoder-decoder setups). CNNs don't inherently reconstruct inputs and they are mostly discriminative, not generative. Autoencoders are better for fusion models, especially for MRI+CT or PET+MRI fusion, where combining modalities and reconstructing a high-quality image is key. Since, Unsupervised method offers more advantages compared to the conventional methods, we emphasized on this technique to implement effective implementation of fusion method of two diverse medical images such as MRI and PET/SPECT. Four existing methods are used to assess the efficacy of the suggested approach: Gaussian-PCNN-LP [23], Hahn-PCNN-CNN [5], CNP-MIF [12] and GFF [3]. For experimentation, a standardized pair of MRI and PET/SPECT scans from the whole brain atlas dataset were utilized as benchmarks. The suggested approach and other methods from the literature were run on an

Intel Core i3-4010U CPU at 1.70GHz with 4GB RAM using python.

The whole brain atlas [9] is an online resource for imaging of brain diseases developed together by the American Academy of Neurology, Harvard Medical School, Countway Library of Medicine, Radiology Department, and the Neurology Departments at Brigham and Women's Hospital. The website is structured into six main sections:

1) An atlas of normal anatomy.
2) A neuroimaging primer for those who are unfamiliar with imaging terminology.
3) Cerebrovascular illness.
4) Neoplastic disease.
5) Inflammatory disease.
6) Degenerative disease. It contains more than 13,000 brain images from MRI, CT, SPECT, and PET scans of 30 brain disorders.

Four brain disorders have been taken into consideration for our experiments.

a) Glioma.
b) Alzheimer's disease.
c) Metastatic bronchogenic.
d) Huntington's disease.

4.1. Parameters and Hyper-Parameters Analysis

In the process of fusing MRI and PET/SPECT images using an unsupervised autoencoder-based framework, accurate spatial alignment of the two modalities is a critical pre-processing step. Due to differences in acquisition geometry, resolution, and contrast, PET/SPECT and MRI images are often not inherently aligned. Therefore, we applied multimodal image registration prior to feeding the data into the autoencoder. Multimodal image registration is the process of aligning images from different imaging modalities such as MRI (structural) and PET/SPECT (functional/metabolic) so that corresponding anatomical or functional regions overlap accurately.

Specifically, the PET/SPECT images were registered to the corresponding MRI images using a rigid (or affine) registration approach, ensuring that both images represent the same anatomical structures in the same spatial configuration. We used Mutual Information (MI) as the similarity metric, as it is well-suited for multimodal registration tasks due to its robustness to differences in intensity distributions across imaging modalities. In order to achieve corresponding overlapping of anatomical and functional regions of MRI and PET/SPECT, both the images are resized to 256x256 and images of same brain tumour disease of same slice of the brain are used.

Regarding dataset, MRI and PET/SPECT images of each 2000 were used as dataset for training the network with 1240 MRI-PET images and 480 MRI-SPECT images of Glioma disease; 400 MRI-PET images of Alzheimer's disease; 440 MRI-SPECT images of metastatic bronchogenic disease; 480 MRI-SPECT images of hypertensive encephalopathy; 560 MRI-SPECT images and 400 MRI-SPECT images of motor neuron disease and normal aging. Out of these 4000 MRI-PET/SPECT images, 3200 MRI-PET/SPECT images were used for training, 400 MRI-PET/SPECT pairs for validation and remaining 400 images for testing were used.

On experimentation, the proposed deep convolutional autoencoder architecture is tested with 50 and 100 epochs and tested with batch sizes of 32, 16, 8 and 4 with different learning rates such as 0.001 and 0.0001. Since these are the important hyperparameters required for training of autoencoders to get good fused image with low loss and good structural similarity, experimentation was carried out for different hyperparameters and results are saved. The network is trained with 2000 MRI and PET images of different brain tumour diseases which are available in Whole Brain Atlas, Harvard University.

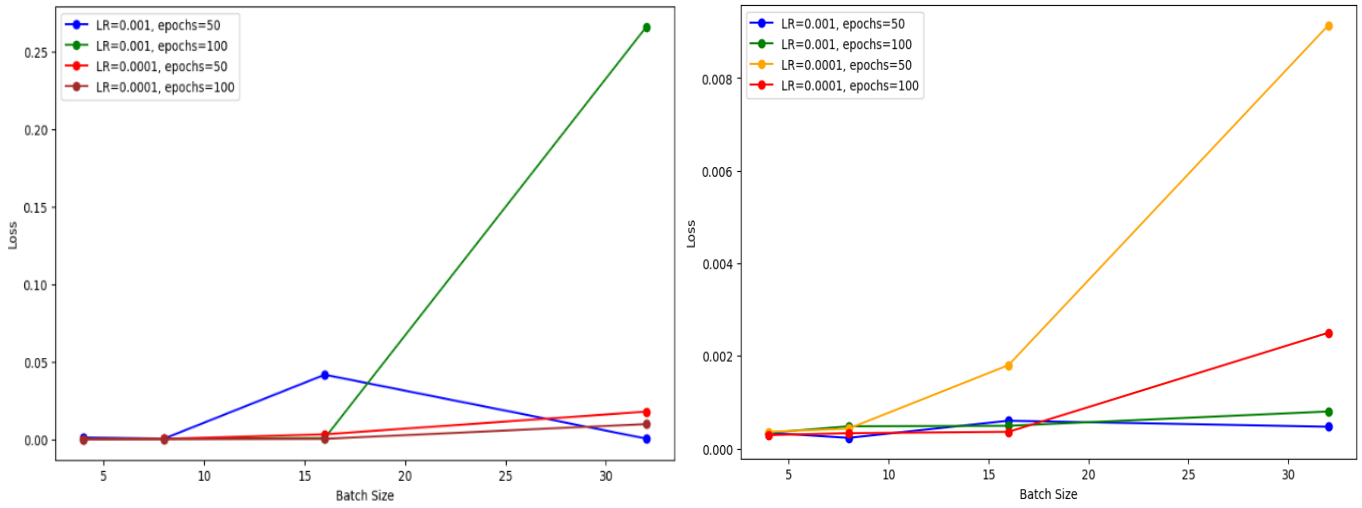
Changing the number of epochs in training an autoencoder directly affects its ability to extract meaningful features from images. Here's how it works and what to consider when using 50 epochs vs. 100 epochs for feature extraction. Lower epochs aid faster training time, which is useful for initial testing or when computational resources are limited. It prevents overfitting if the model starts memorizing the data instead of learning generalizable features. The autoencoder might underfit, meaning it has not fully captured the patterns in the data. Extracted features might not be as robust or detailed.

Training model with higher epochs provides more training time for the model to learn and adapt, leading to better feature extraction. Useful for more complex datasets where patterns are harder to learn. Risk of overfitting, especially if there's no validation set to monitor performance. Increased training time and computational cost. Ensure the learning rate is appropriately tuned. If it's too high, increasing epochs might not improve performance. If it's too low, training might require more epochs to converge. Monitor metrics like reconstruction loss during training to determine the optimal number of epochs. The learning rate has a major impact on how well autoencoders perform when extracting features from images. During training, it establishes the step size at which the model changes its weights.

Here's how learning rates of 0.001 vs. 0.0001 can impact the training and feature extraction capabilities of an autoencoder. The model with higher learning rate learns more quickly since the updates to weights are larger. Can quickly learn significant patterns in large datasets. Rapidly decreases reconstruction loss, especially in the first few epochs. The drawbacks are it may overshoot the optimal weights, resulting in unstable

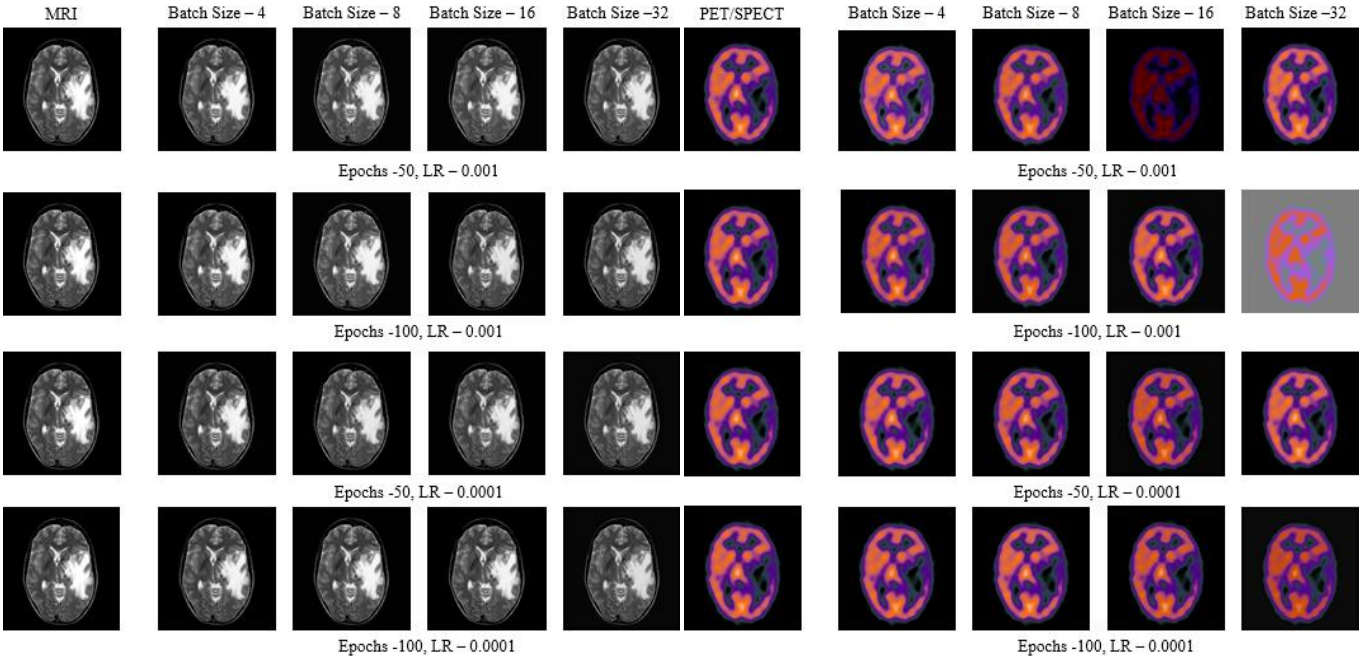
training or poor convergence. The extracted features may be less precise due to insufficient fine-tuning. If combined with a large batch size or unnormalized input, the training might diverge.

Model with lower learning rate will reduce the risk of overshooting, leading to smoother convergence. The model has more time to adjust weights and capture subtle patterns in the data. Especially useful for sensitive architectures like Variational Auto-Encoders (VAEs) or when the data contains high variance. The drawbacks are training takes longer to reach the optimal solution, increasing computation time. The model might converge to a suboptimal solution if the learning rate is too small. A minimal number of epochs could prevent the model from learning the data distribution completely.



a) Loss v. batch sizes for different learning rates and epochs/PET/SPECT image. b) Loss v. batch sizes for different learning rates and epochs/MRI image.

Figure 5. Batch size vs loss curve for different epochs, batch sizes and learning rates [results obtained from program].



a) Loss v. batch sizes for different learning rates and epochs/PET/SPECT image. b) Loss v. batch sizes for different learning rates and epochs/MRI image.

Figure 6. Hyperparameter tuning with different epochs, learning rate and batch sizes for MRI and PET/SPECT images.

Table 3. Search space for optimization of hyperparameter tuning of autoencoders for MRI images with different batch sizes and epochs for learning rates -0.001 and 0.0001.

MRI							
Learning rate-0.001				Learning rate-0.0001			
Epochs=50		Epochs=100		Epochs=50		Epochs=100	
Batch size	Loss	Batch size	Loss	Batch size	Loss	Batch size	Loss
4	0.00034	4	0.00034	4	0.00036	4	0.00029
8	0.00023	8	0.00048	8	0.00043	8	0.00033
16	0.0006	16	0.00049	16	0.0018	16	0.00036
32	0.00047	32	0.0008	32	0.00914	32	0.0025

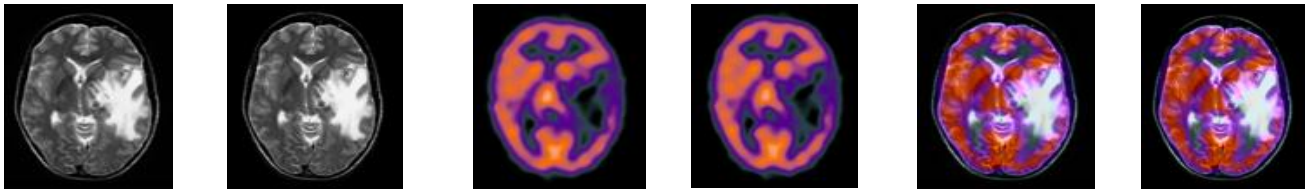
Table 4. Search space for optimization of hyperparameter tuning of autoencoders for PET/SPECT images with different batch sizes and epochs for learning rates -0.001 and 0.0001.

PET/SPECT							
Learning rate-0.001				Learning rate-0.0001			
Epochs=50		Epochs=100		Epochs=50		Epochs=100	
Batch size	Loss	Batch size	Loss	Batch size	Loss	Batch size	Loss
4	0.0012	4	0.00021	4	0.00014	4	0.00012
8	0.0005	8	0.00043	8	0.00034	8	0.00018
16	0.0417	16	0.00084	16	0.0033	16	0.00037
32	0.0005	32	0.2658	32	0.0179	32	0.0099

Figure 5 illustrates how loss varies with batch sizes for various epochs and learning rates. Figure 6 shows the reconstructed images of MRI and SPECT obtained for different epochs, learning rates and batch sizes in order to achieve optimal values of low loss and good structural similarities. Tables 3 and 4 show the search space for optimization of hyperparameter tuning of autoencoders for MRI and PET/SPECT images with different batch sizes and epochs for learning rates -0.001 and 0.0001. In order to get the optimal values, Table 5 shows the final optimized values chosen for our proposed work.

Table 5. Finalized parameters of autoencoders after hyper-tuning.

Parameter	Recommended value	Reason
Batch size	4 or 8	Balances computational efficiency, stability in training, and ability to retain fine details.
Epochs	100	Provides sufficient time for learning intricate medical image details without overfitting.
Learning rate	0.0001	Ensures stable updates and avoids overshooting, critical for preserving subtle patterns.
Optimizer	Adam	Combines benefits of momentum and adaptive learning, improving convergence for complex images.
Loss function	Mean Squared Error (MSE)	Ideal for reconstruction tasks to minimize pixel-wise differences.
Activation function	ReLU (encoder), Sigmoid (decoder)	ReLU avoids vanishing gradients, Sigmoid ensures output values match grayscale range (0-1).



a) Reconstructed MRI image with no skip connection. b) Reconstructed MRI image with skip connection. c) Reconstructed SPECT image with no skip connection. d) Reconstructed SPECT image with skip connection. e) Final fused image with no skip connection. f) Final fused image with skip connection.

Figure 7. Ablation study for with and without skip connection in network.

Table 6. Objective assessments of ablation study with and without skip connection.

	FMI	VIFF	SSIM	Q^f_{ab}
Without skip connection	0.85	0.50	0.80	0.49
With skip connection	0.85	0.72	0.85	0.67

Ablation study was carried out to insist the importance of adding skip connections in the network. The experiment was carried out by training the network with and without skip connections. The subjective assessments of these studies are shown in Figure 7. Subjectively, the clarity and sharpness of the images are improved by including skip connections. This has proven using the objective assessments of these studies which are presented in Table 6. It is clear that increase in Visual Information Fidelity Factor (VIFF) and Gradient based Quality Index (Qab/f) indicate that the sharpness and visual clarity of the image are enhanced by using skip connections. Changing the latent dimension have not affected the output.

4.2. Image Quality Assessment (IQA)

Subjective and objective assessments are the two categories of evaluations available for fused images. Human visual perception is the foundation for subjective evaluation of image quality. Although its benefits, such speed and convenience, cannot be overlooked, it can occasionally be subjective and quite sensitive to light environments and even human emotions. Furthermore, various people have diverse opinions about the same image. Evaluating the image in accordance with IQA is the objective assessment. Despite the obvious limitations of the evaluation results and the requirement to apply codes, the benefit is that the outcome is objective and unaffected by subjective factors. Both of these assessments are used in this paper. The metrics used are listed below in Table 7:

Table 7. IQA metrics for image fusion.

Metrics	Description and formula	
Feature MI [6]	represents data that corresponds to the attributes of input images required in obtaining fused image	$FMI^{AB}=I_{FA}+I_{FB}$
Visual Information Fidelity Factor (VIFF) [7]	relates image information with visual features of fused image.	$VIFF(I_A, I_F) = \frac{1}{2}\, log_2\left(\frac{\left\|g^2_{i,n}s^2_{i,n}C_u + (\sigma^2_{i,n} + \sigma^2_N)I\right\|}{\left\|(\sigma^2_{i,n} + \sigma^2_N)I\right\|}\right)$
Structured Similarity Index (SSIM) [20]	used to describe features that show how the brightness and contrast of the source as well as fused images are similar.	$SS(A,B) = \frac{(2\mu_A\mu_B + r_1)(2\sigma_{AB} + r_2)}{(\mu_A^2 + \mu_B^2 + r_1)(\sigma_A^2 + \sigma_B^2 + r_2)}$
$\mathbf{Q^f_{ab}}$ [21]	is a measure of image quality that evaluates the sharpness and contrast of edges in the image. It is calculated utilizing the gradient magnitude of the image.	$Q^{ab/f} = \frac{\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}(Q^A(m,n)W^A(m,n) + Q^B(m,n)W^B(m,n))}{\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}(W^A(m,n) + W^B(m,n))}$

The effectiveness of the proposed fusion algorithms and quality of the fused images may be objectively assessed using performance indicators such as Qab/f, Feature Mutual Information (FMI), VIFF, and SSIM are employed. MSE and SSIM are the quality evaluation metrics that, as an output, represents the image quality as perceived by human eyes.

4.2.1. Feature Mutual Information (FMI)

Measures how much shared information is retained

from the source images MRI and PET in the fused image. MI is computed as the mutual information between image features (e.g., gradient maps or Laplacian energy maps) of the fused image and each source image. Computation is lightweight (~tens of ms per image pair). Memory usage is minimal as only local feature maps (e.g., gradients) are stored.

4.2.2. Visual Information Fidelity Factor (VIFF)

Evaluates visual fidelity by comparing the information content between the source and fused images using natural scene statistics. Based on multi-scale analysis of wavelet coefficients and human visual system modeling.

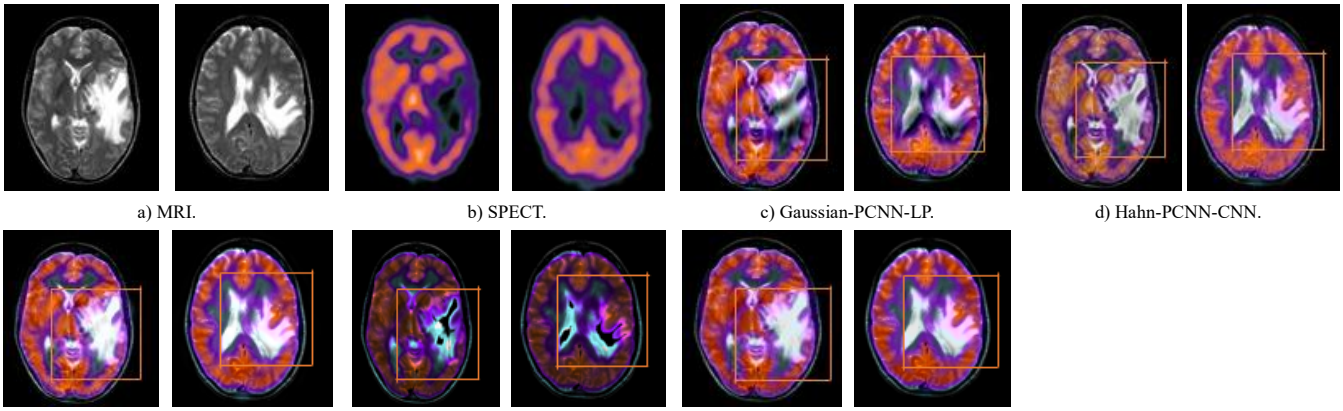
It evaluates how much perceptual information is preserved. More computationally expensive due to multi-scale decomposition. For 256×256 images, typical computation time is around 200-400 ms. Moderate memory use (~20-30 MB per image pair), depending on the number of decomposition levels.

4.2.3. Structural Similarity Index (SSIM)

Assesses how well the fused image retains the structure, luminance, and contrast of the source images. SSIM is computed between the fused image and each source image. Very fast (few ms per image), very low memory usage and suitable for real-time evaluation.

4.2.4. Gradient-Based Quality Index (Qab/f)

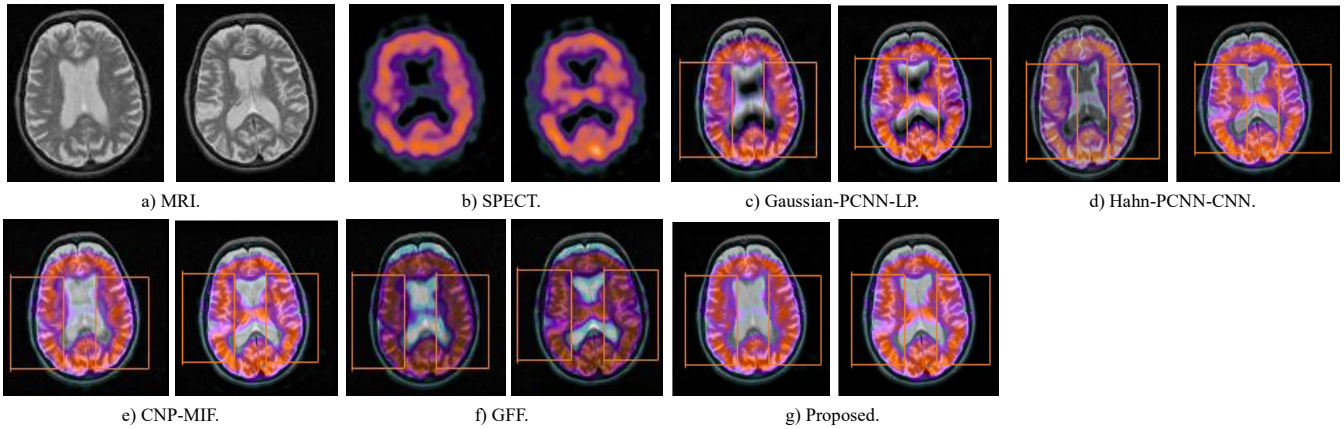
Measures edge preservation and contrast sharpness using gradient magnitudes. First compute gradients (e.g., using Sobel filters) of all images and then measure correlation between gradients of fused and source images. Slightly more computationally intensive than SSIM but efficient. Execution time per image pair: ~50–100 ms. Suitable for analyzing spatial detail and sharpness retention.



a) MRI. b) SPECT. c) Gaussian-PCNN-LP. d) Hahn-PCNN-CNN.

e) CNP-MIF. f) GFF. g) Proposed.

Figure 8. Fusion results of MRI-SPECT pairs of metastatic bronchogenic (slice-10 and slice-13).

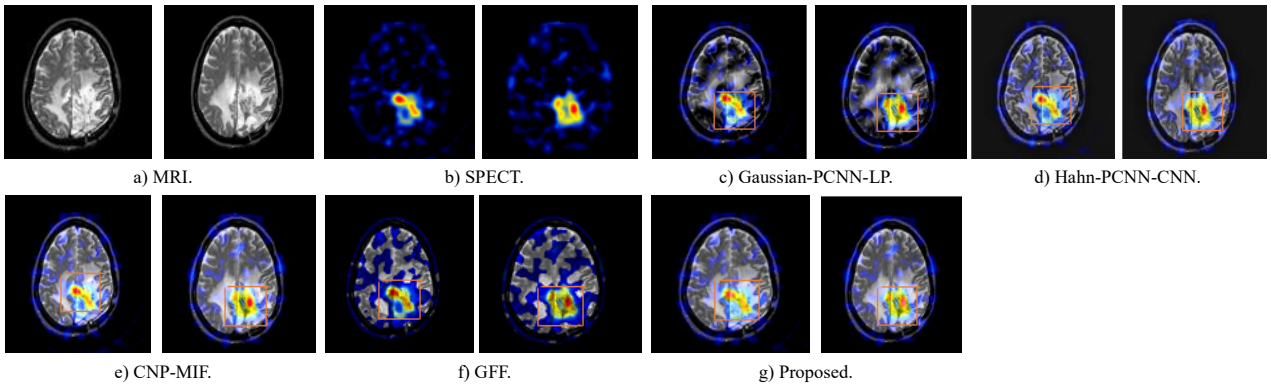


a) MRI. b) SPECT. c) Gaussian-PCNN-LP. d) Hahn-PCNN-CNN.

e) CNP-MIF. f) GFF. g) Proposed.

Figure 9. Fusion results of MRI-SPECT pairs of Huntington's disease (slice-15 and slice-16).

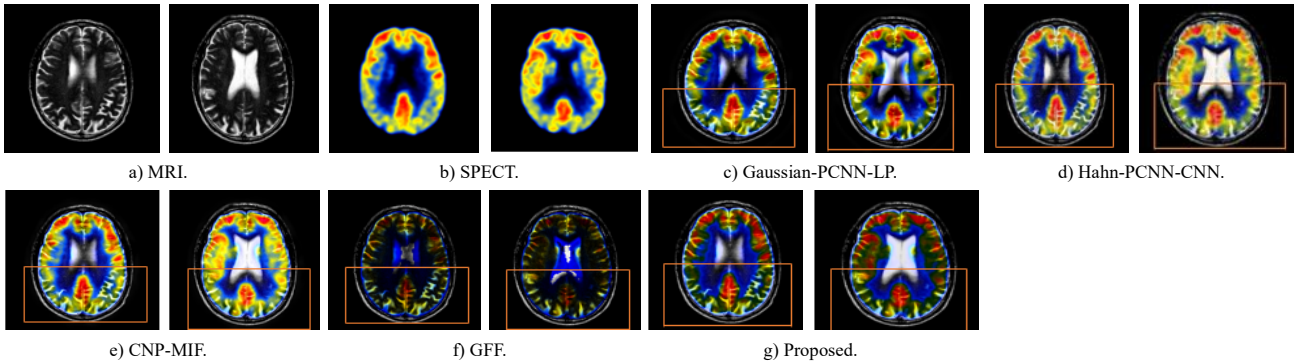
Figure 8 shows the MRI-SPECT pairs of different slices of Metastatic Bronchogenic (slice 10 and slice 13). Figure 9 shows the fused images of MRI-SPECT pairs of Huntington's disease (slice 15 and slice 16). Figures 10 and 11 show the fused images of MRI-SPECT pairs of Glioma (slice 36 and slice 38) and MRI-PET pairs of Alzheimer's disease (slice 15 and slice 16) respectively of existing methods and ours along with source images. The fused images are critically reviewed by a radiologist, at SRM Medical College Hospital and Research Centre, Kattankulathur. Decreased metabolic activities are seen clearly as red patches. Clear delineation of tumour affected area and a clear view of tumour and Edema regions are the salient features of our proposed method.



a) MRI. b) SPECT. c) Gaussian-PCNN-LP. d) Hahn-PCNN-CNN.

e) CNP-MIF. f) GFF. g) Proposed.

Figure 10. Fusion results of MRI-SPECT pairs of Glioma (slice-36 and slice-38).



a) MRI. b) SPECT. c) Gaussian-PCNN-LP. d) Hahn-PCNN-CNN.

e) CNP-MIF. f) GFF. g) Proposed.

Figure 11. Fusion results of MRI-PET pairs of Alzheimer's disease (slice-15 and slice-16).

Table 8. Objective Assessments of proposed and existing methodologies.

Alzheimer [MRI-PET pair-slice-16]					
FMI	0.69	0.87	0.88	0.86	**0.89**
SSIM	0.63	0.81	0.81	0.85	**0.94**
VIFF	0.54	0.38	0.40	0.33	**0.63**
Q^f_{ab}	0.53	0.55	0.59	0.46	**0.64**
Metastatic Bronchogenic disease [MRI-SPECT pair-slice 10]					
FMI	0.69	**0.88**	**0.88**	0.86	0.87
SSIM	0.76	0.88	0.90	0.84	**0.98**
VIFF	0.62	0.52	0.57	0.34	**0.72**
Q^f_{ab}	0.57	0.58	0.63	0.44	**0.67**
Metastatic Bronchogenic disease [MRI-SPECT pair-slice 13]					
FMI	0.70	0.88	**0.89**	0.87	0.88
SSIM	0.85	0.92	0.94	0.86	**0.98**
VIFF	0.68	0.56	0.64	0.35	**0.74**
Q^f_{ab}	0.57	0.61	0.66	0.47	**0.69**
Huntington's disease [MRI-SPECT pair-slice 11]					
FMI	0.39	**0.87**	**0.87**	0.84	**0.87**
SSIM	0.79	0.86	**0.94**	0.86	0.92
VIFF	0.54	0.60	0.69	0.46	**0.80**
Q^f_{ab}	0.55	0.62	**0.67**	0.53	0.66
Huntington's disease [MRI-SPECT pair-slice 12]					
FMI	0.37	0.87	**0.88**	0.85	**0.88**
SSIM	0.76	0.85	**0.93**	0.86	0.92
VIFF	0.60	0.60	0.69	0.48	**0.80**
Q^f_{ab}	0.57	0.62	**0.67**	0.54	0.64
Glioma disease [MRI-SPECT pair-slice 36]					
FMI	0.70	0.87	**0.89**	0.83	0.88
SSIM	0.64	0.84	**0.96**	0.76	**0.96**
VIFF	0.70	0.45	0.74	0.12	**0.75**
Q^f_{ab}	0.70	0.60	**0.78**	0.36	**0.78**
Glioma disease [MRI-SPECT pair-slice 38]					
FMI	0.69	0.87	**0.89**	0.84	0.88
SSIM	0.78	0.85	**0.96**	0.76	0.95
VIFF	0.67	0.46	**0.76**	0.11	0.75
Q^f_{ab}	0.68	0.60	**0.79**	0.32	0.78
Alzheimer [MRI-PET pair-slice-15]					
Metrics	**Ref [30]**	**Ref [16]**	**Ref [5]**	**Ref [23]**	**Our proposal**
FMI	0.87	0.68	0.88	0.86	**0.89**
SSIM	0.76	0.61	0.78	0.85	**0.92**
VIFF	0.35	0.45	0.37	0.33	**0.61**
Q^f_{ab}	0.52	0.54	0.56	0.43	**0.60**

Table 8 presents the objective metrics of the MRI-PET pair and MRI-SPECT pairs of different brain tumour diseases taken into consideration. The suggested approach outperforms the others in terms of VIFF, SSIM, Qab/f, and FMI. The most valuable and edge information was transformed into the final output by our method, as evidenced by the highest FMI and Qab/f values. The highest SSIM and VIFF values also demonstrate the superiority of the suggested approach when it comes to structural similarity.

In the case of MRI-PET pair of Alzheimer's disease-slice 15, our proposed method has outperformed well with 1% increase in FMI, 82% increase in SSIM, 35% increase in VIFF and 7% increase in Qab/f with reference to the second highest value achieved by existing methods and in the case of Alzheimer's disease-slice 16, 1% increase in FMI, 11% increase in SSIM, 17% increase in VIFF and 8% increase in Qab/f respectively.

In the case of metastatic bronchogenic disease, CNP based fusion model has 1% increase in FMI compared to proposed method. Our suggested method performed well in terms of visual quality, edge retention, and structural similarity index, indicating that the image was reconstructed similarly to the original source images. In the case of Huntington's disease-slice 12, CNP-based fusion model has 1% increase in SSIM and 4% increase in Qab/f compared to proposed method. Our suggested approach performed better in terms of visual quality and information retention, and the CNP-based fusion model demonstrated a 2% increase in SSIM and a 1% rise in Qab/f in the case of slice 11.

In the case of MRI-SPECT pair of glioma slice-36, our proposed method outperformed well with respect to all metrics but second highest value in the case of FMI with marginal decrease of 1% with respect to existing CNP model which has exhibited highest values. In the case of MRI-SPECT pair of glioma slice-38, our proposed method achieved second highest value with the difference of 0.001 in the case of FMI, SSIM, VIFF and Qab/f with reference with CNP model.

The effective region of the MRI is the intracerebral tissue with the better resolution. PET images highlight the region where the blood flow is reduced. The figures exhibit the subjective assessment of proposed method with the existing methods. The proposed technique accurately recognizes and combines the effective regions of the two source images, owing to the network's learning during the training phase. Tumour is clearly delineated and the metabolic activities are clearly correlated. Averagely, our proposed method performed well in improving the structural similarity index and visual information fidelity for all diseases taken into consideration compared to existing methods. This implements that our proposed method has improved the visual quality of our fused images and edges are retained effectively thereby helping the radiologists to quick detection of tumour affected portions which in turn will result in quicker diagnosis.

Even though the objective assessments of the proposed method are as good as with the existing methods, our method has improved the subjective assessments of the fused images which will assist the radiologists or doctors to locate disease affected portions effectively and plan treatment accordingly thereby resulting in quicker diagnosis of diseases. For instance, in the case of Alzheimer's disease, highlighted blue portions clearly indicates the decrease in metabolic activities of cells in the parietal cortex and in our proposed method it is clearly delineated.

5. Computation Costs Analyses

The number of times the network is used, which is equal to the total number of frequency subbands in the algorithm, determines the computational complexity of our proposed method as well as PCNN and CNN-based algorithms.

Table 9. Computational complexity of proposed and existing methods.

	Gaussian-CNN [5]	Hahn-PCNN-CNN [23]	CNP-MIF [12]	GFF [3]	Our proposal
Number of mathematical computations involved	25, 493, 504	3, 101, 443	48, 889, 856	2, 621, 440	1, 530, 626

It also requires the loading of multiple network weights and parameters. The number of multiplications involved for existing and our proposed method are shown in Table 9 and our suggested method outperformed well both subjectively and objectively with less number of computations 3 depicts the detailed architectures of encoder-decoder parts of our proposed method.

6. Conclusions

Our approach effectively preserves anatomical and functional image textures, colors, and contrast during fusion, according to current studies. Our model is significantly good with the other four algorithms across four representative evaluation metrics. Our algorithm's robustness is enhanced by the diversity of images used. Because of its high quality and low weight, our algorithm has several potential uses in intelligent medicine. The outcomes of the experiment show that the suggested method faithfully captures significant features in the original photographs with good brightness and contrast. With fewer artifacts, precise details are recovered while texture and edge information are maintained. According to a comparative analysis, the method has good objective indicators. Additional picture data sets can be used to improve the model's performance. This work will be expanded further with real-time hardware implementation and test cases.

Author Contributions

All authors contributed to the study conception and design. In the current study, formulation of concept, conduction of literature review, examination of the work and the manuscript preparation were carried out by V. Saravanan and review of technical aspect and providing insights into formulation and results and manuscript preparation were performed by S Malarvizhi.

Data Availability

The dataset available in Whole Brain Atlas database (http://www.med.harvard.edu/aanlib/home.html).

References

[1] Ahamed B., Baskar R., and Nalinipriya G., "Enhanced Brain Tumor MRI Scan Reconstruction via the EI-Fusion-Net Model," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 4, pp. 704-7013, 2024. DOI: 10.22266/ijies2024.0831.53

[2] Azarang A., Manoochehri H., and Kehtarnavaz N., "Convolutional Autoencoder-based Multispectral Image Fusion," *IEEE Access*, vol. 7, pp. 35673-35683, 2019. DOI: 10.1109/ACCESS.2019.2905511

[3] Bavirisetti D., Kollu V., Gang X., and Dhuli R., "Fusion of MRI and CT Images Using Guided Image Filter and Image Statistics," *International Journal Imaging Systems and Technology*, vol. 27, no. 3, pp. 227-237, 2017.

https://doi.org/10.1002/ima.22228

[4] Cheng C., Xu T., and Wu X., "MUFusion: A General Unsupervised Image Fusion Network Based on Memory Unit," *Information Fusion*, vol. 92, pp. 80-92, 2023. https://doi.org/10.1016/j.inffus.2022.11.010

[5] Guo K., Li X., Hu X., Liu J., and Fan T., "Hahn-PCNN-CNN: An End-to-End Multi-Modal Brain Medical Image Fusion Framework Useful for Clinical Diagnosis," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1-22, 2021. https://doi.org/10.1186/s12880-021-00642-z

[6] Haghighat M., Aghagolzadeh A., and Seyedarabi H., "A Non-Reference Image Fusion Metric Based on Mutual Information of Image Features," *Computers and Electrical Engineering*, vol. 37, no. 5, pp. 744-756, 2011. https://doi.org/10.1016/j.compeleceng.2011.07.012

[7] Han Y., Cai Y., Cao Y., and Xu X., "A New Image Fusion Performance Metric Based on Visual Information Fidelity," *Information Fusion*, vol. 14, no. 2, pp. 127-135, 2013. https://doi.org/10.1016/j.inffus.2011.08.002

[8] Ibrahim S., Makhlouf M., and El-Tawel G., "Multimodal Medical Image Fusion Algorithm Based on Pulse Coupled Neural Networks and Nonsubsampled Contourlet Transform," *Medical and Biological Engineering and Computing*, vol. 61, no. 1, pp. 155-177, 2023. https://doi.org/10.1007/s11517-022-02697-8

[9] Johnson K. and Becker J., The Whole Brain Atlas, https://www.med.harvard.edu/aanlib/, Last Visited, 2025.

[10] Kumari B., Nandal A., and Dhaka A., "Breast Tumor Detection Using Multi-Feature Block Based Neural Network by Fusion of CT and MRI Images," *Computational Intelligence*, vol. 40, no. 3, pp. e12652, 2024. https://doi.org/10.1111/coin.12652

[11] Li B. and Lima D., "International Journal of Cognitive Computing in Engineering Facial Expression Recognition via ResNet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 57-64, 2021. https://doi.org/10.1016/j.ijcce.2021.02.002

[12] Li B., Peng H., Luo X., Wang J., Song X., Perez-Jimenez M., and Riscos-Nunez A., "Medical Image Fusion Method Based on Coupled Neural P Systems in Nonsubsampled Shearlet Transform Domain," *International Journal of Neural Systems*, vol. 31, no. 1, pp. 1-17, 2021. https://doi.org/10.1142/S0129065720500501

[13] Li S., Kang X., and Hu J., "Image Fusion with Guided Filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864-2875, 2013. DOI: 10.1109/TIP.2013.2244222

[14] Li Y., Zhao J., Lv Z., and Pan Z., "Multimodal Medical Supervised Image Fusion Method by CNN," *Frontiers in Neuroscience*, vol. 15, pp. 1-10, 2021. https://doi.org/10.3389/fnins.2021.638976

[15] Liu Y., Chen X., Cheng J., and Peng H., "A Medical Image Fusion Method Based on Convolutional Neural Networks," *in Proceedings of the 20th International Conference on Information Fusion*, Xi'an, pp. 1-7, 2017. DOI: 10.23919/ICIF.2017.8009769

[16] Liu Y., Chen X., Peng H., and Wang Z., "Multi-Focus Image Fusion with a Deep Convolutional Neural Network," *Information Fusion*, vol. 36, pp. 191-207, 2017. https://doi.org/10.1016/j.inffus.2016.12.001

[17] Luo X., Li X., Wang P., Qi S., Guan J., and Zhang Z., "Infrared and Visible Image Fusion Based on NSCT and Stacked Sparse Autoencoders," *Multimedia Tools and Applications*, vol. 77, no. 17, pp. 22407-22431, 2018. https://doi.org/10.1007/s11042-018-5985-6

[18] Munawwar S. and Rao P., "An Efficient Deep Learning Based Multi-Level Feature Extraction Network for Multi-Modal Medical Image Fusion," *The International Arab Journal of Information and Technology*, vol. 22, no. 3, pp. 429-447, 2025. https://doi.org/10.34028/iajit/22/3/2

[19] Panigrahy C., Seal A., and Mahato N., "MRI and SPECT Image Fusion Using a Weighted Parameter Adaptive Dual Channel PCNN," *IEEE Signal Processing Letters*, vol. 27, pp. 690-694, 2020. DOI: 10.1109/LSP.2020.2989054

[20] Piella G. and Heijmans H., "A New Quality Metric for Image Fusion" *in Proceedings of the International Conference on Image Processing*, Barcelona, pp. 173-176, 2003. DOI: 10.1109/ICIP.2003.1247209

[21] Saleh M., Ali A., Ahmed K., and Sarhan A., "A Brief Analysis of Multimodal Medical Image Fusion Techniques," *Electron*, vol. 12, no. 1, pp. 1-30, 2023. DOI: 10.3390/electronics12010097

[22] Sebastian J. and King G., "A Novel MRI and PET Image Fusion in the NSST Domain Using YUV Color Space Based on Convolutional Neural Networks," *Wireless Personal Communications*, vol. 131, no. 3, pp. 2295-2309, 2023. https://doi.org/10.1007/s11277-023-10542-w

[23] Vajpayee P., Panigrahy C., and Kumar A., "Medical Image Fusion by Adaptive Gaussian PCNN and Improved Roberts Operator," *Signal, Image and Video Processing*, vol. 17, no. 7, pp. 3565-3573, 2023. https://doi.org/10.1007/s11760-023-02581-4

[24] Wang K., Zheng M., Wei H., Qi G., and Li Y., "Multi-Modality Medical Image Fusion Using Convolutional Neural Network and Contrast Pyramid," *Sensors*, vol. 20, no. 8, pp. 1-17, 2020. DOI: 10.3390/s20082169

[25] Wei Y. and Ji L., "Multi-Modal Bilinear Fusion with Hybrid Attention Mechanism for Multi-Label Skin Lesion Classification," *Multimedia Tools and Applications*, vol. 83, pp. 65221-65247, 2024. https://doi.org/10.1007/s11042-023-18027-5

[26] Xu H. and Ma J., "EMFusion: An Unsupervised Enhanced Medical Image Fusion Network," *Information Fusion*, vol. 76, pp. 177-186, 2021. https://doi.org/10.1016/j.inffus.2021.06.001

[27] Zhang Y., Liu Y., Sun P., Yan H., Zhao X., and Zhang L., "IFCNN: A General Image Fusion Framework Based on Convolutional Neural Network," *Information Fusion*, vol. 54, pp. 99-118, 2020. https://doi.org/10.1016/j.inffus.2019.07.011

[28] Zhang Y., Nie R., Cao J., and Ma C., "Self-Supervised Fusion for Multi-Modal Medical Images via Contrastive Auto-Encoding and Convolutional Information Exchange," *IEEE Computational Intelligence Magazine*, vol. 18, no. 1, pp. 68-80, 2023. DOI: 10.1109/MCI.2022.3223487

[29] Zhao H., Cai H., and Liu M., "Transformer Based Multi-Modal MRI Fusion for Prediction of Post-Menstrual Age and Neonatal Brain Development Analysis," *Medical Image Analysis*, vol. 94, pp. 103140, 2024. https://doi.org/10.1016/j.media.2024.103140

[30] Zhou T., Cheng Q., Lu H., Li Q., Zhang X., and Qiu S., "Deep Learning Methods for Medical Image Fusion: A Review," *Computers in Biology and Medicine*, vol. 160, pp. 106959, 2023. DOI: 10.1016/j.compbiomed.2023.106959

Saravanan Vijayan is a research scholar doing his research in the area of medical image processing in the Department of Electronics and Communication Engineering in SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Chennai-603203, TamilNadu, India. He received his BE degree in Electronics and Instrumentation Engineering from Madras University, Chennai, India in 2004, the M.Tech Degree in Digital Communication and Networking from SRM University, Kattankulathur, Chennai, India in 2010 and pursuing his PhD in Dept. of ECE at SRM Institute of Science and Technology, Kattankulathur, Chennai. His area of research includes Medical Image Processing and Deep Learning.

Malarvizhi Subramani is working as a Professor in the Department of Electronics and Communication Engineering in SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Chennai-603203, TamilNadu, India. She received the BE degree in Electronics and Communication Engineering from Madras University, Chennai, India, in 1989, the M.Tech degree in Applied Electronics from the Government College of Technology, Coimbatore, India, in 1991, and the Ph.D. degree in Wireless Communication from Anna University, Chennai, under Faculty of Information Communication Engineering in 2006. In 1992, she joined SRM Engineering College, Kattankulathur, Chennai, India, as a Lecturer. She was with Pondicherry Engineering College in 1999. Since 2005, she has been a professor in the Department of ECE, SRM Institute of Science and Technology, Kattankulathur, Chennai.