INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Medical Image Fusion Using CNN with Automated Pooling

Gargi J Trivedi[1]*, **Rajesh Sanghvi**[2]

1 Research Scholar, Department of Mathematics, G H Patel College of Engineering & Technology, CVM University, 388120, Vallabh Vidhyanagar, India
2 Assistant professor, Department of Mathematics, G H Patel College of Engineering & Technology, CVM University, Vallabh Vidhyanagar, 388120, India

OPEN ACCESS

Received: 07-09-2022

Accepted: 10-10-2022

Published: 15-11-2022

Citation: Trivedi GJ, Sanghvi R (2022) Medical Image Fusion Using CNN with Automated Pooling. Indian Journal of Science and Technology 15(42): 2267-2274. https://doi.org/10.17485/IJST/v15i42.1812

* **Corresponding author**.

gargi1488@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objectives : The purpose of this study is to demonstrate that the most appropriate pooling is selected among different types of pooling techniques in CNN, for different applications related to image fusion. The work presents a novel method by which pooling is selected automatically as per the needs of the application. **Methods**: This study presents the use of multiple functions for pooling techniques which are selected automatically based on the input images to produce fused optimum output. This will make the classification process much easier and better end results are obtained. **Finding:** The developed automated pooling method performs better concerning processing time and produces optimal output for parameters like Peak Noise Signal Ratio, Mean Square Error, Fusion Factor, Fusion Fidelity, and Visual Information Fidelity in comparison to existing methods like Discrete Wavelet Transformation, Non-Subsampled Contourlet Transform, and Principal Component Analysis. **Novelty**: The novel technique presented in this paper for automated selection of pooling method provides optimal dimension reduction in both the phases of CNN, and hence allows CNN to converge faster and optimally.

Keywords: Image Fusion; CNN; Machine Learning; Pooling; Medical Image Analysis

1 Introduction

Image fusion is gaining importance in modern-day medical analysis and diagnosis. The fused images produce by various medical scanning techniques such as computerized axial tomography scan, positron emission tomography, magnetic resonance imaging, and single photon emission computed tomography are used either individually or combined to give a detailed insight into underlying diseases or foreign objects in biological organisms. Multimodal image fusion is used for medical diagnosis[(1,2)]

Convolution Neural Network (CNN) is used predominantly in medical diagnosis applications. The pooling layers of CNN are responsible for matrix reduction during iterations. For reducing the matrix either maximum or average pooling methods are used. However, there is no automatic approach to telling which pooling technique to select that will provide an optimal solution for CNN to converge. Currently, in most, the image fusion applications method for pooling is either selected randomly or a trial-and-error approach is used for selecting a suitable method for pooling. This paper presents a novel algorithm for the automatic selection of the optimal pooling method. The presented novel algorithm attempts to fill the research gap for optimal selection of pooling methods in CNN, which in turn will increase efficiency for convergence of CNN and reduces the errors for a fused image.

CNN for medical image fusion [3] focuses on map creation. The fusion of pictures may be seen as a classification problem using CNN for identification [4] by measuring activity levels and creating local filters, spatial-domain fusion techniques are used to extract high-frequency information. Convolution is the fundamental operation in a CNN model [5]. Fusion based on CNN outperforms other fusion techniques because it meets the challenges of creating filters to extract useful information. To assess and fuse activity levels, the learning of a CNN is undertaken. Therefore, the fusion approach based on CNN produces results of greater quality than those produced by the standard fusion method. Figure 1 depicts technologies used to generate multimodal medical images [6] of human body parts.

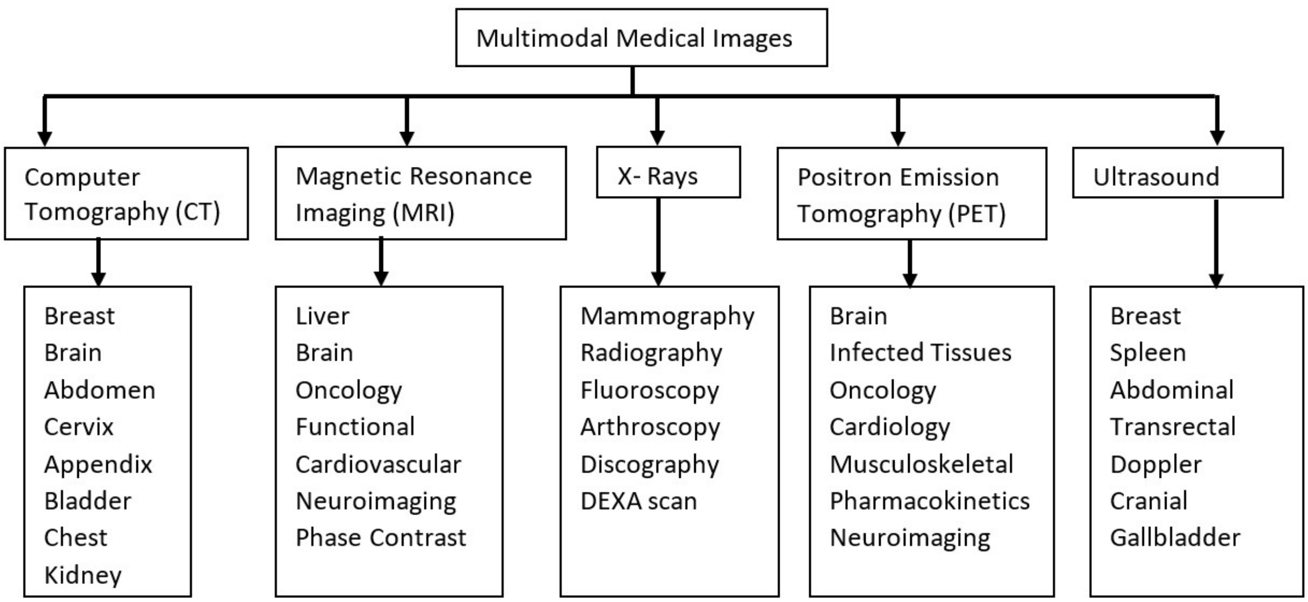

Fig 1. Different types of multimodal medical images

CNN consists of a fully connected network where one or more convolution layers abstract the features of an image and forwards it to the pooling layer. The job of the pooling layer is to optimally reduce the image dimension and pass the image further for output analysis. Hence the output of the CNN depends on the pooling layer and the techniques used in pooling. The most common pooling method is MAX and AVG pooling [7]. Researchers have discovered many other pooling techniques which are beneficial in different application areas. Local pooling is a technique that uses down sampling with feature maps. Global pooling is a technique that uses the full characteristics of maps to obtain a scalar value that provides an image description [8]. This is further passed to connected layers for further classifications. When Max pooling [9] is included in CNN's convolutional layers max pooling decreases the dimensionality of images by reducing the number of pixels inside the output from the preceding convolutional layer. Max pooling chooses just the strongest activation inside the pooling region [10]. Figure 2 represents an example of max pooling. Average pooling [11] obtains an average value of the selected matrix and output is produced. Figure 3 represents average pooling.

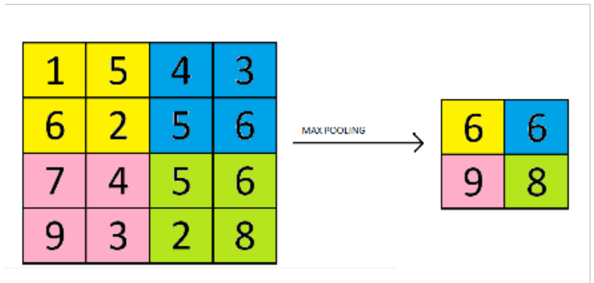

Fig 2. Max pooling technique applied on 4 × 4 matrix to reduce it to 2 × 2

With this notation, max- pooling (max) can be written as $f_{max}(x) = max\{x_i\}_{i=1}^{N}$
For the very alike average-pooling (avg), we have

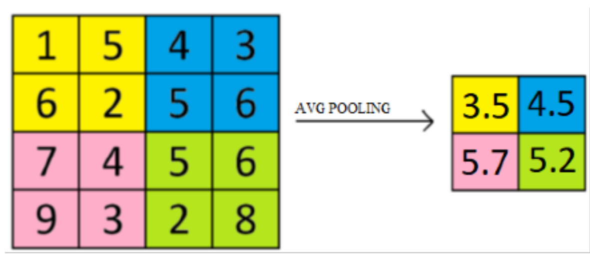

Fig 3. Avg pooling technique applied on 4 × 4 matrix to reduce it to 2 × 2

$$f_{avg}(x) = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

Convolutions with weights w_i can be implemented with

$$f_{conv}(x) = \sum_{i=1}^{N} w_i x_i$$

Different conditions allow for the application of maximum as well as average pooling. In the field of medical image analysis if a lesion exists in a tiny portion of the medical image, then in pooling part region that matches the background pixels will tend to dominate the pooled representation in this situation, so the average pooling may not be the good option. But by using average pooling we can differentiate normal and aberrant images. Max pooling increases the CNN's nonlinearity compared to average pooling(12).

Hence it is valid to conclude that both max and average pooling techniques are needed, however, due to the distinct characteristics of the two methods, an advantage of any one of the methods can be obtained. The following method depicts a way to obtain the dual advantage of max and average pooling methods(13).

Blending method for image fusion(14) utilizes both SR and SCNN, and it consists of three phases. In first phase, complete source images are fed into the traditional orthogonal matching pursuit (OMP), which produces the SR-fused image utilizing the max-rule to enhance pixel localization. During second phase, for each source picture, a brand-new SCNN-based K-SVD dictionary learning method is used once again. To generate the final fused picture, in third phase the fusion rule step lastly uses a linear combination of the above two steps.

Morphological pre-processing method for noise reduction(15) presents the bottom-hat-top-hat technique as a morphological pre-processing method to deal with noise and non-uniform lighting. Then, RGB images are converted into grayscale images that may retain fine details using grey-PCA. The images are then divided into the low-pass and high-pass sub-bands using the local shift-invariant haar wavelet transform approach, which effectively restores all relevant properties in different dimensions and orientations. To efficiently capture smooth edges and textures, the high-pass sub-bands are supplied to two branches of the siamese convolutional neural network. This is done through a process of feature identification, initial segmentation, and consistency verification. While the low-pass sub-bands are fused using existing energy fused, the energy information is recovered utilizing the averaging and selection technique.

Once employed in training and testing as a classifier of medical pictures, convolutional neural networks (CNN) with the Euclidean distance method yield roughly (98.18%) accuracy[(16)].

Multiple convolution layers to extract visual characteristics[(17)] is a technique to extract the key visual characteristics from the various input images, two convolutional layers are used. Following that, a suitable fusion rule (elementwise-max, elementwise-min, or elementwise-mean) is chosen based on the kind of input images to combine the convolutional features of numerous input images. Eventually, two fully connected layers rebuild the fused features to create the informative fused image.

2 Material and methods

The traditional existing methods of image fusion are discussed in this section and each method is simulated for pre-processed images of CT, MRI, and MR along with SPECT and datasets from Kaggle[(18)] and Harvard[(19)] libraries. The same images and data set are also simulated on the developed method for automatic pooling. The result of the simulations is mentioned in Table 1.

2.1 Principal Component Analysis (PCA)

It is a mathematical tool of linear algebra which is used to obtain important data from complex large datasets. This is a typical method for identifying patterns in data of high dimensions, multivariate analysis data analysis is used for a variety of other things[(20)]. Following are the steps of the PCA algorithm:

Step 1: Convert the two input photos to column vectors and use these two column vectors to create the matrix 'B.'

Step 2: Calculate the observational mean vector alongside every column and subtract it from each of the columns of the matrix.

Step 3: Calculate the resultant matrix's covariance matrix, 'C.'

Step 4: Determine the covariance matrix's eigenvalues K and eigenvectors E.

Step 5: Divide each element by the mean of the eigenvector that corresponds to a well-built eigenvalue. This results in the first main component P1. To obtain the second principal component P2, repeat the technique with an eigenvector corresponding to a lesser eigenvalue.

$P_1 = \frac{C(1)}{\Sigma C} P_2 , = \frac{C(2)}{\Sigma C}$

Step 6: The fused picture is augmented by

$I_f = P_1 I_1 + P_2 I_2$

2.2 Discrete Wavelet Transformation (DWT)

Firstly, two same-sized input images are registered then using discrete wavelet transformation, individual images are decomposed. One low-frequency band and three high-frequency bands are present in the reconstructed images (L-H, H-L, H-H bands). The transformation of input image coefficients is fused using the low and high sub-bands fusion rule. Then, the fused image is generated by applying an inverse wavelet transform based on the resultant transform coefficients[(21)].

2.3 Non-Subsampled Contourlet Transform (NSCT)

This transformation creates an image with a multiscale and multidirectional decomposition using non-subsampled pyramids (NSP) and non-subsampled directional filter banks (NSDFB). In contrast to the contourlet transform, the NSCT is a shift-invariant transformation[(22)]. As big as the original image is the multiscale and multidirectional image from NSCT are classified.

2.4 Proposed Method for Automated Pooling

In our method, we follow the flow of a traditional convolutional neural network (CNN) with a modified pooling stage. The input image is fed to the convolutional layer to extract the feature map. The matrices from the convolution layer are fed into the pooling layer for dimension reduction. After dimension reduction again it is passed through the second stage of the convolution layer and finally one more time pooling is applied before passing input to a fully connected feed-forward network to generate output.

Considering Matrix M, we obtain R_{MAX} and R_{AVG} as presented in Figure 4:

Here M_1 to M_n, N_1 to N_n, O_1 to O_n, and P_1 to P_n assume values from 0 to 9.

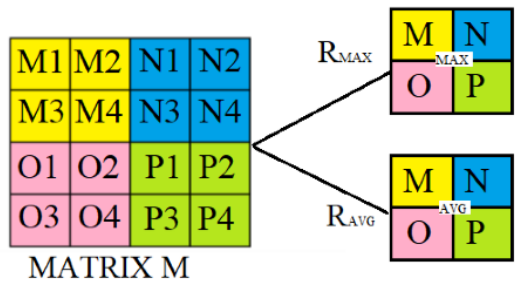

Fig 4. Max and Average pooling on Matrix M to obtainRMAX and RAVG respectively

The algorithm for automated pooling is as follows:
Step 1: For each Matrix M
a. Let R_{MAX} be the matrix obtained by reducing matrix M by applying max pooling
$f_{max}(x) = max(x_i\}_{i=1}^{N}$.
b. Let R_{AVG} be the matrix obtained by reducing matrix M by applying average pooling
$f_{avg}(x) = \frac{1}{N}\sum_{i-1}^{N}|x_i|$.
Step 2: Obtain the result matrix R_{RES} as the difference between R_{MAX} andR_{AVG} as per the following conditions:
a. If R_{MAX} -R_{AVG} = 0, Then R_{RES} = R_{MAX} = R_{AVG}
b. If R_{MAX} -R_{AVG} > 0 and R_{MAX} -R_{AVG} <= 5, Then consider R_{RES} = R_{MAX}
c. If R_{MAX} -R_{AVG} > 5 and R_{MAX} -R_{AVG} <= 10, Then consider R_{RES} = (R_{MAX} + R_{AVG})/2
d. If R_{MAX} -R_{AVG} > 10, Then consider R_{RES} = R_{AVG}
Step 3: Use calculated R_{RES} as a pooling method for dimension reduction in both passes of the CNN method.

3 Discussion

Three distinct datasets with a range of inputs were used to assess the accuracy and performance of the validated proposed automated pooling algorithm.

3.1 Image Quality Assessment Parameters

Most frequently, the Peak Signal Noise Ratio (PSNR) and Mean Square Error (MSE) are utilized to assess the effectiveness of loss compression codec from reconstruction (e.g., for image compression)[23]. Because of this, one reconstruction may look more accurate than the other even though it has a lower PSNR[24]. In Fusion Symmetry (FS) evaluation and asses[25], the Mutual Information (MI) of the input and output images is employed[26]. (The output is of greater quality if the value of obtaining fusion symmetry is smaller.) while Fusion Factor (FF)[27] shows the variance between the input and output of two medical images. Lastly, Visual Information Fidelity (VIF)[28] is a crucial index for assessing Image quality to measure visual distortion.

The first dataset contains CT MRI images. The second dataset contains MRI and PET images. The third dataset contains MRI and PET images. Figure 5 represents fused Different Multimodal images of tumors in the human brain.[18,19]

3.2 Experimental Analysis

The experimental analysis is categorized into a qualitative and quantitative analysis of different fusion algorithms.

3.2.1 Visual Quality Analysis

This part presents a qualitative analysis of fused images from various approaches, as well as the implemented method for automated pooling in CNN. Three image datasets are used in the tests[18,19] (CT-MRI, MRI-PET, MRI-SPECT). All the datasets in Figure 5 are subjected to visual qualitative analysis.

The sub-images of Figure 5(C)–(H) show the fusion result of figures (A)and (B), (B) and (E), (B) and (F). these are the pair of slices scanned along the axial plane of the human brain. Ideally, the fusion of figures (A) and (B) should include the bright skull feature of the CT image and textural tissue information of the MRI. Our developed method for automated pooling has successfully combined the silent features of CT and MRI images but other methods suffered from blurring. PET-MRI images are

Table 1. Quality Assessment parameters for Different Algorithms

Test Dataset	Algorithms	PSNR	MSE	FF	FS	VIF	Processing Time (in Second)
Dataset 1 CT + MRI	**Developed Automated Pooling Method**	**33.44**	**0.3561**	**3.4889**	**0.1021**	**0.922**	**31**
	DWT	33.42	0.3896	3.4563	0.1154	0.836	29
	NSCT	32.55	0.5663	3.3289	0.1123	0.775	35
	PCA	34.58	0.8023	3.4063	0.1232	0.579	32
Dataset 2 MR+PET	**Developed Automated Pooling Method**	**37.57**	**0.5739**	**4.5879**	**0.0988**	**0.913**	**30**
	DWT	31.48	0.4569	4.5892	0.172	0.845	27
	NSCT	36.85	0.6923	3.9897	0.198	0.789	29
	PCA	38.75	0.5567	4.3256	0.210	0.678	35
Dataset 3 MR+SPECT	**Developed Automated Pooling Method**	**39.45**	**0.6872**	**5.7998**	**0.1154**	**0.923**	**28**
	DWT	38.42	0.5689	5.9754	0.1322	0.693	24
	NSCT	39.58	0.6320	5.4323	0.2135	0.785	29
	PCA	34.52	0.6568	5.4689	0.2356	0.856	25

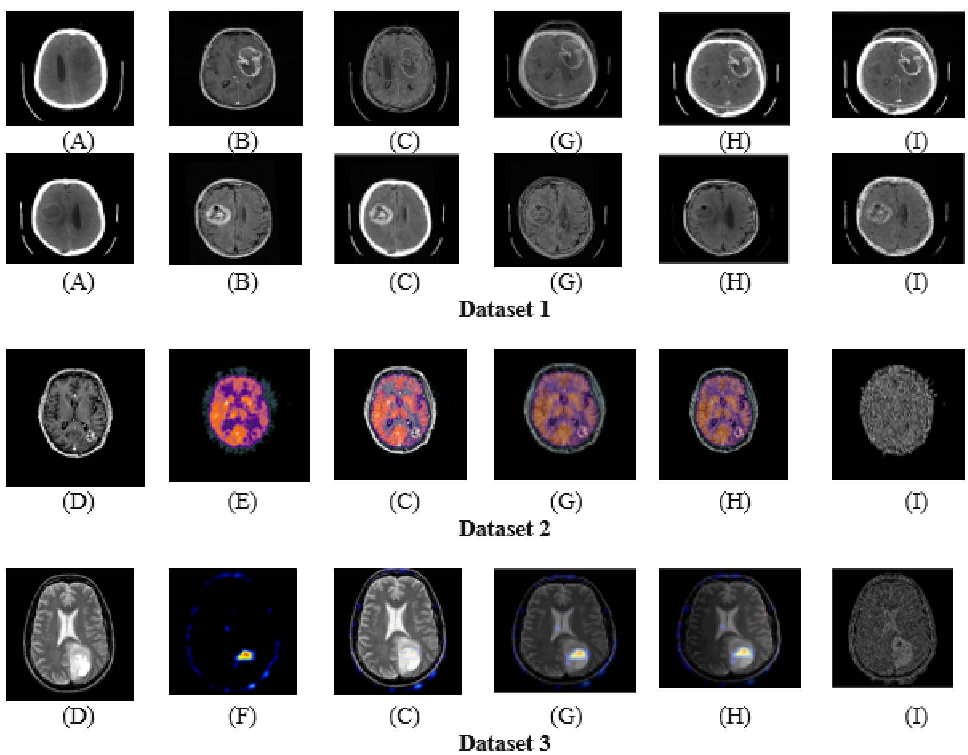

Fig 5. Experiment results on images: (A) CT image; (B) MRI -image;(C) Fused Image Using Developed Automated Pooling Method ; (D) MR-Image (E) PET- Image; (F) SPECT – Image ; (G) Fused Image Using NSCT; (H) Fused Image Using DWT; (I) Fused Image Using PCA

included in the second dataset. PET imaging is frequently used to determine whether lesions are cancerous or benign. Surgery for the lesion is virtually always based on the structural assessment provided by MRI. The acquired findings are depicted in sub image of Figure 5, where the fused picture (C) maintains both MRI and PET data characteristics. The third dataset MRI-SPECT fusion is intended to provide anatomical information with high soft-tissue contrast for reliably detecting regions of physiological problems. Figure 5 (C) sub image shows that our developed CNN with an automatic pooling approach generates fused pictures with greater contrast and intensity than PCA, DWT, and NSCT fusion methods.

3.2.2 Quantitative Analysis

Table 1 displays quantitative evaluation-based image assessment parameters on three distinct datasets, with the implemented technique CNN using automated pooling to obtain the greatest PSNR values. For MSE, DWT performs best, while our technique comes in second. Because of its decomposing image approach, we believe the DWT technique is good at maintaining edge information. DWT outperforms our technique in FF but falls short in FS also in VIF. Our implemented approach outperforms PCA and NSCT. In summary, the suggested automated poling in CNN successfully fuse distinct multimodal images and assuring the quality of the fused images with maintaining an acceptable computing efficiency.

4 Conclusion

The developed automated pooling method when simulated on different datasets mentioned in table 1 generates average peak signal noise ratio of 36.82, average mean square error of 0.53, average fusion factor of 4.62, average fusion symmetry of 0.10, average visual information fidelity of 0.91 and average processing time of 29.66 seconds which is better in comparison to average quality factors and processing time of existing DWT, NSCT and PCA methods. From the results achieved it is clear that the developed method for automated pooling works better and converges faster. Based on image quality assessment parameters simulated on different data sets it can be concluded that each of the existing standard methods is not suitable for all kinds of data sets and hence these methods cannot be generalized multi-modal image fusion. In contrast to these methods, the developed methods work well for every type of data set due to the selection of its automated pooling technique. Developed method enhances the convergence ratio of CNN used for image fusion applications. The presented work on image fusion by CNN with automated pooling demonstrates a bright direction in the field of picture fusion with plenty of scope for advancement in the future. In future automated pooling can add new optimal pooling methods when they are developed. Further similar studies will likely carry on in the upcoming years to support the growth of image fusion.

References

1) Shibu DS, Priyadharsini SS. Multi scale decomposition based medical image fusion using convolutional neural network and sparse representation. *Biomedical Signal Processing and Control*. 2021;69:102789. Available from: https://doi.org/10.1016/j.bspc.2021.102789.
2) Zhang H, Xu H, Tian X, Jiang J, Ma J. Image fusion meets deep learning: A survey and perspective. *Information Fusion*. 2021;76:323–336. Available from: https://doi.org/10.1016/j.inffus.2021.06.008.
3) Wang L, Chang C, Liu Z, Huang J, Liu C, Liu C. A Medical Image Fusion Method Based on SIFT and Deep Convolutional Neural Network in the SIST Domain. *Journal of Healthcare Engineering*. 2021;2021:1–8. Available from: https://doi.org/10.1155/2021/9958017.
4) Liu S, Wang M, Yin L, Sun X, Zhang YDD, Zhao J. Two-Scale Multimodal Medical Image Fusion Based on Structure Preservation. *Frontiers in Computational Neuroscience*. 2022;15. Available from: https://doi.org/10.3389/fncom.2021.803724.
5) Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: A survey. *IET Image Processing*. 2022;16(5):1243–1267. Available from: https://doi.org/10.1049/ipr2.12419.
6) Wang K, Zheng M, Wei H, Qi G, Li Y. Multi-Modality Medical Image Fusion Using Convolutional Neural Network and Contrast Pyramid. *Sensors*. 2020;20(8):2169–2169. Available from: https://doi.org/10.3390/s20082169.
7) Wang Z, Li X, Duan H, Su Y, Zhang X, Guan X. Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform. *Expert Systems with Applications*. 2021;171:114574. Available from: https://doi.org/10.1016/j.eswa.2021.114574.
8) Rajendran N, Siyamalan M, Ramanan A, Wang R. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing and Applications*. 2022;34(1):5321–5347. Available from: https://doi.org/10.1007/s00521-022-06953-8.
9) Kaur M, Singh D. Fusion of medical images using deep belief networks. *Cluster Computing*. 2020;23(2):1439–1453. Available from: https://doi.org/10.1007/s10586-019-02999-x.
10) Wei Z, Zhang J, Liu L, Zhu F, Shen F, Zhou Y, et al. Building Detail-Sensitive Semantic Segmentation Networks With Polynomial Pooling. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019;p. 7108–7116. Available from: https://doi.org/10.1109/CVPR.2019.00728.
11) Xu H, Ma J, Fusion EM. An unsupervised enhanced medical image fusion network. *Information Fusion*. 2021;76:177–186. Available from: https://doi.org/10.1016/j.inffus.2021.06.001.
12) Yousif AS, Omar Z, Sheikh UU. An improved approach for medical image fusion using sparse representation and Siamese convolutional neural network. *Biomedical Signal Processing and Control*. 2022;72(Part B):103357. Available from: https://doi.org/10.1016/j.bspc.2021.103357.
13) Abas AI, Koçer HE, Baykan NA. Medical image fusion with convolutional neural network in multiscale transform domain. *Turkish Journal of Electrical Engineering and Computer Sciences*. 2021;29(SI-1):2780–2794. Available from: https://doi.org/10.3906/elk-2105-170.

14) Abdelrahman L, Ghamdi AM, Collado, Mesa F, Mottaleb A, M. Convolutional neural networks for breast cancer detection in mammography: A survey. *Comput Biol Med*. 2021;131:104248. Available from: https://doi.org/10.1016/j.compbiomed.2021.104248.
15) Bhutto JA, Tian L, Du Q, Sun Z, Yu L, Tahir MF. CT and MRI Medical Image Fusion Using Noise-Removal and Contrast Enhancement Scheme with Convolutional Neural Network. *Entropy*. 2022;24(3):393. Available from: https://doi.org/10.3390/e24030393.
16) amer mohamed N, Al-Tamimi MSH. Image Fusion Using A Convolutional Neural Network. *Solid State Technology*. 2020;63(6):13149–13162.
17) Zhang Y, Liu Y, Sun P, Yan H, Zhao X, Zhang L. IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion*. 2020;54:99–118. Available from: https://doi.org/10.1016/j.inffus.2019.07.011.
18) Brain MRI Images for Brain Tumour Detection. . Available from: https://www.kaggle.com/datasets.
19) 16.Harvard Medical School lecture notes: Introduction to Neuroimaging . . Available from: https://www.med.harvard.edu/aanlib/.
20) Polinati S, Bavirisetti DP, Rajesh NVPSK, Dhuli R. Multimodal Medical Image Fusion Based on Content-based and PCA-sigmoid. *Current Medical Imaging*. 2022;18(5):548–556. Available from: http://dx.doi.org/10.2174/1573405617666211004114726.
21) Rao BSN, Raju NVK, Dhanush M, Harshith PNSM, Mehdi MJ. MRI and Spect Medical Image Fusion using Wavelet Transform. *2022 7th International Conference on Communication and Electronics Systems (ICCES)*. 2022;2022:1690–1696. Available from: https://doi.org/10.1109/ICCES54183.2022.9835857.
22) Iqbal C, Ghafoor A, Bakhshi AD, Saghir NJ. Medical image fusion using non subsampled contourlet transform and iterative joint filter. *Multimedia Tools Applications*. 2022;81:4495–4509. Available from: https://doi.org/10.1007/s11042-021-11753-8.
23) Li Y, Zhao J, Lv Z, Pan Z. Multimodal Medical Supervised Image Fusion Method by CNN. *Frontiers in Neuroscience*. 2021;15(638976). Available from: https://doi.org/10.3389/fnins.2021.638976.
24) Parvathy VS, Pothiraj S, Sampson J. A novel approach in multimodality medical image fusion using optimal shearlet and deep learning. *International Journal of Imaging System and Technology*. 2020;23(4):847–859. Available from: https://doi.org/10.1002/ima.22436.
25) Chang J, Zhang L, Gu N, Zhang X, Ye M, Yin R, et al. A mix-pooling CNN architecture with FCRF for brain tumor segmentation. *Journal of Visual Communication and Image Representation*. 2019;58(1):316–322. Available from: https://doi.org/10.1016/j.jvcir.2018.11.047.
26) Li Y, Zhao J, Lv Z, Li J. Medical image fusion method by deep learning. *International Journal of Cognitive Computing in Engineering*. 2021;2(1):21–29. Available from: https://doi.org/10.1016/j.ijcce.2020.12.004.
27) Mohan R, Ganapathy K, Arunmozhi R. Brain tumor classification of magnetic resonance Images using novel CNN-based medical image Analysis and Detection network in comparison with AlexNet. *Journal of Population Therapeutics and Clinical Pharmacology*. 2022;29(1):e97–e108. Available from: https://doi.org/10.47750/jptcp.2022.898.
28) Almasri MM, Alajlan AM. Artificial Intelligence-Based Multimodal Medical Image Fusion Using Hybrid S2 Optimal CNN. *Electronics*. 2022;11(14):2124. Available from: https://doi.org/10.3390/electronics11142124.