International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Advancements in Transfer Learning Strategies for PET and MRI Brain Image Fusion

Sonia Panesar[1] and Amit Ganatra[2]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: Multimodal medical picture fusion is a trending research topic, in which research is happening. Multimodal medical image fusion is a procedure that integrates information from several medical imaging modalities to provide a more useful and complete visual content in the same picture to do subsequent operations like image segmentation etc. Medical image fusion may be particularly advantageous for biomedical research and medical image analysis and to minimize both the scan duration and motion artifacts in scan. The merging of neuroimaging data may lead to new insights into brain function and structure. In this article, multiple deep learning techniques including pre-trained VGG19 model, ALEXNET model and DENSENET model are applied utilizing transfer learning methodology to merge MRI (Magnetic Resonance Imaging) and PET (Positron Emission Tomography) neuroimaging. As the availability to medical data is restricted, transfer learning is employed for feature extraction and save training time. The features are blended using a pre-trained VGG19 model, ALEXNET model and DENSENET model. The experimental findings of all the three models include both quantitative and qualitative assessment metrics analysis for fused picture and achieves superior overall performance than unimodal and feature-level fusion approaches, and that it beats state-of-the-art methods.

Keywords: *Fusion, Transfer Learning, PET, MRI, VGG19, ALEXNET, DENSENET, Image Fusion, Multimodal Medical image.*

1. Introduction

Medical imaging scans are diagnostic procedures used to provide visual reconstructions of the inside of the human body. These scans enable healthcare providers to observe and analyze the structure and function of organs, tissues, and systems without the need for intrusive treatments [1,2]. These plays a significant part in the diagnosis, monitoring, and treatment planning of numerous medical disorders. Neuroimaging is crucial in the area of neuroscience and clinical practice. It enables for the diagnosis and monitoring of numerous neurological and psychiatric diseases, such as brain tumors, stroke, epilepsy, Alzheimer's disease, schizophrenia, and more. Neuroimaging essentially separates into two categories- structural neuroimaging, offers the anatomical information of the organ on the other hand functional neuroimaging, depicts and analyzes the change in metabolism and blood flow. The structural and functional neuroimaging fusion permits medical practitioners to concurrently observe soft tissues and molecular processes altogether and delivers more constructive knowledge about the same thing. Brain tumor segmentation may be properly conducted by leveraging the integration of structural neuroimaging like MRI (Magnetic Resonance Imaging) with functional neuroimaging like PET (Positron Emission Tomography). Each modality gives unique information about the human body, and fusion is applied to maximize the strengths of each modality while accounting for their specific limits. For instance, the combined employment of positron emission tomography (PET) and computed tomography (CT) has become a regular practice in the clinical field for several neurological illnesses [3,4]. The major objective of combining various modalities is to build a system that smoothly combines diverse diagnostic procedures to provide a single picture. This integrated depiction becomes helpful in supporting experts such as radiologists, oncologists, and interventionists throughout their diagnostic procedures and decision-making. Deep learning (DL) has produced amazing breakthroughs in numerous computer vision and image processing difficulties. Numerous DL-based approaches for multimodal image fusion, addressing issues including multi-focus picture fusion, multi-exposure image fusion, and multi-sensor image fusion, have been presented. These approaches significantly enhance the efficacy of applications such as image-guided disease analysis, medical diagnostics, automatic change detection, navigation assistance, military operations, remote sensing, digital imaging, aerial and satellite imaging, microscopic imaging, and concealed weapon detection in satellite images [5,6].

1.1. Types of Multimodal Image Fusion

In Literature, there are various distinct kinds of multimodal picture fusion approaches, each with its own benefits and uses as outlined as follows:

1. Pixel Level Fusion:- This sort of fusion is directly

[1] Assistant Professor, Computer Sci. & Eng., Babaria Institute of Technology, Bits Edu Campus, Vadodara, Gujarat, India
Email ID: soniafpanesar@gmail.com,
ORCID ID : 0009-0006-3544-7285
[2] Provost, Parul University, Vadodara, Gujarat, India
ORCID ID : 0000-0002-9993-9901

applied to the pixel values of the pictures, independent of the features being examined. The focus is on merging the full pixel, including all its attributes, without any specific emphasis on the intensity or color elements. For example, intensity-level fusion is a special sort of pixel-level fusion that highlights the intensity values or grayscale properties of the pixels [2,3,7]. It has the power to integrate a high-resolution grayscale picture, such as X-ray, with a color image from a different modality. This results in a merged picture that contains specific anatomical information from the X-ray and color or contextual information from the other modality.

2. Feature Level Fusion:- Feature-level fusion involves merging information from several sources or modalities at a low level, generally after the extraction of fundamental features from separate sources but before higher-level analysis or decision-making procedures. It is often used with machine learning models. The purpose of feature-level fusion is frequently to increase the information available for further processing by merging complimentary features from diverse sources and enhancing overall system performance in tasks like tumor classification etc. It also offers issues, such as dealing with heterogeneous data, aligning distinct feature spaces, and addressing changes in size and resolution across the input modalities.

3. Decision Level Fusion:- Decision-level fusion, also known as late fusion or post-processing fusion, is a method in which the outputs or decisions from several sources or classifiers are joined at a later stage. These sources may act on the same or separate modalities. The fusion process happens after the separate sources have made their conclusions or projections. Decision-level fusion is typically applied when dealing with heterogeneous data or when the sources use distinct techniques, models, or representations. The fusion procedure helps resolve the disparities in the results [8,13]. Decision-level fusion may increase the resilience and dependability of a system. By merging judgments from numerous sources, the entire system may be more resistant to mistakes or uncertainty associated with individual inputs. Redundancy in information may be utilized to promote dependability. There are several approaches for decision-level fusion, including voting systems (e.g., majority voting, weighted voting), averaging, and more complex techniques such as ensemble methods like bagging or boosting. The selection of the approach depends on the features of the data and the special needs of the application. Challenges related with decision-level fusion include assuring compatibility between the outputs of diverse sources, addressing uncertainties, and defining suitable weighting or voting algorithms.

1.2. Applications of PET/MRI Fusion

Positron Emission Tomography (PET) , a functional neuroimaging and Magnetic Resonance Imaging (MRI), a structural neuroimaging are sophisticated medical neuroimaging methods that play significant roles in the diagnosis, staging, and monitoring of different medical disorders. Each modality has its distinct strengths and uses [8,9]. PET-MRI fusion provides a strong and synergistic technique in medical imaging. By combining the capabilities of both modalities, PET-MRI gives complementary information regarding both anatomy and metabolic activity. The Key applications of PET-MRI fusion are as following :

1. Oncology:-

• Tumor Localization and characterisation: PET-MRI fusion enables for exact localization and characterisation by integrating the anatomical features from MRI with the metabolic information from PET.

• Staging and Restaging: In cancer staging, PET offers information regarding the metabolic activity of tumors, Cells that are quickly developing or are metabolically active take up the glucose and light up on the scan. In general, cancer cells are more metabolically active than normal cells and tend to "light up" whereas MRI delivers precise anatomical imaging, which includes data about lesion location, size, shape, and structural alterations to nearby tissues . This assists in correct staging and restaging of cancer [1,3].

2. Pediatric Imaging :-

• Pediatric Oncology: PET-MRI is especially effective in pediatric oncology for limiting radiation exposure while giving thorough anatomical and functional information.

• Congenital Abnormalities: The combination of PET and MRI is advantageous in the evaluation of congenital abnormalities, giving thorough structural and metabolic information.

3. Neurology :-

• Brain Imaging: PET-MRI is important in neuroimaging, delivering rich structural information from MRI with functional and metabolic data from PET. It assists in the diagnosis and monitoring of neurodegenerative illnesses. Degenerative brain illnesses, such as moderate cognitive impairment, Alzheimer's, and Parkinson's, originate from deteriorating neuronal function and diminished neuron numbers in the central nervous system. These degenerative disorders, impairing memory, speech, and mobility, represent considerable issues as the aging population rises, with no known solution and severe repercussions on people, families, and society [12].

• Epilepsy assessment: For the assessment of epilepsy, several imaging methods like MRI, PET, SPECT, EEG, MEG etc may be utilized to determine the underlying causes and find the epileptogenic center. The choice of imaging modality relies on the individual clinical setting and the information sought. The combination of PET and MRI is utilized in the examination of epilepsy patients to locate the epileptogenic center and analyze structural abnormalities.

4. Cardiology:-

• Myocardial Perfusion and Viability: Myocardial illness, the primary cause of mortality in humans, includes anomalies in the heart muscle. Stunned myocardium refers to a situation when there is wall dysfunction, but the perfusion (resting and stress) remains normal. Myocardial ischemia occurs when there is diminished perfusion of the myocardium during stress (e.g., during exercise) but normal perfusion at rest, characterized as a reversible perfusion abnormality. Patients with reversible perfusion abnormalities may considerably improve from therapy. Hibernating myocardial has reduced perfusion in both stress and resting phases, manifesting as a permanent defect. Despite the diminished perfusion, the myocytes remain alive and may benefit from revascularization. In the event of myocardial infarction, there is a lack of perfusion both under stress and at rest, resulting in a permanent defect, and the myocytes are not viable. Revascularization does not give any advantage in such circumstances. The fusion of PET-MRI is applied in measuring myocardial perfusion and viability, delivering full information on blood flow, metabolism, and heart anatomy.

• Cardiac Tumor Detection: In situations of suspected cardiac malignancies, PET-MRI may aid in localizing and defining the lesions.

1.3. Organization of Paper

The paper follows a systematic format as indicated below: The introduction gives an overview, explaining several forms of multimodal image fusion and stressing the applications of PET/MRI fusion. Section 2 investigates the numerous strategies deployed, with a special emphasis on deep learning and transfer learning techniques. Subsections under transfer learning study particular models, including VGG19, ALEXNET, and DENSNET. Section 3 explains the proposed multimodal fusion procedure, followed by a comprehensive discussion of experimental parameters in Section 4. The analysis of the acquired findings is reported in Section 5. The report finishes by summarizing the results and analyzing their ramifications. Finally, a list of references is supplied to recognize the sources and background material that informed the research.

2. Different Methodology

2.1. Deep Learning

Before deep learning became popular, there was a lot of study done on image fusion. Image fusion issues are further subdivided into visible/infrared image fusion, multi-focus image fusion techniques, multi-exposure picture fusion, multi-temporal image fusion, and remote sensing image fusion, among other subproblems, based on different application domains. Traditional fusion techniques refer to the early approaches to these image fusion challenges that used mathematical transformations to manually examine activity levels and create fusion rules in the transformation or spatial domain.

These conventional techniques, which are designed to satisfy the unique needs of various applications, include multi-scale transform-based approaches, sparse representation-based techniques, methods based in the spatial domain, hybrid transform-based methods, and methods based on total variation or remote sensing image fusion [10,11]. But it's becoming more and more clear how limited these conventional approaches are. To guarantee the viability of feature fusion later, conventional approaches are limited to using the same transformation for feature extraction across various source pictures. Creating efficient picture representation techniques and fusion rules to provide cutting-edge outcomes is becoming more difficult, therefore.

Because of its strong capabilities in feature extraction, representation, fusion, and reconstruction, deep learning (DL) has had a considerable impact on the discipline since its introduction. The main benefit of deep learning (DL) is its ability to create hierarchical representations by separating high-level features from low-level features. This allows DL to achieve state-of-the-art performance in many computer vision and image fusion problems.

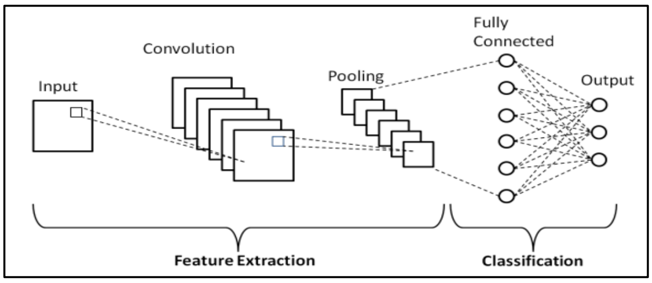

Fig. 1. CNN network architecture [12]

The Convolutional Neural Network (CNN), as seen in Fig. 1, is the best deep learning model for computer vision because of its many architectural features, such as reduced parametrization, weight sharing, hierarchical feature learning, local connection, and parameter sharing. Furthermore, pre-trained CNN models that have been trained on huge datasets like ImageNet are easily accessible and adaptable to particular applications. Even with a little amount of labeled data, deep networks may be trained effectively via the use of transfer learning using pre-trained models. In a work by [10], the trainable layer for extracting features from infrared pictures was fine-tuned using transfer learning, which included redesigning the CNN architecture.

A typical convolutional neural network's design consists of many layers placed in a certain order. Three primary layers make up a conventional CNN architecture, while modifications may exist depending on particular CNN models and tasks:

1. Input Layer: Takes unprocessed pictures.

2. Convolutional Layers: To capture characteristics that are more and more complicated, many layers are layered. These layers use filters or kernels to apply convolution operations on the incoming data.

3. Activation Function: Often known as ReLU (Rectified Linear Unit), this function is used elementwise to provide the model non-linearity so that it may pick up on more intricate linkages.

4. Layers for pooling (subsampling or down sampling): These layers minimize the input volume's width and height. In order to down sample the data and improve computing performance, a popular approach called max pooling keeps the largest value from a set of values.

5. Fully Connected (Dense) Layers: Utilized at the conclusion of the network to aggregate high-level features and generate predictions, these layers link every neuron in one layer to every other layer's neuron.

6. Flatten Layer: This layer is used before the fully connected layers and helps with input into the fully connected layers by converting the multi-dimensional output of the convolutional and pooling layers into a one-dimensional vector.

7. Output Layer: Generates the ultimate forecasts.

2.2. Transfer Education

using training on the target task, a pre-trained model on a large dataset may be further refined using transfer learning, a potent deep learning approach that updates the weights depending on input from the target task. Three main issues with conventional machine learning methods are usually addressed by transfer learning (TL): (1) insufficient labeled data; (2) insufficient processing capacity; and (3) distribution mismatches. Transfer learning (TL) may be broadly divided into four categories: four types of learning are transudative, inductive, unsupervised, and negative [5,18]. In addition, there are four different forms of learning within each category: relation-based learning, feature-based learning, parameter-based learning, and instance-based learning. For many industries, transfer learning is beneficial to multimodal learning. Through performance optimization, data scarcity surmounting, and the use of pre-trained models on a variety of data sources, transfer learning improves multimodal learning in the healthcare industry. It enables the integration of data from several modalities, including patient records and medical imaging, resulting in more reliable and precise healthcare forecasts. This method speeds up model training, reduces the amount of labeled data required, and eventually improves diagnostic and prognostic performance in medical applications.

VGG19

The deep convolutional neural network architecture known as VGG-19 was first created for image categorization applications. Due to its capacity to extract hierarchical features and patterns from pictures, VGG-19 may be used in the context of fusion tasks even if it is not specifically designed for multimodal medical image fusion [16]. Effective spatial hierarchy capture is facilitated by the VGG-19 architecture's widespread usage of tiny 3x3 convolutional filters.

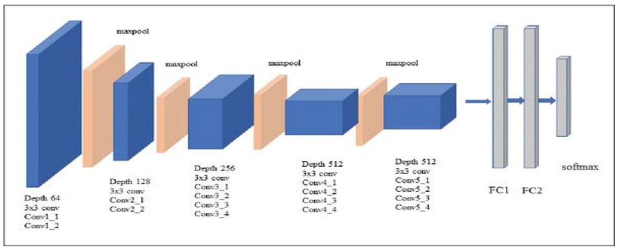

Fig. 2. VGG19 deep convolutional neural network architecture [18]

Every convolutional layer in VGG-19 extracts features at a distinct scale, as shown in fig. 2. The network's flexibility to adjust to varied scales makes it appropriate for managing a range of resolutions and modalities' properties, which aids in efficient fusion. VGG-19 is a pre-trained model that can be used to big datasets like ImageNet. The model can learn general characteristics from a variety of photos because to this pre-training. For the sake of the particular fusion challenge, transfer learning entails optimizing the previously trained VGG-19 on a smaller dataset of medical images. This facilitates the use of natural image characteristics that have been learnt for medical image fusion. This is especially useful in situations when there are few medical imaging datasets available, as in MRI and PET scans of the same patient with a brain problem. It successfully addresses data scarcity problems. Qualitative assessment measures show state-of-the-art performance in the use of transfer learning in the integration of CT-MRI data.

AlexNet

AlexNet is a convolutional neural network (CNN) architecture distinguished by its eight-layer deep structure. This architecture, which is designed especially for image classification problems, consists of five convolutional layers followed by three fully linked layers.

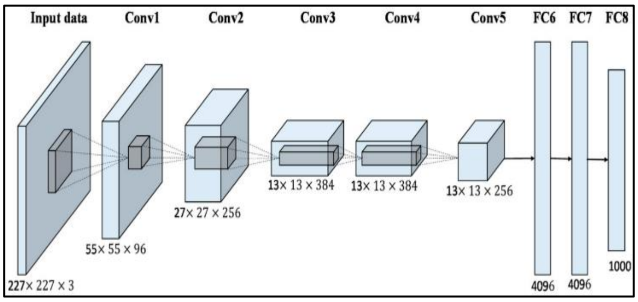

Fig. 3. ALEXNET deep convolutional neural network architecture [5]

The model was created by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, as seen in Fig. 3. It received a lot of praise for its performance in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Being among the first models to use Graphics Processing Units (GPUs) for training deep neural networks, AlexNet was crucial in making convolutional neural networks and deep learning more widely used. Transfer learning may be based on pre-trained models, such as AlexNet [17, 18]. Utilizing the information gained from large-scale image classification tasks, a pre-trained AlexNet may be fine-tuned for a particular medical image fusion job using a smaller dataset. For medical imaging researchers and practitioners, having access to such pre-trained models streamlines the process and makes it easier for them to modify the architecture to suit their unique needs rather than having to start from scratch.

DENSENET

Densely Connected Convolutional Networks, or DenseNet for short, is a unique kind of convolutional neural network (CNN) architecture that departs from traditional CNN topologies. DenseNet, developed by Gao Huang, Zhuang Liu, and Laurens van der Maaten, is characterized by a dense connection pattern in which all layers are feed forwardly linked to all other layers. By encouraging feature reuse and improving gradient flow during training, this architecture facilitates smooth information flow across the network. Multiple dense blocks, each including a group of densely linked layers, make up DenseNet. The feature maps from earlier layers are concatenated inside each dense block and used as input for later layers. Compared to conventional topologies, this dense connection results in a more parameter-efficient and computationally powerful network. Like AlexNet, DenseNet has shown successful in a variety of computer vision applications, most notably picture categorization. Its design improves gradient flow, reduces the vanishing gradient issue, and makes the model more compact.

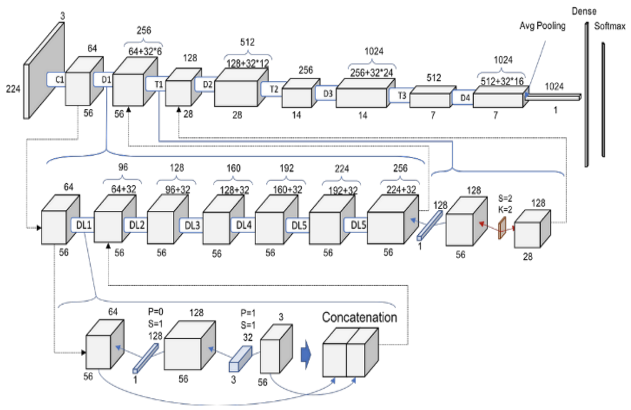

Fig. 4. DENSENET deep convolutional neural network architecture [18]

As shown in fig.4 model is application in medical image fusion tasks, particularly in conjunction with transfer learning, can capitalize on its inherent benefits. Pre-trained DenseNet models can serve as a robust starting point for medical imaging tasks, allowing researchers and practitioners to fine-tune the network on a smaller dataset related to specific fusion requirements. Leveraging the knowledge encoded in the pre-trained models aids in achieving better performance and faster convergence on medical image fusion tasks, like the role played by pre-trained models such as AlexNet in other domains.

3. Proposed Multimodal Fusion Process

Transfer learning techniques have shown to be very effective in automatically extracting representative and hierarchical characteristics at different levels of abstraction. VGG19, AlexNet, and DenseNet are standard models for image fusion that are often used. Using certain feature extraction methods, feature-level fusion entails removing features from input pictures to fuse them. This procedure isolates and uses pertinent features or qualities from the original pictures for the fusion process.

Our study intends to develop a hybrid approach that incorporates structural and functional information from PET and MRI scans, producing a unique fused picture, to produce a more robust and controlled fusion choice. The suggested method combines the transfer learning VGG19 architecture with the discrete wavelet transform for multimodal medical picture fusion. Additionally, the AlexNet architecture applies the concepts of transfer learning networks.

picture processing entails running the picture through a series of filters to calculate the Discrete Wavelet Transform (DWT). The picture is first passed through a low-pass filter and then a high-pass filter. The objective of this dual-filtering process is to extract from the picture the detailed (d), diagonal (h), vertical (v), and approximation (a) characteristics. While the high-pass filter concentrates on smaller, more precise information, the low-pass filter catches wider, more general qualities. When combined, these filters aid in the image's breakdown into its component parts, enabling a depiction that incorporates both the general structure and more subtle details.

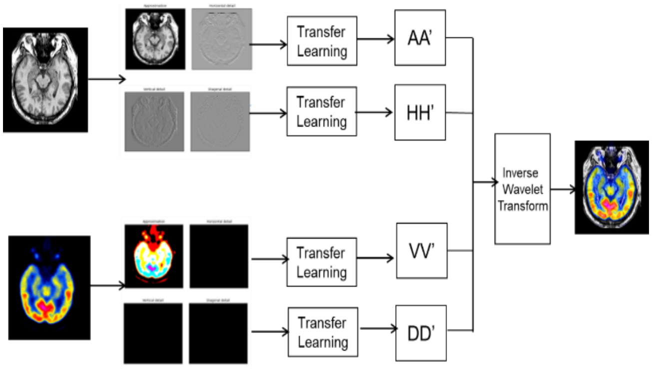

Fig. 5. PET-MRI Medical Image Fusion System

In this work, the pre-trained VGG19, AlexNet, and DenseNet models were trained using multimodal medical data, namely PET and MRI, as input source pictures. The PET and MRI images were first broken down using the Discrete Wavelet decomposition approach in the transform domain. Approximation (a), horizontal (h), vertical (v), and diagonal (d) coefficients were obtained from the decomposition of MRI images. Similarly, as shown in Fig. 5, the decomposition of PET images yielded the coefficients a' (approximation), h' (horizontal), v' (vertical), and d' (diagonal).

The lower-level coefficients (a and a'), horizontal coefficients (h and h'), vertical coefficients (v and v'), and diagonal coefficients (d and d') of the CT and MRI pictures were then used to extract features. The VGG19 model was then utilized to merge these characteristics. From the input coefficients, the VGG19, AlexNet, and DenseNet models produced the A, H, V, and D bands. Essential information from the original picture was caught in the A band, while details were acquired in the H, V, and D bands, which corresponded to horizontal, vertical, and diagonal directions.

Lastly, the inverse discrete wavelet transform was used to create the fused picture after the frequency sub-bands A, H, V, and D were integrated using the VGG19, AlexNet, and DenseNet architecture.

4. Experimental Parameters

Image fusion is a process that involves combining information from multiple images to create a single, more informative image. To assess the quality of fused images, various evaluation metrics are employed [14,15]. The selection of a specific metric depends on the goals and requirements of the fusion task. Commonly used evaluation metrics for image fusion include:

Entropy: Entropy measures the amount of information or uncertainty in an image. Higher entropy indicates greater information content. Entropy-based metrics can be used to evaluate the amount of information preserved in the fused image.

Mutual Information (MI): Mutual information measures the statistical dependence between two variables. In image fusion, it assesses the shared information between the source images and the fused image. Higher mutual information values indicate better fusion.

Structural Similarity Index (SSI): SSI compares the structural information between the source and fused images. It considers luminance, contrast, and structure, providing a more comprehensive assessment of image quality.

Peak Signal-to-Noise Ratio (PSNR): PSNR (Peak Signal-to-Noise Ratio) measures the ratio of the maximum possible power of a signal to the power of corrupting noise. Higher PSNR values are indicative of better image quality.

Mean Squared Error (MSE): For multimodal image fusion, when information from multiple images is combined, MSE (Mean Squared Error) can be calculated to quantify the difference between the fused image and a reference image.

Root Mean Square Error (RMSE): RMSE (Root Mean Square Error) calculates the average difference between the pixel values of the fused and reference images. Lower RMSE values are indicative of better image fusion.

RESULTS ANALYSIS

We used assessment metrics such PSNR, MSE, RMSE, SSI, MI, and EN for our extensive experimental tests on different PET and MRI test pictures. Taking into account the deep learning model that was used, we also incorporated the time required for the fusion of pictures.

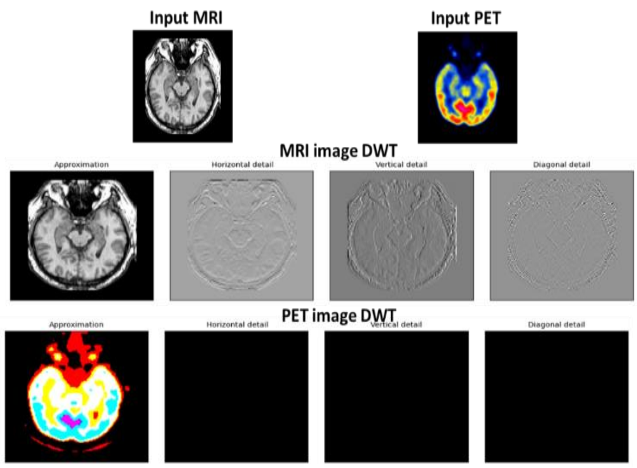

Fig. 6. DWT of PET & MRI

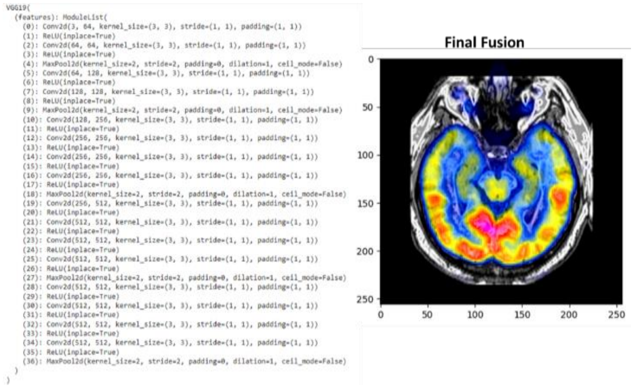

Fig. 7. VGG19 Fusion

As seen in Table 1, the fusion results are compared using MRI and PET input pictures as reference images. The comparison shows that the performance is better. The computed entropies for the input PET and MRI pictures, the fused image, and the combined entropy of the input and fused images are shown in Table 2.

When compared to other models that are currently available, our suggested technique, DWT+VGG19, performs very well in evaluation metrics like PSNR and SSIM, successfully collecting a significant amount of structural information in the pictures. Consequently, in terms of picture quality, our suggested technique for medical image fusion outperforms previous wavelet-based and neural network-based algorithms.

Table 1. DWT+VGG19 Fusion

Image Type	Parameters	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6
MRI	**MSE**	2317.09	2464.75	2929.01	2667.17	2046.22	1051.61
	PSNR	14.4	14.21	13.46	13.87	15.02	17.91
	RMSE	0.41	0.44	0.49	0.47	0.45	0.41
	SSIM	0.75	0.75	0.74	0.75	0.80	0.90
	MI	0.90	0.889	0.885	0.86	0.890	0.91
PET	**MSE**	5510.0	4197.8	3707.8	3415.20	3590.57	3389.92
	PSNR	10.7	11.9	12.43	12.79	12.57	12.82
	RMSE	0.89	0.71	0.64	0.65	0.69	0.92
	SSIM	0.60	0.61	0.63	0.65	0.69	0.72
	MI	0.43	0.440	0.486	0.468	0.432	0.247
	Time	3.36 sec	2.85 sec	4.43 sec	3.41 sec	3.60 sec	3.85 sec

Table 2. DWT+VGG19 Fusion Entropy Analysis

No.	MRI Entropy	PET Entropy	Fusion Entropy	Joint Entropy
Img1	4.32	2.51	5.041	7.569
Img2	4.36	2.63	5.016	7.602
Img3	4.25	2.42	4.761	7.251
Img4	3.99	2.45	4.590	6.894
Img5	3.69	1.98	4.127	6.082
Img6	2.98	1.24	3.327	4.510

5. Conclusion

Finally, evaluations of the DWT+VGG19 fusion approach were conducted over a range of parameters for both MRI and PET imaging. Improvements in Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Root Mean Square Error (RMSE), Structural Similarity Index (SSIM), Mutual Information (MI), and computational time demonstrate the positive effects of our suggested fusion strategy. Notably, for both MRI and PET modalities, the fusion procedure performed better at keeping structural information and picture quality. The entropy study shown in Table 2 adds further evidence to the fusion technique's efficacy. It is crucial to remember that even if the suggested approach showed encouraging outcomes, there is still space for development.

To improve overall diagnosis accuracy, future research areas may concentrate on deep learning architecture exploration, fusion algorithm optimization for particular medical imaging applications, and integration of other modalities. Translating these discoveries into useful medical imaging solutions would also need a thorough validation study and an investigation of real-world clinical circumstances.

Author contributions

Conceptualization—Sonia Panesar and Dr. Amit Ganatra; Methodology—Sonia Panesar; Software—Sonia Panesar; Validation—Dr. Amit Ganatra; Formal Analysis— Sonia Panesar and Dr. Amit Ganatra; Investigation—Dr. Amit Ganatra; Resources—Sonia Panesar; data curation— Sonia Panesar; Writing—review and editing— Sonia Panesar and Dr. Amit Ganatra; Visualization—Sonia Panesar; supervision—Dr. Amit Ganatra; Project administration—Sonia Panesar.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Li Y, Zhao J, Lv Z, Li J. Medical image fusion method by deep learning. International Journal of Cognitive Computing in Engineering. 2021 Jun 1; 2:21-9.

[2] Zhang YD, Dong Z, Wang SH, Yu X, Yao X, Zhou Q, Hu H, Li M, Jiménez-Mesa C, Ramirez J, Martinez FJ. Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. Information Fusion. 2020 Dec 1; 64:149-87.

[3] Zeng, Guixi, et al. "Medical image fusion using a deep convolutional neural network." Journal of Healthcare Engineering, Volume 2018, Article ID 6417203, 2018.

[4] Jin, Y. et al. "Deep learning for medical image fusion: A multi-task network for improved diagnosis and operative planning." In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018.

[5] Ghafoorian, M. et al. "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation." In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017.

[6] Song X, Zhou F, Frangi AF, Cao J, Xiao X, Lei Y, Wang T, Lei B. Multicenter and Multichannel Pooling GCN for Early AD Diagnosis Based on Dual-Modality Fused Brain Network. IEEE Transactions on Medical Imaging. 2022 Jun 29;42(2):354-67.

[7] Faul, D. D., Varasteh, Z., Anselm, K., Kuhlmann, M. T., Giesel, F. L., Kratochwil, C., & Haberkorn, U. (2020). PET/MRI and SPECT/MRI in oncology: a guide to hybrid imaging. Cancer Imaging, 20(1), 52. doi:10.1186/s40644-020-00332-4

[8] Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., & Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. Neuroimage, 52(1), 290-301.

[9] Dou, Q., Chen, H., Yu, L., Qin, J., & Heng, P. A. (2019). Multilevel contextual 3D CNNs for false positive reduction in pulmonary nodule detection. IEEE Transactions on Biomedical Engineering, 66(12), 3434-3445.

[10] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600-612. doi:10.1109/TIP.2003.819861

[11] Zhang, Y., Wei, W., Shen, W., & Zhang, G. (2019). A Survey of Deep Learning-Based Multimodal Medical Image Fusion. Neurocomputing, 338, 321-333.

[12] Chen, W., Wang, L., Shen, L., Wang, J., & Yang, X. (2018). Deep Learning-based Multi-modal Fusion for Diagnosis of Alzheimer's Disease. Information Fusion, 41, 268-278.

[13] Moeskops, Pim, et al. "Deep learning for brain MRI segmentation: State of the art and future directions." NeuroImage, Volume 202, 2019, Article 116091.

[14] Fu J, Li W, Du J, Xu L. DSAGAN: A generative adversarial network based on dual-stream attention mechanism for anatomical and functional image fusion. Information Sciences. 2021 Oct 1; 576:484-506.

[15] Finn C, Xu K, Levine S. Probabilistic model-agnostic meta-learning. Advances in neural information processing systems. 2018;31.

[16] Li P, Gu J, Kuen J, Morariu VI, Zhao H, Jain R, Manjunatha V, Liu H. Selfdoc: Self-supervised document representation learning. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021 (pp. 5397-5407).

[17] Hermessi H, Mourali O, Zagrouba E. Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain. Neural Computing and Applications. 2018 Oct; 30:2029-45.

[18] Huang P, Tan X, Zhou X, Liu S, Mercaldo F, Santone A. FABNet: Fusion Attention Block and Transfer Learning for Laryngeal Cancer Tumor Grading in P63 IHC