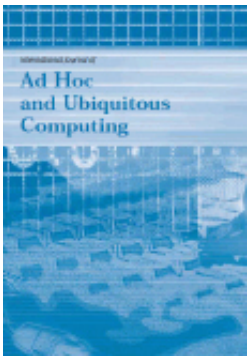


International Journal of Ad Hoc and Ubiquitous Computing

ISSN online: 1743-8233 - ISSN print: 1743-8225
https://www.inderscience.com/ijahuc

Multimodal fusion of different medical image modalities using optimised hybrid network

Tanima Ghosh, N. Jayanthi

DOI: 10.1504/IJAHUC.2024.10067151

Article History:

Received:	14 October 2023
Last revised:	17 October 2023
Accepted:	20 February 2024
Published online:	30 December 2024

Multimodal fusion of different medical image modalities using optimised hybrid network

Tanima Ghosh* and N. Jayanthi

Department of Electronics and Communication Engineering,
Delhi Technological University,
Bawana Road, Delhi-42, India
Email: tanima284691@gmail.com
Email: njayanthi@dce.ac.in
*Corresponding author

Abstract: Image fusion leverages the strengths of various imaging modalities to create a more complete and informative picture of medical conditions, which leads to better identification and treatment. Accordingly, this paper implements a new multimodel image fusion approach, named pelican optimisation algorithm-based DenseNet and ResidualNet (POA+Dense-ResNet) for multimodel image fusion. Here, the POA is used to train the Dense-ResNet, which is the combination of ResidualNet and DenseNet. The input images from different modalities are pre-processed and then the transformation of the spatial domain to the spectral domain is done by dual-tree complex wavelet transform (DTCWT). These transformed images are segmented by edge-attention guidance network (ET-Net). Then, the fusion is done by the POA+Dense-ResNet. The POA+Dense-ResNet achieved minimum root mean square error (RMSE), mean square error (MSE), and maximum peak signal to noise ratio (PSNR) of 0.650, 0.423, and 53.525 dB.

Keywords: ResidualNet; edge-attention guidance network; pelican optimisation algorithm; DenseNet; dual-tree complex wavelet transform; mean square error; MSE; peak signal to noise ratio; PSNR.

Reference to this paper should be made as follows: Ghosh, T. and Jayanthi, N. (2025) 'Multimodal fusion of different medical image modalities using optimised hybrid network', *Int. J. Ad Hoc and Ubiquitous Computing*, Vol. 48, No. 1, pp.19–33.

Biographical notes: Tanima Ghosh is a research scholar in the Department of Electronics and Communication Engg at Delhi Technological University (DTU) in Delhi-110042. She is working as an Assistant Professor in the Department of Electronics and Communication Engineering at Bhagwan Parshuram Institute of Technology (Affiliated to GGSIPU) Delhi, since 2011. She has done her BTech in Electronics and Communication Engineering from, Delhi College of Engineering (currently named as DTU) in Delhi. She completed her MTech from Indira Gandhi Delhi Technological University for Women (IGDTUW), Delhi. She qualified GATE exam in 2008. She has many journal and conference paper publications.

N. Jayanthi is an Assistant Professor in the Department of Electronics and Communication Engineering at Delhi Technological University, since 2007. She has done her BE in Electronics and Communication Engineering from Bharathidasan University, Trichy, and MTech in Communication Systems from National Institute of Technology, Trichy. She has completed her PhD at Delhi Technological University. Her area of research interest includes image processing, pattern recognition, machine learning and communication systems. She published many technical papers in international conferences and journals. She is a member in professional societies such as IEEE and Intelligent Transport System (ITS) and Pattern Recognition and Artificial Intelligence (IUPARI).

1 Introduction

Recently, medical imaging has become an indispensable tool for diagnosing a wide range of diseases (Kaur and Singh, 2021). Despite significant advancements in imaging technologies, medical experts often face challenges when relying on a single imaging modality due to the limitations in the amount and type of information each modality can provide. Image fusion has emerged as a promising solution to address these challenges by combining images from different modalities, thereby integrating anatomical and physiological data to facilitate more accurate and comprehensive diagnosis (Wang et al., 2021). By merging images from modalities such as computerised tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and ultrasound, image fusion provides a more holistic view of the organ. This integration helps in correlating structural details with functional information, enhancing the accuracy of diagnosis

(Nandhini Abirami et al., 2022). Fused images offer improved visualisation of disease characteristics, such as tumor boundaries, tissue composition, and metabolic activity, which aids in more precise detection and characterisation of diseases (Almasri and Alajlan, 2022). This kind of fusion is a transformative scheme in medical imaging, aimed at preserving and integrating complementary details from various modalities into a single, comprehensive image. This fusion enhances diagnostic accuracy, which aids in better treatment planning, and ultimately improves patient care (Ding et al., 2023).

Image fusion techniques are pivotal in enhancing the quality and information of images across numerous domains, like remote sensing and surveillance to medical imaging and machine learning. The traditional and deep learning-based schemes have their respective strengths and applications in medical image fusion (Azam et al., 2022; Nagaraja Kumar et al., 2023a). Traditional methods are useful for their simplicity and interpretability, while deep learning-based methods offer superior performance and adaptability at the cost of increased complexity and resource requirements (Ding et al., 2023). Extracting complementary features from source images is a crucial step in the fusion of images, ensuring that the fused result contains the most informative and comprehensive data from each modality (Guo et al., 2023; Nagaraja Kumar et al., 2023b). Thus, these methods are useful for medical professionals to achieve high-quality fused images (Zhang et al., 2023).

Traditional-based medical image fusion methods have indeed demonstrated excellent performance in various applications. However, one of the significant challenges associated with these methods is the manual design of complex activity level measurements, which can be time-consuming and labour-intensive (Wei et al., 2023). Complex activity level measurements can significantly increase the computational burden, making the fusion process less efficient, especially with high-resolution medical images (Huang et al., 2020; Meher et al., 2019). Addressing these challenges through automated and adaptive methods, hybrid techniques, and advanced parameter optimisation algorithms can significantly enhance the efficiency, which leading to better diagnostic and treatment outcomes (Li et al., 2012). Utilising machine learning techniques to automate the design of activity-level measurements can significantly reduce the need for manual parameter tuning. These methods can learn optimal parameters directly from data. The deep learning approaches, particularly convolutional neural networks (CNNs), can automatically learn to extract relevant features and determine activity levels, adapting to different types of medical images and fusion tasks.

This paper aims to develop POA+Dense-ResNet for multimodel image fusion in different modalities. The proposed method leverages pre-processing, spectral domain transformation using DTCWT, segmentation with ET-Net, and fusion using the POA (Trojovský and Dehghani, 2022) and Dense-ResNet to produce high-quality fused images. The combination of optimisation algorithms, and deep learning architectures ensures that the effectiveness of the proposed method.

The contribution of this research is:

- *POA+Dense-ResNet:* the POA+Dense-ResNet is developed in this research for an effective multimodel image fusion from different modalities. Here, the Dense-ResNet is implemented by the combination of ResidualNet and DenseNet, and the POA (Trojovský and Dehghani, 2022) is used to train the Dense-ResNet.

The balance sections of this research article are provided as follows: the comprehensive details of existing schemes are given in Section 2, the details of the newly proposed POA+Dense-ResNet are provided in Section 3, the outcomes of experiments conducted and the discussion of these results are depicted in Section 4, and the research findings and directions for future research are elaborated in Section 5.

2 Motivation

The goal of image fusion is to combine strengths from different imaging modalities to create more comprehensive and informative images. The common issues in similar existing schemes are: The differences in resolution, scale, and orientation pose significant challenges. Also, each imaging modality may introduce specific artefacts that can interfere with the fusion process. Effective techniques are needed to manage and minimise these artefacts to maintain the quality of the fused image. Thus, to overcome these problems an effective multimodel image fusion approach is developed in this research.

2.1 Related work

Kaur and Singh (2021) established a model, in which the deep neural network was trained using multi-objective differential evolution to fuse medical multi-modal images. Here, the network architecture and parameters were optimised to increase the quality of the fused image according to predefined criteria. By combining complementary information from multiple modalities, it provided more accurate and comprehensive diagnostic information. However, this model required high-performance systems for computation and significant training times. Wang et al. (2021) implemented a Gabor representation of multi-CNN combination (G-CNN) and fuzzy neural network for fusing medical multimodal images. Here, the Gabor filters extracted texture and spatial frequency features. Also, multiple CNNs were employed to independently process the Gabor representations from various modalities, extracting and learning hierarchical features. The outputs from the multi-CNN stage were combined using a fuzzy neural network. This approach addressed the challenges associated with integrating information from diverse imaging modalities, potentially advancing medical image analysis capabilities. However,

the training and inference times were prolonged, particularly when processing large volumes of high-resolution medical images.

Nandhini Abirami et al., (2022) established a generative adversarial network (GAN)-based approach for fusing PET and MRI images. In this model, the specific loss functions were utilised to train the GAN, which are crucial for ensuring the accuracy of the fused images. However, it was challenging due to the differences in resolution, noise characteristics, and modality-specific artefacts of the PET and MRI images. Almasri and Alajlan (2022) implemented a modified discrete wavelet transform (MDWT) for the fusion of multimodel images. The classification was done by convolutional neural network-based hybrid optimisation dynamic algorithm (CNN-HOD). This approach enhanced the quality, accuracy, and usefulness of medical images for diagnostic purposes, which potentially improved medical decision-making and patient outcomes. However, the ability to generalise to new datasets or different medical conditions was limited.

Ding et al., (2023) established a specific neural network architecture called M4FNet for medical image fusion. Here, this network used filters or convolutional layers with various sizes to capture information at different spatial scales. Also, combining features extracted at different scales ensured that both fine-grained information and broader context were preserved in the fused image. However, this model was not generalised well on new variations in imaging conditions. Zhang et al. (2023) established a pair feature difference guided network (FDGNet) to perform medical multimodal image fusion. The differences between feature pairs were extracted from the input images and the feature difference loss and weighted fidelity loss were utilised for the network training. It potentially improved the accuracy and utility of fused medical images. However, the insufficient training data led to poor performance and generalisation issues.

Li et al. (2023) implemented a model, named DFENet by integrating transformers and CNNs to scale up the process of medical multimodal image fusion. Here, a dual-branch architecture typically consists of two parallel pathways or branches that process different aspects of the input data. It achieved better performance in terms of accuracy, robustness, and clinical utility. However, it required a large amount of images, which limited the practical application of the model in real-world scenarios. Ye et al. (2023) established an unsupervised foveated differentiable architecture search (F-DARTS) for the multimodel image fusion. It incorporated foveated mechanisms to prioritise different regions of interest (ROIs) within the images, optimising resource allocation and enhancing fusion quality. It effectively handled the complex and varied information present in multimodal medical images. However, the small perturbations in input data could potentially affect fusion quality and diagnostic outcomes.

2.2 Challenges

The challenges of the already implemented techniques are discussed below:

- Medical images can contain noise and artefacts due to various reasons such as patient motion, scanner imperfections, or contrast agent issues. Some models were unable to handle such noise and artefacts without compromising the quality of the fused images (Kaur and Singh, 2021).
- Image fusion involves complex algorithms and significant computational resources, necessitating robust hardware and software solutions to process and integrate the images efficiently (Ding et al., 2023).
- Developing standardised metrics and methods for evaluating the quality of fused images is essential. However, assessing factors such as spatial resolution, contrast, and the preservation of important diagnostic information from the original images was difficult (Li et al., 2023).
- Labelled datasets for multimodal medical images are often limited and can be expensive to acquire due to the need for expert annotations (Ye et al., 2023).

3 Implemented POA-Dense-ResNet to fuse multimodel images

Multimodal image fusion is a powerful tool that enhances the information content and quality of images, making it invaluable in various fields requiring detailed analysis and decision-making. In this research, POA+Dense-ResNet is developed for multimodel image fusion. Here, the combination of ResidualNet and DenseNet, called Dense-ResNet, which is trained using the POA (Trojovský and Dehghani, 2022). The DTCWT is used to perform the spatial domain to spectral domain transformation after the input images from various modalities have undergone pre-processing. The ET-Net segments these transformed images. The POA+Dense-ResNet achieve the fusion in the last stage. Figure 1 shows the block diagram of the POA+Dense-ResNet.

3.1 Image acquisition

The images are obtained from three different modalities for multimodel image fusion and it is given as:

$$N = \{N_1, N_2, N_3, ...N_A, ..., N_Z\} \tag{1}$$

where N denotes the database considered for this research, A^{th} T1 image in the database is denoted as N_A and entire images in the database expressed as z.

Figure 1 Block diagram of the POA+Dense-ResNet (see online version for colours)

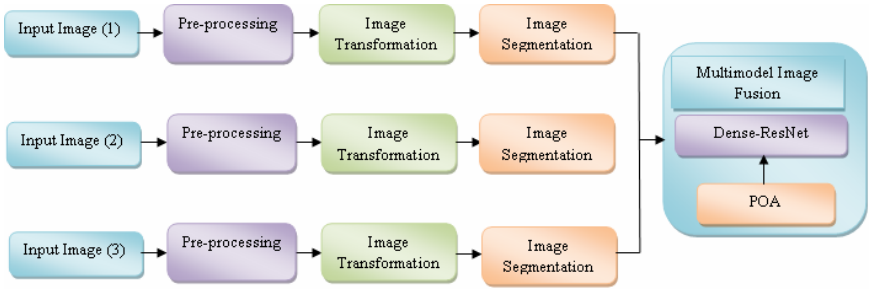


Figure 2 Structure of ET-Net (see online version for colours)

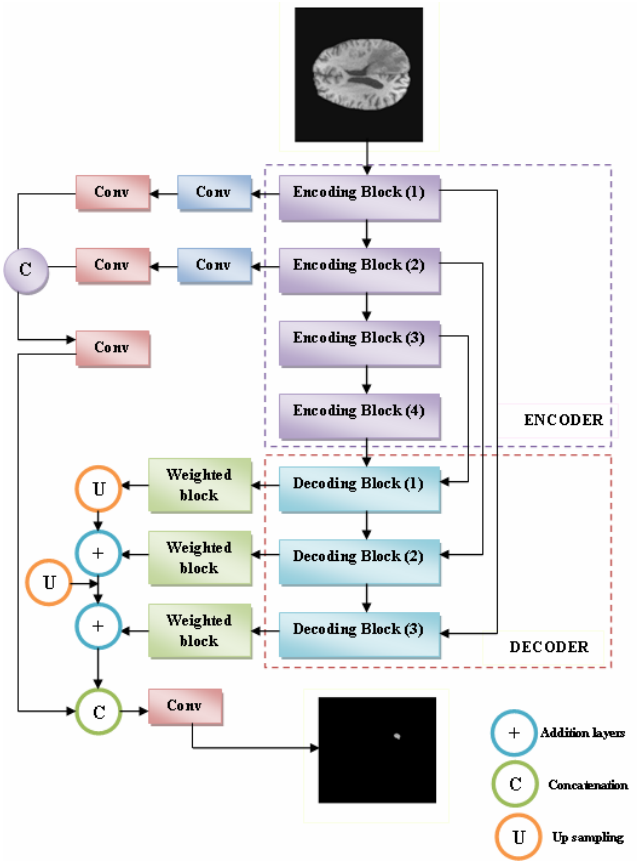


Figure 3 Structural model of hybrid Dense-ResNet (see online version for colours)

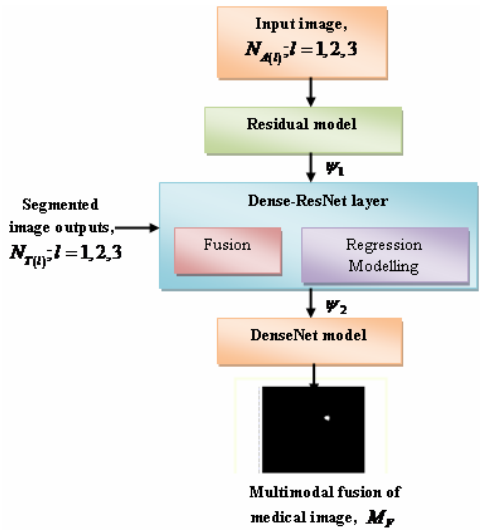


Figure 4 Residual model structure (see online version for colours)

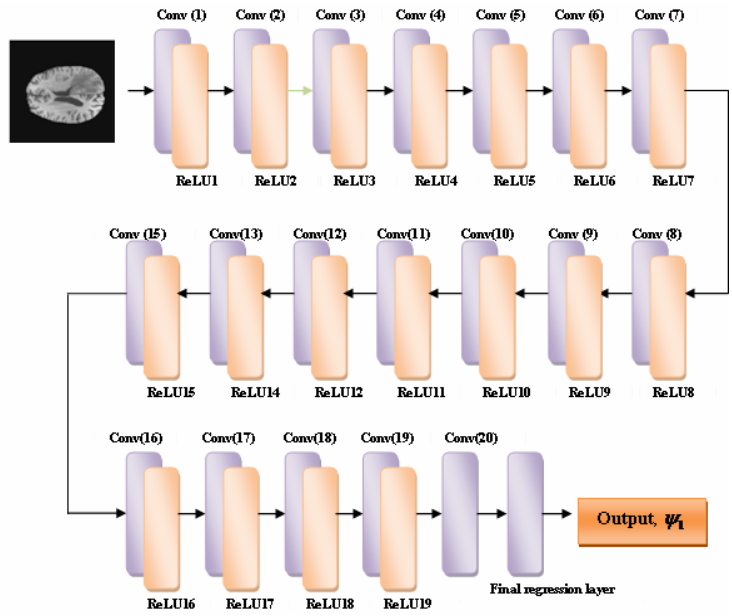
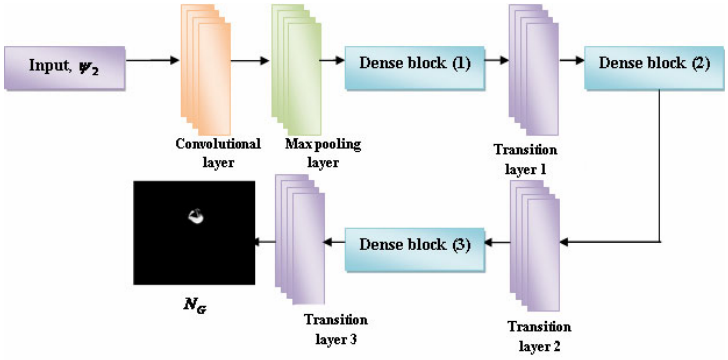


Figure 5 Structure of DenseNet (see online version for colours)



3.2 Pre-processing

In this stage, noise removal is a crucial step to ensure that the subsequent image fusion process works with clean and high-quality data. This paper used a median filter (Maheshan and Prasanna Kumar 2020) for de-noising the T1 input image N_A, which is a common practice in image processing due to its effectiveness in preserving edge information while eliminating noise. It is a nonlinear filtering technique, which replaces each pixel's value with the median value of the pixel's neighbourhood. Unlike linear filters, the median filter does not blur edges. This is critical in medical imaging, where retaining edge information. The formula for applying this filter is provided as:

$$N_Q = \underset{A \in S_Q}{Medi}(N_A) \tag{2}$$

where S_Q denotes the neighbourhood sub image and N_Q is the pre-processed outcome.

3.3 Transformation of image

In the image fusion approach discussed, transforming the pre-processed T1 image N_Q from the spatial domain to the spectral domain is a crucial step. This transformation is achieved using the DTCWT (Slimen et al., 2020), which offers significant improvements over the traditional discrete wavelet transform (DWT). DTCWT is an advanced version of DWT, providing better directional selectivity and shift-invariance. Unlike DWT, DTCWT is nearly shift-invariant, meaning the transformed image does not change significantly with small shifts in the input image. This is crucial for maintaining consistency in feature extraction and it can capture features in multiple directions, which is beneficial for analysing complex structures in medical images.

$$\kappa(m) = \kappa_b(m) + k\kappa_c(m) \tag{3}$$

$$\chi(m) = \chi_b(m) + k\chi_c(m) \tag{4}$$

where the term for complex imaginary is denoted as k, complex wavelet is presented as $\kappa(m)$, the function for complex scaling is denoted as $\chi(m)$, and b and c represents

the top and bottom tree output coefficient. The transformed outcome is indicated as N_U and it is given to the next stage.

3.4 Segmentation

In the proposed multimodal image fusion approach, the image segmentation phase is essential for accurately extracting ROIs from the transformed T1 images. This is achieved using the ET-Net (Zhang et al., 2019), which is specifically designed for precise medical image segmentation. ET-Net is used to segment the transformed images by identifying and extracting primary regions while ignoring edge information that may introduce noise or irrelevant details. The network uses edge attention mechanisms to enhance the accuracy of the segmentation. By focusing on the primary regions and managing edge information effectively, ET-Net ensures that the segmentation is precise.

3.4.1 Architecture of ET-Net

The ET-Net (Zhang et al., 2019) is an advanced encoder-decoder network that incorporates specific modules designed to enhance the segmentation process by effectively managing edge information and multi-scale feature aggregation. It contains two key components, such as weighted aggregation module (WAM) and edge guidance module (EGM). The EGM preserves the characteristics of local edges during the encoding process, which ensures that important edge information is maintained. The WAM aggregates multi-scale side outputs from the network, which enhances the ability to integrate features at various scales. Here, the encoder network has ResNet-50, which contains four encoding blocks to mine features from the input image. The decoder network consists of three decoding blocks to form the decoder path. It is used in decoding blocks to enhance the representation of both high and low-level features. The EGM ensures that critical edge information is not lost during encoding, which is vital for accurate segmentation. Figure 2 illustrates the organisation of ET-Net, which shows how the transformed T1 image is processed through the network to produce a segmented output N_T.

Here, the segmented outcome of T1 image is denoted as $N_{U(1)}$. Likewise, the pre-processing, transformation, and segmentation of the other two modalities of medical images, like T2, and T1Gd are done and the corresponding outcomes are denoted as $N_{U(2)}$ and $N_{U(3)}$. The segmented outcomes of different modalities of images are permitted for the Dense-ResNet scheme.

3.5 Dense-ResNet scheme for medical image fusion

The proposed scheme combines different or the same modalities into a single, clearer, and more informative image. This process enhances the quality of the resultant image. The segmented outcomes of the T1 image $N_{U(1)}$, T1Gd image $N_{U(2)}$, and T2 image $N_{U(3)}$ are permitted to do the multimodal image fusion by the hybrid Dense-ResNet approach. This is the merging of DenseNet (Vulli et al., 2022) and ResNet (Chen et al., 2019). It is formulated using fractional calculus (FC) (Bhaladhare and Jinwala, 2014), which provides a mathematical framework for modelling and controlling dynamic systems with memory and hereditary properties. The Dense-ResNet contains three modules, such as residual, DenseNet, and Dense-ResNet layer. The input medical image is denoted as $\{N_{A(l)}\}$; $l = 1, 2, 3$ with $(r \times s)$ dimension, which is given as the input to the residual model, which contains ResNet. Then, the image dimension is changed as $(r \times s) \times l$, and l is the entire input image. Then, the segmented outcome $\{N_{T(l)}\}$; $l = 1, 2, 3$ and the ResNet outcome are given to the Dense-ResNet layer for regression and fusion. The combined features from the Dense-ResNet model are then passed through a DenseNet model. This final step in the fusion process ensures that the fused image retains high clarity and enhanced content by making use of DenseNet's efficient feature propagation and layer connectivity. Figure 3 denotes the structure of the Dense-ResNet approach.

3.5.1 Residual model

ResNet (Chen et al., 2019) is a widely used deep learning architecture known for its ability to train very deep networks without encountering the vanishing gradient problem. This section provides a detailed breakdown of the components of ResNet, which collectively ensure effective feature extraction, gradient flow, and high training speed, leading to high-quality, informative fused images.

- *Convolutional layer:* it is a fundamental component of CNNs, including ResNet. It plays a crucial role in reducing the number of parameters, increasing the performance by weight sharing, and focusing on local receptive fields. The detailed computation process, involving filters and dot products, enables the network to learn and extract meaningful features from the input images.

$$\psi_1 = \sum_{s=0}^{x-1} \sum_{t=0}^{x-1} O_{f,t} * N_{A(l)(h+s),(i+t)} \tag{5}$$

  where $N_{A(l);\, l=1,2,3}$ conveys the input image, h and i are the coordinates, O is the kernel matrix with $x \times x$ dimension, s and t are the kernel matrix index location and * is the operator utilised in cross-correlation without zero padding.

- *Average pooling layer:* the pooling layer in ResNet significantly contributes to the network's efficiency by reducing the spatial size of feature maps and controlling overfitting. By summarising the presence of features and preserving important spatial information, pooling layers play a critical role in extracting meaningful representations. The formula for this layer is,

$$C_{out} = \frac{C_{in} - u_s}{e} + 1 \tag{6}$$

$$I_{out} = \frac{I_{in} - u_t}{e} + 1 \tag{7}$$

where two-dimensional input matrix height and width are I_{in} and C_{in} and the outputs are C_{out} and I_{out}. Also, u_s is the kernel width and u_v is the kernel height.

- *Activation function:* rectified linear unit (ReLU) activation function is a fundamental component in neural networks that enables effective learning of nonlinear relationships in data. Its simplicity, efficiency in computation, and ability to mitigate gradient vanishing issues make it a preferred choice in modern deep-learning models. The formula is,

$$ReLU(\psi_1) = \begin{cases} 0 & N_{A(l)} < 0 \\ N_{A(l)} & N_{A(l)} \geq 0 \end{cases} \tag{8}$$

- *Batch normalisation:* batch normalisation is a powerful technique in neural networks that normalises inputs within mini-batches, stabilising and accelerating the training process. By reducing internal covariate shifts, adjusting activations, and improving training speed and stability, batch normalisation significantly enhances performance.
- *Residual blocks:* residual blocks with shortcut connections are fundamental to modern deep learning architectures like ResNet. By facilitating gradient flow and enabling the training of deep networks, they have revolutionised the field by allowing for the development of highly accurate and scalable models capable of solving complex tasks across various domains. The formula is given as:

$$Y = G(N_A) + N_A \tag{9}$$

The formula for the matching dimension factor is,

$$Y = G(N_A) + \delta N_A \tag{10}$$

where the residual block is given as Y, the mapping function is G, and δ is the matching factor.

- *Linear classifier:* the linear classifier, implemented through fully connected layers and often coupled with softmax activation, is a fundamental component in neural networks for classification tasks. The dot product is used in the evaluation of a fully connected layer and the formula is,

$$Y = \delta_{v \times w} N_{A(q \times r)} + z_{v \times w} \tag{11}$$

With the help of the Softmax function,

$$\varsigma = \frac{f^{\varsigma_m}}{\sum_{u=1}^{\lambda} f^{\varsigma_u}} \tag{12}$$

where $\delta_{v \times w}$ denotes the matrix weight with $v \times w$ size, $M_{A(q \times r)}$ denotes the input with dimension $q \times r$, z is the bias, any single component of the output layer is denotes as ς_m, and the output is provided as λ.

- *Regression layer:* this layer with a half-mean-squared-error loss function is vital for tasks requiring continuous value predictions. It helps normalise and stabilise the training process, ensuring the neural network learns to make accurate and reliable predictions. The output is denoted as ψ_1 and Figure 4 is the structure of residual structure.

3.5.2 Dense-ResNet layer

The outcome of the residual approach ψ_1 and the segmented output of the images $N_{t(l)}$ are given as the input of Dense-ResNet. It is a variant of the ResNet that includes dense connections. This architecture ensures that features from all layers are used, promoting better feature reuse and improving the model's ability to capture complex patterns. Also, it takes both the feature maps from the residual model and the segmented images as inputs. This combination helps in determining the relationships between the inputs more effectively. Regression modelling is performed to predict continuous values based on the features extracted by the Dense-ResNet model. In this context, FC (Bhaladhare and Jinwala, 2014) is used for regression. The FC is employed to solve various integral and derivative equations, aiding in capturing the intricate dependencies and relationships within the data.

The outcome of this layer at l^{th} interval is,

$$i_1 = \sum_{y=1}^{v} \sum_{z=1}^{w} N^1_{T(l)_{y,z}} * X_{y,z} \tag{13}$$

where the coefficient of weight is given as X, and the segmented image input $N_{t(l)}$ with y and z rows and columns.

The outcome of this layer at $(l-1)^{\text{th}}$ interval is,

$$i_2 = \sum_{y=1}^{v} \sum_{z=1}^{w} N^2_{T(l)_{y,z}} * X_{y,z} \tag{14}$$

The outcome of this layer at $(k-1)^{\text{th}}$ interval is,

$$i_3 = \sum_{y=1}^{v} \sum_{z=1}^{w} M^3_{T(l)_{y,z}} * W_{y,z} \tag{15}$$

The FC (Bhaladhare and Jinwala, 2014) concept for regression modelling is applied and the formula is,

$$\begin{aligned} z(q+1) = \rho \cdot z(q) + \frac{1}{2}\rho \cdot z(q-1) + \frac{1}{6}(1-\rho) z(q-2) \\ + \frac{1}{24}\rho(1-\rho)(2-\rho) \cdot z(q-3) \end{aligned} \tag{16}$$

The outcomes of the Dense-ResNet layer at different time intervals are applied and the expression becomes,

$$\begin{aligned} \psi_2 = \rho \cdot i_1 + \frac{1}{2}\rho \cdot .i_2 + \frac{1}{6}(1-\rho).i_3 \\ + \frac{1}{24}\rho(1-\rho)(2-\rho) \cdot \psi_1 \end{aligned} \tag{17}$$

The z_1, z_2, and z_3 values are applied and expression is given as:

$$\psi_2 = \rho \cdot \sum_{y=1}^{v}\sum_{z=1}^{w} N^1_{T(l)_{y,z}} * X_{y,z} + \frac{1}{2}\rho \cdot . \sum_{y=1}^{v}\sum_{z=1}^{w} N^2_{T(l)_{y,z}} * X_{y,z} + \frac{1}{6}(1-\rho). \sum_{y=1}^{v}\sum_{z=1}^{w} N^3_{T(l)_{y,z}} * X_{y,z} \frac{1}{24}\rho(1-\rho)(2-\rho)\cdot\psi_1 \quad (18)$$

where ρ is the order of the derivative, and the fusion outcome achieved by the Dense-ResNet layer is denoted as ψ_2.

3.5.3 DenseNet model

The DenseNet architecture is indeed a popular choice for deep learning classification tasks due to its unique design and several advantages. It addresses the vanishing gradient problem by ensuring shorter connections between layers close to the input and those close to the output. Also, it minimises the number of trainable parameters through feature reuse. Moreover, strong feature propagation is achieved by directly connecting each layer to every other layer in a feed-forward fashion. The outcome of the Dense-ResNet layer ψ_2 is given to the DenseNet approach. The key components of this model are discussed below:

- *Convolutional layer:* the convolutional layers in a DenseNet model play a crucial role in extracting features from the input images. This is the first layer that processes the input image. It typically uses a relatively large kernel size to capture broad features. The formula for this layer is given as:

$$F^k_{VW} = \sum_{D=0}^{g-1}\sum_{E=0}^{g-1} \alpha_{DE} G^k_{(V+D)(W+E)} \quad (19)$$

where, the nonlinear input is given as F^k_{VW}, V and W are the coordinates, α is the kernel matrix with $g \times g$ size, D and F are the kernel matrix index location, * is the operator. The nonlinearity expression is,

$$G^k_{VW} = \sigma\left(F^k_{VW}\right) \quad (20)$$

where σ is the previous layer parameter.

- *Max pooling layer:* it significantly minimises the spatial dimensions of the feature maps, which helps in managing the model's complexity and computational requirements. By selecting the maximum value within a filter over the feature map, max pooling retains the most prominent features. Assume the feature map with $Q \times R \times S$ dimension, which denotes the height, width, and channels of the feature map. The expression obtained after applying max pooling in feature map dimension is,

$$Maxpool = \frac{(Q-\omega+1)}{\varepsilon} \times \frac{(R-\omega+1)}{\varepsilon} \times S \quad (21)$$

Here, the filter size and stride are denoted as ω and ε.

- *Dense layer:* each layer within a dense block receives input from all preceding layers and passes on its feature maps to all subsequent layers within the same block. This dense connectivity ensures that feature maps are reused and concatenated throughout the network. Each layer performs a convolution operation which can be seen as a form of matrix manipulation. The expression is given as:

$$\beta.\sigma = \begin{matrix} w_{11} & w_{12} & \cdots & w_{1x} & \eta_1 \\ w_{21} & w_{22} & \cdots & w_{2l} & \eta_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ w_{x1} & w_{x2} & \cdots & w_{xy} & \eta_x \end{matrix} \quad (22)$$

Here, the dimension of the matrix N is $x \times z$, η is a matrix with $1 \times x$ dimension. The corresponding weights are expressed as:

$$\chi^{ky} = \chi^{ky} - \mu \times U\chi^{ky} \quad (23)$$

$$\vartheta^{ky} = \vartheta^{ky} - \mu \times U\vartheta^{ky} \quad (24)$$

where the ky layer corresponding weights are given as χ^{ky} and the corresponding bias is signified as ϑ^{ky}, and μ is the learning rate. The partial derivatives loss function is $U\chi$ and $U\vartheta$, which is determined by the chain rule. The expression for this is,

$$U\chi^{ky} = \frac{\partial m}{\partial \chi^{ky}} = \frac{1}{\beta} UM^{ky} J^{(ky-1)H} \quad (25)$$

$$U\vartheta^{ky} = \frac{\partial m}{\partial \vartheta^{ky}} = \frac{1}{\beta} \sum_{V=1}^{\beta} UM^{ky(V)} \quad (26)$$

$$UJ^{(ky-1)} = \frac{\partial m}{\partial J^{(ky-1)}} = Z^{ky} UM^{ky} \quad (27)$$

$$UM^{ky} = UJ^{ky} \times d\cdot\left(M^{ky}\right) \quad (28)$$

where M^{ky} is the linear activation at ky layer, $d\cdot\left(M^{ky}\right)$ is the nonlinear function on the basis of M^{ky}, and J^{ky} is the nonlinear activation function.

- *Transition layer:* transition layers help in reducing the dimensions (height and width) of the feature maps, which control the spatial resolution of the features. They also reduce the number of channels, which helps in managing the growth of the network size due to the dense connections. It reduces the spatial dimensions of the feature maps by half by 2×2 average pooling with a stride of 2.

The fusion outcome of this model is denoted as N_G and architecture of the DenseNet is given in Figure 5.

3.6 Training of Dense-ResNet using POA

The Dense-ResNet is trained by the POA (Trojovský and Dehghani, 2022) to improve the efficiency of Dense-ResNet. The POA is the natural hunting strategy of

pelicans to perform optimisation in a population-based framework. It uses mechanisms inspired by pelican behaviours for both local and global search, ensuring a balance between exploration and exploitation. POA has potential applications in a wide range of engineering fields, offering a novel and efficient approach to solving complex optimisation problems. Pelicans exhibit unique hunting behaviours, such as cooperative hunting and diving to catch fish. These behaviours are modelled in the algorithm to explore and exploit the search space efficiently. The steps of POA are explained below:

1 Initialisation: the POA is a population-based optimisation algorithm where the pelicans in the population represent candidate solutions to the optimisation problem. The position of each pelican in the search space corresponds to a set of values for the optimisation problem's variables. The initial positions of the pelicans are randomly generated within the specified lower and upper bounds of the problem's variables. This ensures that the initial candidate solutions cover a broad area of the search space. The standard formula for initialising positions is given below:

$$y_{j,k} = m_k + r.\left(up_k - low_k\right), \quad j = 1, 2, 3, ..., n. \tag{29}$$

where the j^{th} candidate solution in k^{th} variable value is given as $y_{j,k}$, P is the entire population, n is the entire problem variables, random variable is denoted as r in [0, 1] interval, k^{th} upper bound is denoted as up_k, and k^{th} lower bound is indicated as low_k.

The population matrix is given below,

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_j \\ \vdots \\ Z_P \end{bmatrix} = \begin{bmatrix} z_{1,1} & \cdots & z_{1,k} & \cdots & z_{1,n} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ z_{j,1} & \cdots & z_{j,k} & \cdots & z_{j,n} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ z_{P,1} & \cdots & z_{P,k} & \cdots & z_{P,n} \end{bmatrix}_{P\times n} \tag{30}$$

Here, row denotes the candidate solution and columns represent the problem variables proposed values. The population matrix is indicated as z and j^{th} pelican is denoted as z_j.

The objective function is used for determining the objective function, which is given as:

$$O = \begin{bmatrix} O_1 \\ \vdots \\ O_j \\ \vdots \\ O_P \end{bmatrix}_{P\times 1} = \begin{bmatrix} O(Z_1) \\ \vdots \\ O(Z_j) \\ \vdots \\ O(Z_j) \end{bmatrix}_{P\times 1} \tag{31}$$

where, the objective function vector is denoted as O, and j^{th} candidate solution objective function value is given as O_j. The hunting process is done in the exploration and exploitation phases.

2 Exploration phase: in the exploration phase of the POA, pelicans move towards randomly generated prey locations within the search space. This phase is crucial for enhancing the exploration capabilities of the algorithm, ensuring a comprehensive search of the problem-solving space. The movement towards prey, combined with random perturbations, allows the pelicans to scan different areas effectively, avoiding local optima and increasing the likelihood of finding the global optimum. The formula is given as:

$$z_{j,k}^{Q_1} = \begin{cases} z_{j,k} + r.\left(q_k - J.z_{j,k}\right), & O_q < Q_j; \\ z_{j,k} + r.\left(z_{j,k} - q_k\right), & else, \end{cases} \tag{32}$$

where J indicates the random number, k^{th} dimension prey is denoted as q_k, and the updated status of k^{th} dimension j^{th} pelican is given as $z_{j,k}^{Q_1}$.

A new position for a pelican is accepted if it results in a better objective function value compared to its current position. This ensures that each update leads to an improvement or maintains the current best solution. It is given as:

$$Z_j = \begin{cases} Z_j^{Q_1}, & O_j^{Q_1} < O_j; \\ Z_j, & else, \end{cases} \tag{33}$$

where j^{th} pelican updated status is denoted as $Z_j^{Q_1}$ and the exploration phase value of the objective function is given as $O_j^{Q_1}$.

3 Exploitation phase: in the local exploitation phase of the POA, pelicans refine their positions by conducting a local search around promising areas identified in the exploration phase. This phase mimics the behaviour of pelicans spreading their wings to move fish upwards and collecting them, representing a focused search for good solutions to improve them further. By evaluating the objective function at neighbouring points and accepting improvements, the algorithm enhances its exploitation ability, converging to better solutions within the search space. This two-phase approach of exploration followed by exploitation ensures a balance between global search and local refinement, leading to effective optimisation. The mathematical formula for this phase is,

$$z_{j,k}^{Q_2} = z_{j,k} + E.\left(1 - \frac{iter_{counter}}{iter_{\max}}\right).(2.r - 1).z_{j,k} \tag{34}$$

where $z_{j,k}^{Q_2}$ denotes the exploitation phase k^{th} dimension j^{th} pelican's updated status, the constant is given as E and the value for this is 0.2, $iter_{counter}$ exhibits the iteration counter and $iter_{\max}$ is the maximum iteration.

The effective updation is based on the rejection and acceptance of the new position of the pelican, which is given in the below equation,

$$Z_j = \begin{cases} Z_j^{O_2}, & O_j^{O_2} < O_j; \\ Z_j, & else, \end{cases} \tag{35}$$

where the exploitation phase value of the objective function is given as $O_j^{O_2}$.

The POA iteratively updates candidate solutions through exploration and exploitation phases, evaluates their objective function values, and tracks the best solution found so far. This process repeats until the stopping criteria are met.

The pseudo-code of the POA is given in Algorithm 1.

Algorithm 1 Pseudo-code of the POA

```
Start
Input of the problem
Evaluation of maximum iterations iter_max and population P
Evaluation of objective function
For iter_counter = 1: iter_max
    Random generation of prey
    For J = 1: P
        Phase 1: Exploration
        For k = 1: n
            Determination of updated status using equation (32)
        End
        Update population member by equation (33)
        Phase 2: Exploitation
        For k = 1: n
            Determination of updated status using equation (34)
        End
        Update population member by equation (35)
        Update optimal candidate result
    End
```

4 Results and discussion

This section elaborates on the results obtained by the devised scheme to verify the improved performance of the POA+ Dense-ResNet.

4.1 Experimental setup

The tool considered for the performance evaluation of the devised scheme is MATLAB.

4.2 Dataset description

The BraTS dataset (BRATS 2020 Dataset, 2023) focused on the development and evaluation of algorithms for the segmentation of brain tumours from MRI scans. The BraTS2020 continued the tradition of pushing forward the boundaries of automated brain tumour segmentation. It mainly focuses on the gliomas, which are a type of brain tumour with high variability in appearance and shape. It contains multi-modal MRI scans, including: T1-weighted (T1), T1-weighted contrast-enhanced (T1Gd), T2-weighted (T2), fluid-attenuated inversion recovery (FLAIR). Also, it includes images with high-grade gliomas (HGG) and low-grade gliomas (LGG).

4.3 Evaluation metrics

1 MSE: it is the average squared variations between predicted and observed values and the formula is,

$$mse = \frac{1}{\mu}\sum_{\tau=1}^{\mu}\left(N_G^* - N_G\right)^2 \tag{36}$$

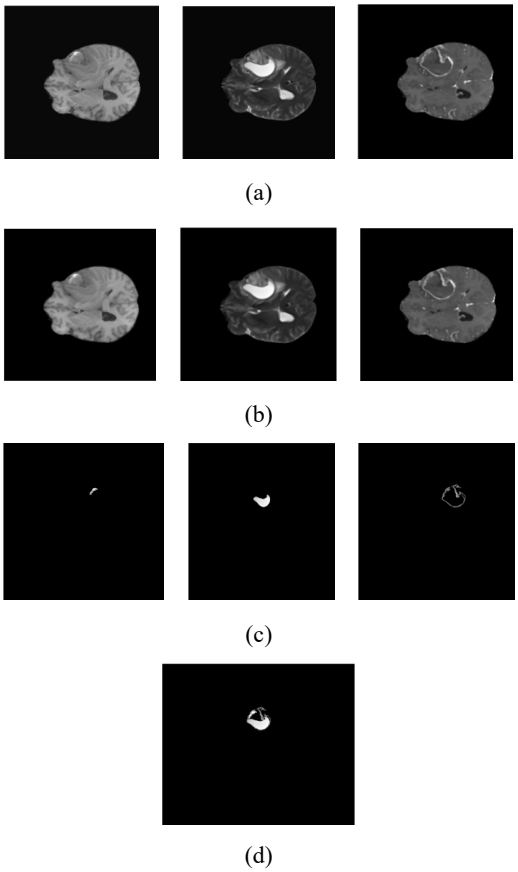
2 PSNR: it is the ratio among the maximum power of the signal and the signal noise power and the expression is given as:

$$psnr = 10 * \log_{10}\left[\frac{(255)^2}{mse}\right] \tag{37}$$

3 RMSE: it is obtained by taking the square root of the MSE and it is expressed by the below equation,

$$rmse = \sqrt{\frac{1}{\mu}\sum_{\tau=1}^{\mu}\left(N_G^* - N_G\right)^2} \tag{38}$$

Figure 6 Experimental results of POA+Dense-ResNet, (a) input images T1, T2, and T1GD, (b) filtered images T1, T2, and T1GD, (c) segmented images T1, T2, and T1GD, (d) multimodal fused image



(a)

(b)

(c)

(d)

4.4 Experimental result

Figure 6 shows the experimental results. The input images, such as T1, T2, and T1GD are given in Figure 6(a). The corresponding filtered and segmentation images are given in Figures 6(b) and 6(c). The final multimodel fused outcome is given in Figure 6(d).

4.5 Performance analysis

The performance analysis of the devised scheme for various epochs is analysed based on K-fold and training data.

4.5.1 K-fold analysis

The K-fold-based performance analysis of POA+ Dense-ResNet is depicted in Figure 7. The analysis on MSE, PSNR, and RMSE is given in Figures 7(a), 7(b), and 7(c). The values attained at K-fold = 9 are discussed as follows: the MSE of POA+Dense-ResNet is 0.851, 0.830, 0.551, 0.141, and 0.056 for 20, 40, 60, 80, and 100 epochs. Also, for 20, 40, 60, 80, and 100 epochs the obtained PSNR of the devised scheme is 43.809 dB, 43.852 dB, 45.634 dB, 48.294 dB, and 51.547 dB. Moreover, the RMSE offered by the devised scheme is 0.923, 0.911, 0.743, 0.376, and 0.236 for 20, 40, 60, 80, and 100 epochs.

4.5.2 Training data analysis

Figure 8 shows the training data-based performance analysis of POA+Dense-ResNet. The performance achieved at 90% training data is discussed in this section. The analysis of MSE is provided in Figure 8(a). The MSE of POA+ Dense-ResNet is 0.789, 0.736, 0.655, 0.597, and 0.472 for 20, 40, 60, 80, and 100 epochs. Figure 8(b) shows the PSNR analysis of POA+Dense-ResNet. For 20, 40, 60, 80, and 100 epochs the obtained PSNR of the POA+Dense-ResNet is 44.231 dB, 45.898 dB, 48.090 dB, 48.360 dB, and 51.365 dB. The RMSE analysis of POA+Dense-ResNet is provided in Figure 8(c). The RMSE offered by POA+Dense-ResNet is 0.888, 0.858, 0.809, 0.773, and 0.687 for 20, 40, 60, 80, and 100 epochs.

4.6 Comparative approaches

The effectiveness of the devised scheme is compared with the deep TL model (Kaur and Singh, 2021), G-CNN and fuzzy neural networks (Wang et al., 2021), GAN (Nandhini Abirami et al., 2022), CNN-HOD (Almasri and Alajlan, 2022), and Dense-ResNet (Ghosh and Jayanthi, 2024).

4.7 Comparative evaluation

The comparison of the POA+Dense-ResNet by varying K-fold and training data is provided here.

4.7.1 K-fold analysis

Figure 9 depicts the K-Fold-based comparative analysis of POA+Dense-ResNet. The performance achieved at K-Fold = 9 is discussed in this section. The analysis of MSE is provided in Figure 9(a). The MSE of Dense-ResNet, G-CNN and fuzzy neural networks, CNN-HOD, GAN, deep TL model, and POA+ Dense-ResNet are 0.720, 0.750, 0.881, 1.066, 1.186, and 0.465. The PSNR analysis is shown in Figure 9(b). The obtained PSNR is 51.592 dB, 49.843 dB, 46.377 dB, 44.780 dB, 44.329 dB, and 52.645 dB for the methods Dense-ResNet, G-CNN and fuzzy neural networks, CNN-HOD, GAN, Deep TL model, and POA+ Dense-ResNet. The RMSE analysis is provided in Figure 9(c). The RMSE achieved by the Dense-ResNet is 0.848, G-CNN and fuzzy neural networks is 0.866, CNN-HOD is 0.939, GAN is 1.033, deep TL model is 1.089, and POA+ Dense-ResNet is 0.682. The POA+ Dense-ResNet outperform all other models in terms of MSE, PSNR, and RMSE, which indicates its potential as a highly effective tool for multimodal image fusion in medical applications.

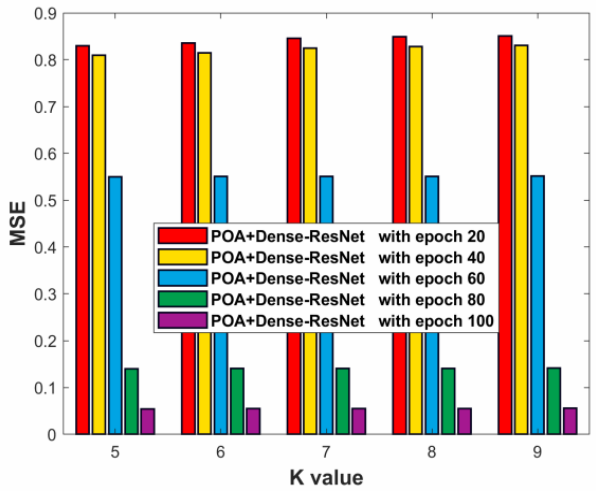
4.7.2 Training data analysis

Figure 10 provides the comparative analysis based on training data. The comparative analysis of MSE, PSNR, and RMSE is depicted in Figures 10(a), 10(b), and 10(c). The values attained at training data 90% are discussed as follows: The MSE of Dense-ResNet, G-CNN and fuzzy neural networks, CNN-HOD, GAN, Deep TL model, and POA+ Dense-ResNet are 0.654, 0.681, 0.801, 0.969, 1.078, and 0.423. The obtained PSNR is 52.454 dB, 50.417 dB, 49.354 dB, 46.164 dB, 43.485 dB, and 53.525 dB for the methods Dense-ResNet, G-CNN and fuzzy neural networks, CNN-HOD, GAN, deep TL model, and POA+Dense-ResNet. The RMSE achieved by the Dense-ResNet is 0.809, G-CNN and fuzzy neural networks is 0.826, CNN-HOD is 0.895, GAN is 0.985, deep TL model is 1.038, and POA+Dense-ResNet is 0.650. The lower MSE and RMSE and higher PSNR values of POA+Dense-ResNet suggest it can more accurately preserve and combine information from different imaging modalities, resulting in a more precise and detailed fused image.

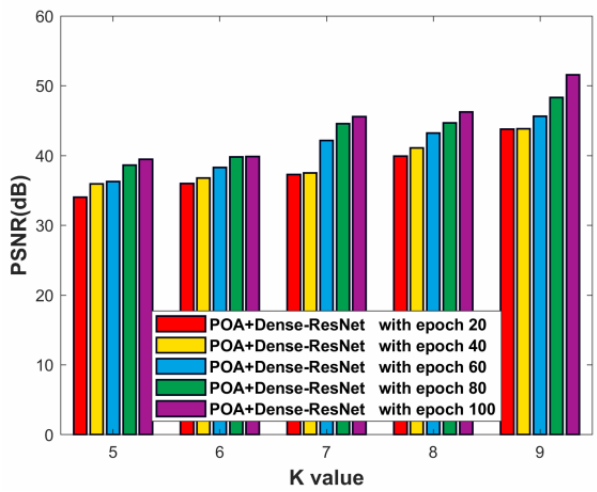
4.8 Comparative discussion

An overview and interpretation of the results obtained from the experiment as displayed in Table 1. The POA+Dense-ResNet consistently outperform other models across all metrics, which indicates its superior capability in the fusion process.

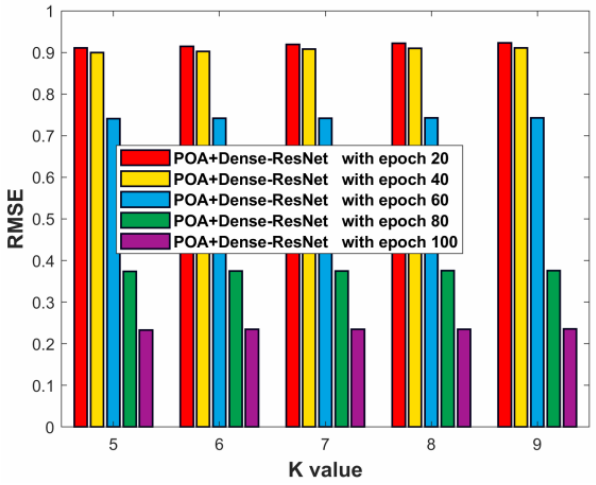
Figure 7 Performance analysis of POA+Dense-ResNet based on K-fold, (a) MSE, (b) PSNR, (c) RMSE (see online version for colours)



(a)

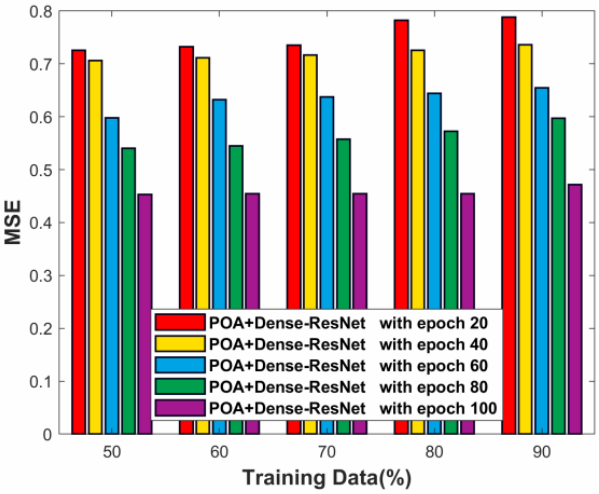


(b)

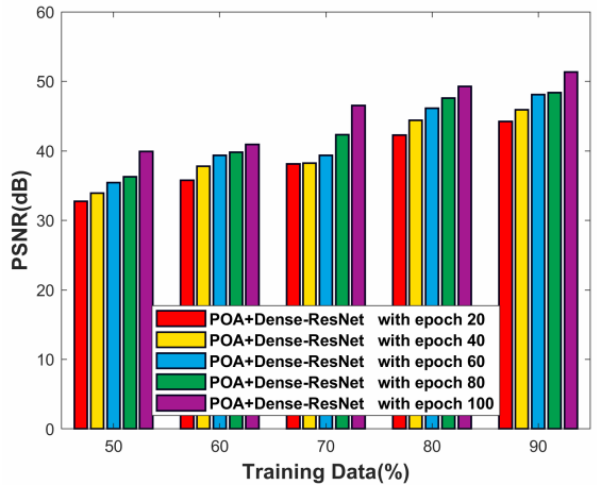


(c)

Figure 8 Performance analysis of POA+Dense-ResNet based on training data, (a) MSE, (b) PSNR, (c) RMSE (see online version for colours)

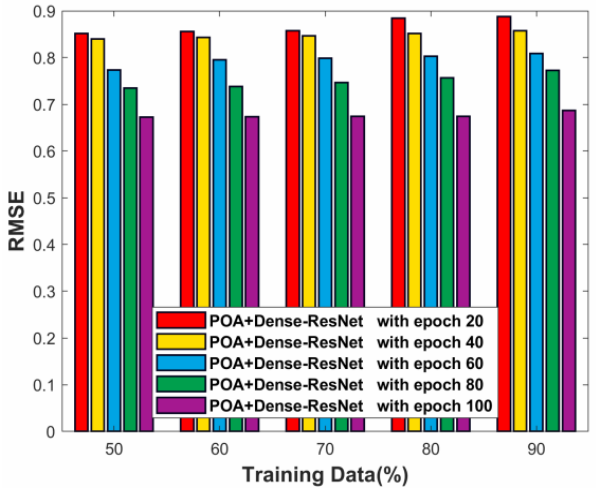


(a)



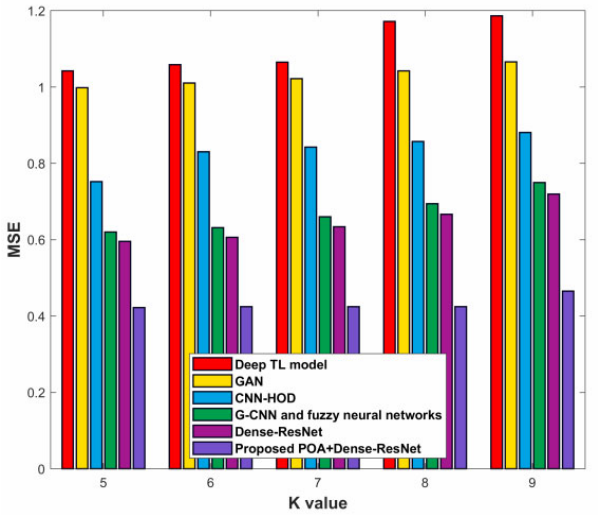
(b)

Figure 8 Performance analysis of POA+Dense-ResNet based on training data, (a) MSE, (b) PSNR, (c) RMSE (continued) (see online version for colours)

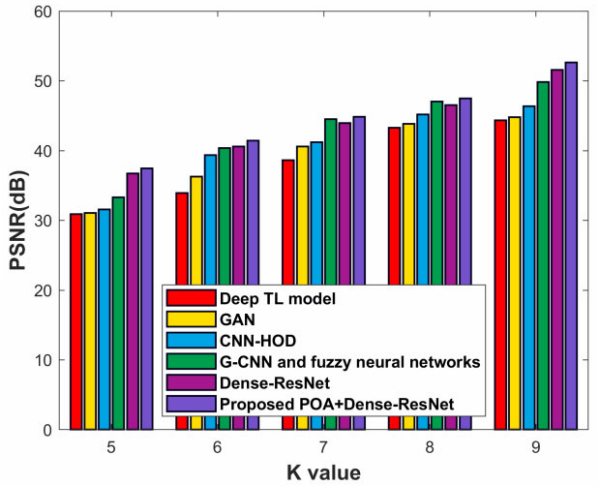


(c)

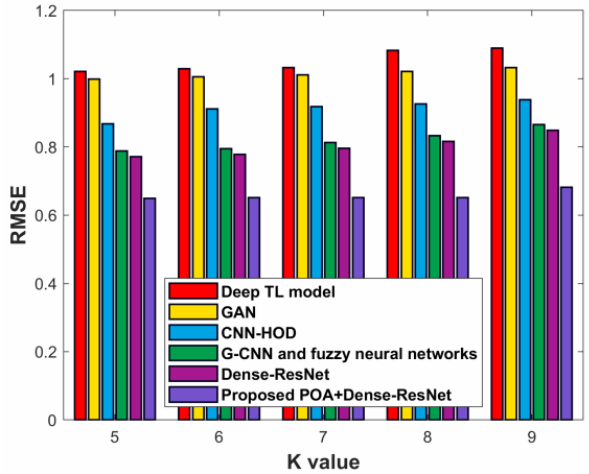
Figure 9 Comparative analysis of POA+Dense-ResNet based on K-fold, (a) MSE, (b) PSNR, (c) RMSE (see online version for colours)



(a)



(b)



(c)

Figure 10 Comparative analysis of POA+Dense-ResNet based on training data, (a) MSE, (b) PSNR, (c) RMSE (see online version for colours)

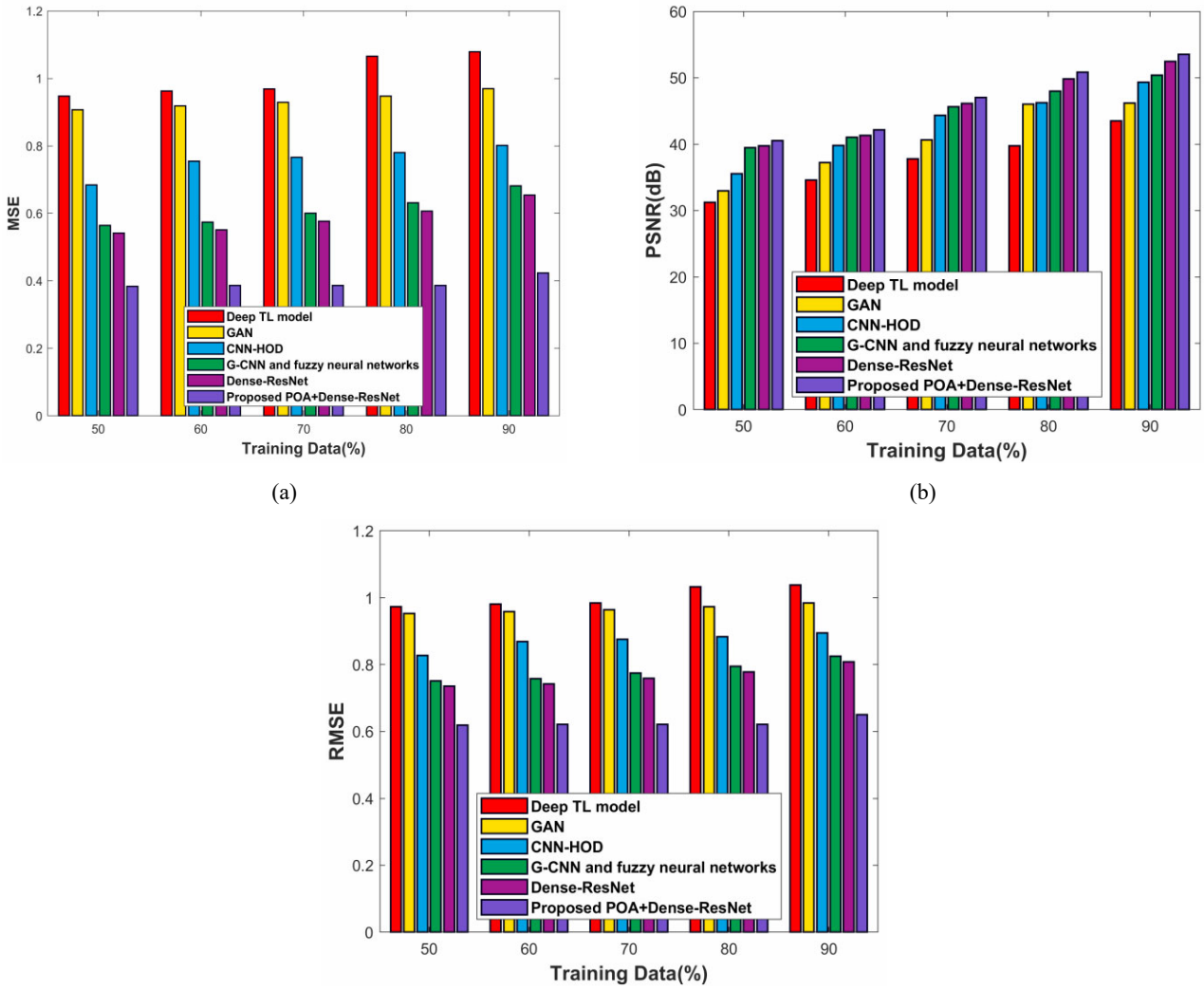


Table 1 Comparative discussion

Evaluation parameters	*Techniques*					
	Dense-ResNet	*G-CNN and fuzzy neural networks*	*CNN-HOD*	*GAN*	*Deep TL model*	*ROA+Dense-ResNet*
			For K-value			
MSE	0.720	0.750	0.881	1.066	1.186	0.465
PSNR (dB)	51.592	49.843	46.377	44.780	44.329	52.645
RMSE	0.848	0.866	0.939	1.033	1.089	0.682
			For training data			
MSE	0.654	0.681	0.801	0.969	1.078	0.423
PSNR (dB)	52.454	50.417	49.354	46.164	43.485	53.525
RMSE	0.809	0.826	0.895	0.985	1.038	0.650

5 Conclusions

This research presents a novel approach to multimodal image fusion in medical imaging, by combining the POA with DenseNet and ResNet, termed POA+Dense-ResNet. Here, the Dense-ResNet is trained by the POA. The images are pre-processed and then the spatial domain to spectral domain transformation is carried out by DTCWT. After that, the ET-Net is used for the segmentation and finally, the fusion is done by the POA+Dense-ResNet. The

POA+Dense-ResNet achieved minimum MSE, RMSE, and maximum PSNR of 0.423, 53.525 dB, and 0.650. The enhanced results achieved by the POA+Dense-ResNet model lead to clearer and more informative medical images. Also, this method has the ability to minimise error and maintain high image quality is crucial for medical applications. In future, implementing and testing the POA+Dense-ResNet model will be done in clinical settings to validate its practical benefits and effectiveness in real-world medical imaging tasks. Also, the application of this fusion technique to other fields beyond medical imaging, such as remote sensing, surveillance, and multimedia will be considered in future work.

References

Almasri, M.M. and Alajlan, A.M. (2022) 'Artificial intelligence-based multimodal medical image fusion using hybrid S2 optimal CNN', *Electronics*, Vol. 11, No. 14, p.2124.

Azam, M.A., Khan, K.B., Salahuddin, S., Rehman, E., Ali Khan, S., Khan, M.A., Kadry, S. and Gandomi, A.H. (2022) '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.

Bhaladhare, P.R. and Jinwala, D.C. (2014) 'A clustering approach for the-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm', *Advances in Computer Engineering*, Vol. 1.

BRATS 2020 Dataset (2023) [online] https://www.kaggle.com/datasets/awsaf49/brats2020-training-data?select=BraTS20+Training+Metadata.csv (accessed 23 July 2023).

Chen, Z., Chen, Y., Wu, L., Cheng, S. and Lin, P. (2019) 'Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions', *Energy Conversion and Management*, Vol. 198, p.111793.

Ding, Z., Li, H., Guo, Y., Zhou, D., Liu, Y. and Xie, S. (2023) 'M4FNet: multimodal medical image fusion network via multi-receptive-field and multi-scale feature integration', *Computers in Biology and Medicine*, Vol. 159.

Ghosh, T. and Jayanthi, N. (2024) 'An efficient Dense-Resnet for multimodal image fusion using medical image', *Multimedia Tools and Applications*, Vol. 83, pp.68181–68208.

Guo, P., Xie, G., Li, R. and Hu, H. (2023) 'Multimodal medical image fusion with convolution sparse representation and mutual information correlation in NSST domain', *Complex & Intelligent Systems*, Vol. 9, pp.317–328.

Huang, B., Yang, F., Yin, M., Mo, X. and Zhong, C. (2020) 'A review of multimodal medical image fusion techniques', *Computational and Mathematical Methods in Medicine*, Vol. 144.

Kaur, M. and Singh, D. (2021) 'Multi-modality medical image fusion technique using multi-objective differential evolution based deep neural networks', *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, pp.2483–2493.

Li, S., Yin, H. and Fang, L. (2012) 'Group-sparse representation with dictionary learning for medical image denoising and fusion', *IEEE Transactions on Biomedical Engineering*, Vol. 59, No. 12, pp.3450–3459.

Li, W., Zhang, Y., Wang, G., Huang, Y. and Li, R. (2023) 'DFENet: a dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion', *Biomedical Signal Processing and Control*, Vol. 80.

Maheshan, C.M. and Prasanna Kumar, H. (2020) 'Performance of image pre-processing filters for noise removal in transformer oil images at different temperatures', *SN Applied Sciences*, Vol. 2, No. 67, pp.1–7.

Meher, B., Agrawal, S., Panda, R. and Abraham, A. (2019) 'A survey on region based image fusion methods', *Information Fusion*, Vol. 48, pp.119–132.

Nagaraja Kumar, N., Jayachandra Prasad, T. and Satya Prasad, K. (2023a) 'An intelligent multimodal medical image fusion model based on improved fast discrete curvelet transform and type-2 fuzzy entropy', *International Journal of Fuzzy Systems*, Vol. 25, pp.96–117.

Nagaraja Kumar, N., Jayachandra Prasad, T. and Satya Prasad, K. (2023b) 'Multimodal Medical image fusion with improved multi-objective meta-heuristic algorithm with fuzzy entropy', *Journal of Information & Knowledge Management*, Vol. 22, No. 1, pp.1–33.

Nandhini Abirami, R., Durai Raj Vincent, P.M., Srinivasan, K., Manic, K.S. and Chang, C.Y. (2022) 'Multimodal medical image fusion of positron emission tomography and magnetic resonance imaging using generative adversarial networks', *Behavioural Neurology*.

Slimen, I.B., Boubchir, L., Mbarki, Z. and Seddik, H. (2020) 'EEG epileptic seizure detection and classification based on dual-tree complex wavelet transform and machine learning algorithms', *Journal of Biomedical Research*, Vol. 34, No. 3, p.151.

Trojovský, P. and Dehghani, M. (2022) 'Pelican optimization algorithm: a novel nature-inspired algorithm for engineering applications', *Sensors*, Vol. 22, No. 3.

Vulli, A., Srinivasu, P.N., Sashank, M.S.K., Shafi, J., Choi, J. and Ijaz, M.F. (2022) 'Fine-tuned DenseNet-169 for breast cancer metastasis prediction using FastAI and 1-cycle policy', *Sensors*, Vol. 22, No. 8, p.2988.

Wang, L., Zhang, J., Liu, Y., Mi, J. and Zhang, J. (2021) 'Multimodal medical image fusion based on Gabor representation combination of multi-CNN and fuzzy neural network', *IEEE Access*, Vol. 9, pp.67634–67647.

Wei, M., Xi, M., Li, Y., Liang, M. and Wang, G. (2023) 'Multimodal medical image fusion: the perspective of deep learning', *Academic Journal of Science and Technology*, Vol. 5, No. 3, pp.202–208.

Ye, S., Wang, T., Ding, M. and Zhang, X. (2023) 'F-DARTS: Foveated differentiable architecture search based multimodal medical image fusion', *IEEE Transactions on Medical Imaging*, Vol. 42, No. 11, pp.3348–3361.

Zhang, G., Nie, R., Cao, J., Chen, L. and Zhu, Y. (2023) 'FDGNet: a pair feature difference guided network for multimodal medical image fusion', *Biomedical Signal Processing and Control*, Vol. 81, p.104545.

Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y. and Shao, L. (2019) 'Et-net: a generic edge-attention guidance network for medical image segmentation', in *Proceedings of 22nd International Conference, in Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Shenzhen, China, Part I, Vol. 22, Springer International Publishing, pp.442–450.