



## Research article

# Alzheimer's disease stage recognition from MRI and PET imaging data using Pareto-optimal quantum dynamic optimization

Modupe Odusami<sup>a</sup>, Robertas Damaševičius<sup>a</sup>, Egle Milieškaitė-Belousovienė<sup>b</sup>, Rytis Maskeliūnas<sup>a,\*</sup>, for the Alzheimer's Disease Neuroimaging Initiative

<sup>a</sup> Faculty of Informatics, Kaunas University of Technology, Kaunas, Lithuania

<sup>b</sup> Department of Intensive Care, University of Health Sciences, Kaunas, Lithuania



## ARTICLE INFO

Dataset link: <http://adni.loni.usc.edu/>

## Keywords:

Alzheimer's disease  
Pareto optimization  
Deep learning  
Classification  
Image fusion  
Multimodal

## ABSTRACT

The threat posed by Alzheimer's disease (AD) to human health has grown significantly. However, the precise diagnosis and classification of AD stages remain a challenge. Neuroimaging methods such as structural magnetic resonance imaging (sMRI) and fluorodeoxyglucose positron emission tomography (FDG-PET) have been used to diagnose and categorize AD. However, feature selection approaches that are frequently used to extract additional data from multimodal imaging are prone to errors. This paper suggests using a static pulse-coupled neural network and a Laplacian pyramid to combine sMRI and FDG-PET data. After that, the fused images are used to train the Mobile Vision Transformer (MViT), optimized with Pareto-Optimal Quantum Dynamic Optimization for Neural Architecture Search, while the fused images are augmented to avoid overfitting and then classify unfused MRI and FDG-PET images obtained from the AD Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS) datasets into various stages of AD. The architectural hyperparameters of MViT are optimized using Quantum Dynamic Optimization, which ensures a Pareto-optimal solution. The Peak Signal-to-Noise Ratio (PSNR), the Mean Squared Error (MSE), and the Structured Similarity Indexing Method (SSIM) are used to measure the quality of the fused image. We found that the fused image was consistent in all metrics, having 0.64 SIMM, 35.60 PSNR, and 0.21 MSE for the FDG-PET image. In the classification of AD vs. cognitive normal (CN), AD vs. mild cognitive impairment (MCI), and CN vs. MCI, the precision of the proposed method is 94.73%, 92.98% and 89.36%, respectively. The sensitivity is 90.70%, 90.70%, and 90.91% while the specificity is 100%, 100%, and 85.71%, respectively, in the ADNI MRI test data.

## 1. Introduction

AD and the neurodegenerative processes that drive dementia have emerged as the leading global health issue, with far-reaching consequences for individuals, families, and healthcare systems. AD, which accounts for 60 - 80% of all cases of dementia, is the most common cause of cognitive decline and memory impairment in the elderly population [1]. Throughout the years, great scientific effort has been dedicated to deciphering the complicated mechanisms behind neurodegeneration [2], investigating possible biomarkers for early diagnosis [3], creating targeted therapies, and promoting risk reduction techniques.

\* Corresponding author.

E-mail address: [rytis.maskeliunas@ktu.lt](mailto:rytis.maskeliunas@ktu.lt) (R. Maskeliūnas).

<https://doi.org/10.1016/j.heliyon.2024.e34402>

Received 7 December 2023; Received in revised form 6 July 2024; Accepted 9 July 2024

Available online 15 July 2024

2405-8440/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

As the population ages, the prevalence of AD and other neurological disorders increases, which encourages scientists to study more deeply the biomarkers and environmental variables that influence the vulnerability of the disease [4]. Early diagnosis of AD is crucial for timely intervention and treatment. The combination of sMRI and FDG-PET through collaborative generative adversarial networks addresses this challenge by enabling a more comprehensive and accurate assessment, potentially facilitating earlier detection and intervention strategies. However, this has remained a challenge due to the complexity of disease progression. Multimodal neuroimaging methods were utilized to solve several limitations of AD diagnosis by combining MRI and FDG-PET through collaborative generative adversarial networks [5]. Although numerous therapeutic trials have produced promising results, the complexity of neurodegeneration requires more research in multimodal neuroimaging techniques to ensure timely initiation of treatment, maximize therapeutic efficacy, and minimize potential adverse effects.

Another important part of neurodegenerative research has been the search for accurate and accessible biomarkers [6]. Biomarkers provide crucial information on disease development, allowing early diagnosis and helping to track therapy responses. Cerebrospinal fluid analysis and neuroimaging methods such as FDG-PET and MRI have shown promise in diagnosis [7]. This, in conjunction with modern machine learning techniques, has enormous potential to enable quick intervention and facilitate data analysis.

FDG-PET and related methods are crucial in the investigation of AD, offering vital insights into the underlying molecular processes of the disease [8]. FDG-PET uses radiotracers to view and quantify particular biomarkers in the brain [9], providing a non-invasive way to investigate AD [10]. FDG-PET biomarkers combined with the power of machine learning help in early identification, disease progression monitoring, and distinguishing AD from other neurodegenerative diseases with comparable clinical signs [11].

The effectiveness of FDG-PET has allowed the identification of those at risk of developing Alzheimer's disease before major cognitive symptoms appear, providing new pathways for early intervention and potential disease-modifying treatment options [12]. Furthermore, the ability of FDG-PET to monitor changes in amyloid-beta and tau deposits over time provides vital information on disease progression, helping researchers describe the evolution of AD and its influence on cognitive decline. Although being a vital tool, this method has drawbacks, such as its high cost, limited availability, and exposure to ionizing radiation. Efforts are being made to provide more accessible and cost-effective PET tracers, as well as to improve image analysis techniques, to increase sensitivity and specificity [13].

FDG-PET imaging offers a way to visualize the metabolic activity of brain cells, helping to locate abnormal regions by measuring the level of glucose metabolism. FDG-PET can help diagnose and choose treatment for various neurological diseases by focusing on these alterations [14], thus becoming more valuable in distinguishing between normal and pathological findings, unlike structural magnetic resonance imaging (sMRI), which provides information on the anatomical structure of the brain exclusively [15]. The regions that are most useful for distinguishing the stages of AD according to their mean relative metabolic rate of cerebral glucose include the midcingulate cortex, the angular, superior temporal, middle frontal, and superior frontal gyri [16]. It is a reliable biomarker for AD, but it is insufficient to identify the earliest phases of MCI [17]. With sMRI, machine learning approaches are increasingly being employed to help analyze AD phases. These technologies could help reveal crucial details about the structure of the brain [18]. The ability of sMRI to show different patterns of brain atrophy makes it useful for identifying AD. The great resolution of the soft tissue on sMRI makes it much more helpful in the diagnosis of AD [19].

Combining FDG-PET imaging with sMRI features offers an additional method to classify AD into different stages [20,21]. Researchers can better understand the underlying mechanisms of the disease by combining these two modalities. This strategy might increase the precision of the early analysis and treatment results [22]. In AD research, multiscale fusion approaches that combine MRI and PET data have been developed to provide a comprehensive understanding of the pathophysiology of AD [23,24]. These techniques take advantage of the strengths of both imaging modalities and promise to reveal previously inaccessible information, leading to a better diagnosis, monitoring, and management of AD. The two-dimensional Fourier transform (FT) and the discrete wavelet transform (DWT) [25] were used in the fusion process, which combined MRI and PET images [26]. Furthermore, the method of medical image fusion was implemented using the multiscale decomposition of the medical image pyramid to improve human visual understanding [27]. SIMM achieved by the method was better compared to other methods.

Our main contribution is a novel technique to integrate Gaussian decomposition and the Laplacian Pyramid (LP) coupled with the Spatial Pulsed Neural Coupling Network (SPCNN) in the image-fusion task for AD stage classification. Before combining the sMRI and FDG-PET images, different weights were assigned to various spatial frequencies. This helped to emphasize the characteristics that were most important for classifying the stages of AD, and MVIT is used to train the fused image and is adapted to work with sMRI and FDG-PET images, rather than having to use separate models for each modality. The architectural hyperparameters of the MVIT model are optimized using Quantum Dynamic Optimisation (QDO) for higher performance in AD stage classification.

## 2. Review of related ML based methods

The stages of AD classification were achieved employing a Support Vector Machine (SVM) model, which was trained using both FDG-PET and sMRI images [28,29]. The SVM was trained using extracted Grey Matter (GM) from FDG-PET as a feature. The GM served as the basis for classification in the AD staging process. Similarly, the accurate inherent similarity that is shared between sMRI and FDG-PET modality data was captured to select complementary features for the SVM-based classifier for the classification of AD stages [30,31]. The characteristics of the brain region of interest (ROI) in the sMRI and FDG-PET data are extracted using a template [32]. Complementary features are further selected on the basis of a consistent metric, and SVM was used for the classification of AD stages. Based on t statistics, ROIs with the highest t values were identified from sMRI and FDG-PET images for the classification of AD [33]. To classify the stages of AD, a new composite image was generated by blending the region of brain GM tissue in sMRI and FDG-PET images using mask and registration coding techniques [34]. Researchers have classified individuals with AD according to

their GM density and glucose based on sMRI and FDG-PET, allowing a more comprehensive and accurate diagnosis of AD [35]. A three-dimensional convolutional network was used for the classification of AD stages, in which a Sparse autoencoder was added to the network to increase the description of the characteristics for classification [36]. Similarly, a hypergraph-based framework was used that introduced a specialized hypergraph-based regularization to explicitly represent the high-order relationship in MRI and FDG-PET modality for the classification of AD stages [35].

The requirement for the selection of multimodal features derives from the fact that the features received from many modalities are inherently interconnected and have a high degree of dimension [37]. To avoid overfitting and improve the generalizability of the results, highly correlated or duplicated features are frequently removed [38]. This is often achieved in supervised learning using feature selection methods. By eliminating irrelevant or redundant traits, these strategies such as the feature-weighted scheme [28], greedy search [39], sparsity regularization [40], relational regularization [41], maximum expectation [42], and random forest incorporation [43] help identify and retain the most informative information in multimodal neuroimaging. The goal of feature selection is to simplify the data and prioritize the most valuable features that can be used as input for machine learning algorithms, enabling the creation of effective classification or prediction models for AD. However, multiple imaging techniques, each with unique scales, resolutions, noise levels, and informational contents, are used in multimodal neuroimaging. It may be difficult to integrate information between modalities due to this variance [44]. Therefore, the distinctive qualities of each modality must be carefully considered when integrating multimodal neuroimaging data.

The current feature selection techniques used to choose complementary data from sMRI and FDG-PET have not been able to handle feature variability well [27]. Additionally, the selected feature subset is also inconsistent. These issues have impacted the general accuracy and reliability of these methods. A multiscale decomposition algorithm can be used to combine spatial information from multimodal images [27]. With the help of the algorithm, which divides the images into various scales, useful information can be extracted at each scale. A more complete comprehension of the images can be achieved by merging the data from each scale. The pyramid is used in the decomposition of medical imaging, which incorporates numerous scales [45]. This method aims to enhance human visual comprehension of medical images. The stages of AD were classified using the initial ResNet model, based on combined sMRI and FDG-PET images that were obtained using the DWT and FT methods, the model was trained [26]. For the purpose of classifying the stages of AD, the demon algorithm and DWT enabled the perfect fusion of sMRI and FDG-PET [46]. Accurate differentiation between various stages of AD has been shown to be possible with the help of this pair of imaging techniques. The distinct advantages and complementary features of each imaging modality are responsible for the effectiveness of this approach. 3D DWT and 3D moment-invariant features of the sMRI and FDG-PET images were fused to classify AD stages [47]. For the analysis of sMRI and FDG-PET data, the DWT was enhanced with the pre-trained VGG16 network. The final fused image was then constructed using the inverse DWT, which was used to evaluate a pre-trained ViT for AD classification [48]. The low frequency subbands and the high frequency subbands are fused using the maximum selection rule and the modified spatial frequency for AD classification [49]. The use of Laplacian LP and SPCNN demonstrated a good use of multimodal information [50]. ViT has achieved excellent results in image recognition and classification with a single transformer encoder to learn global representations [51]. The simple ViT model trains faster and better than the original ViT [52], the performance of ViTs saturates rapidly when scaled to be deeper and improves the accuracy of the image classification [53]. Although MViT allowed for light-weight global processing of information with transformers [54].

### 3. Methods

The proposed method uses a fused image from the T1-weighted MRI of the whole brain and the FDG-PET image from the ADNI-1 data set to identify different stages of AD, including AD in the dementia stage, MCI due to AD, and cognitive normal (CN), which is often known as healthy control. The main goal of this method is to recognize the different phases of AD. The integration of data from both imaging methods aims to improve the precision and dependability of AD diagnosis. sMRI and FDG-PET images are enhanced using the sharpening technique before being reduced to scales. By combining Gaussian and LP (GLP) and SPCNN, it is possible to fuse sMRI and FDG-PET images of the same class. Using fused images, MViT is trained to identify the stages of AD. Unfused sMRI and FDG-PET images are used to test the accuracy of the method. Therefore, by using unfused test data sets, the efficiency of MViT in diagnosing the stages of AD can be assessed. Fig. 1 shows the block diagram of our proposed technique.

#### 3.1. Data and image acquisition

To gather high-quality data for this study, T1-weighted magnetization prepared rapid gradient echo (MPRAGE) techniques, along with radio frequency field (B1) and image intensity nonuniformity correction (N3) imaging methods, were considered while downloading MRI sequences from the ADNI and OASIS databases. Men and women of 42 to 95 years of age were evaluated according to the clinical dementia rating (CDR) and the minimal state assessment, which are clinical selection criteria used in the ADNI database that provide information on overall cognitive status. Because there is no access to postmortem pathological data, in vivo biomarkers, such as amyloid deposition, were not included in our selection criteria. FDG obtained by co-registration of a structural scan anatomical image was considered for FDG-PET images downloaded from the ADNI database. The ADNI and OASIS databases allow users to obtain whole-brain sMRI and PET images in the NIFTI format. For the MRI from ADNI database, 167 subjects across AD, CN, and MCI stages that gives best representation was selected out of 195 AD, 714 CN, and 654 MCI with respect to class bias. Likewise for the FDG-PET from ADNI database 190 subjects across AD, CN, and MCI stages that gives best representation was selected out of 195 AD, 714 CN, and 654 MCI with respect to class bias. Whole-brain magnetic resonance imaging must be analyzed slice

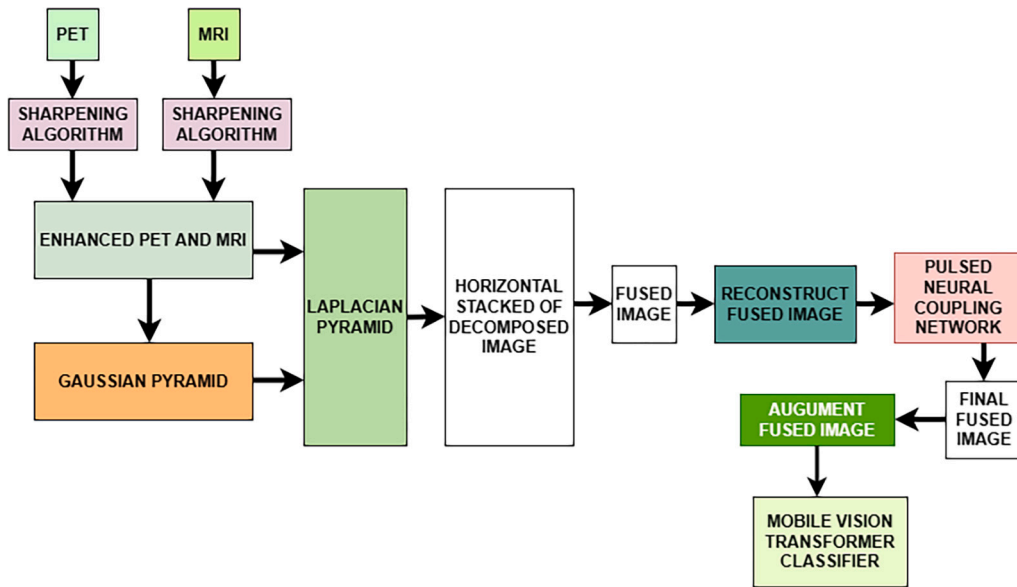


Fig. 1. Proposed Multimodal Approach Framework.

**Table 1**

Total Number of Slices Extracted from PET and MRI NIFTI Images for Fusion.

Group	MRI (ADNI)		
	AD(n=1000)	CN(n=1000)	MCI(n=900)
Sex(M/F)	60 /107	60 /107	60 /107
CDR Score	2	0	0.5, 1
MMSE Score	19 to 23	29 to 30	26 to 28
Age (mean)	65	65	65
Group	FDG-PET (ADNI)		
	AD(n=1400)	CN(n=1400)	MCI(n=800)
Sex(M/F)	80 /110	80 /110	80 /110
CDR Score	2	0	0.5, 1
MMSE Score	19 to 23	29 to 30	26 to 28
Age (mean)	63	63	63

ADNI - AD Neuroimaging Initiative; OASIS - Open Access Series of Imaging Studies; CDR - Clinical dementia rating, CN - Cognitive Normal, AD - Alzheimer Disease in dementia stage, MCI - Mild Cognitive Impairment, M - Male, F - Female, MMSE - Mini Mental State Examination.

by slice. Therefore, the slices needed for this research project are extracted using MRICroGL software, a free 3D viewer for medical imaging data that can display NIFTI files and allow for slice extraction. During the extraction of slices from the volume brain image using MRICroGL, the software preserves the original spatial resolution of the image, i.e. the extracted slices retained the same pixel dimensions and voxel sizes as the original volume. The selection of slices was based on a careful consideration of several factors aimed at ensuring the quality and relevance of the chosen MRI slices. A rigorous selection process was considered to include only the middle slices that contain the most relevant information for our research objectives. Specifically, this study focused on axial slices, which are known to provide a comprehensive view of the brain in a horizontal orientation from the base to the top of the skull. This choice is particularly beneficial for consistent anatomical comparisons across subjects and reduces the risk of partial volume effects that are more prevalent in other planes. The slices were chosen to demonstrate consistency across the class, considering the number of subjects involved in the study. The selected slices were required to exhibit optimal image quality. This criterion was essential to guarantee clear and accurate anatomical details in the MRI data. Slices with any artifacts or degradation in image quality were excluded from the final selection. Focusing on the middle slices of the MRI volume was a deliberate choice. These slices were selected because they offer a good representation of the overall brain anatomy. Middle slices are less susceptible to artifacts commonly found at the edges of the acquisition volume, ensuring a more reliable depiction of the brain structures. 167 subjects are selected per class for ADNI with an average of 6 slices each for MRI while an average of 190 subjects per class are selected per class for PET with an average of 7 slices per class. The total number of slices extracted for each of the classes is shown in Table 1.

## 4. Technical details of the implementation

### 4.1. Application of the Laplacian model for FDG-PET and sMRI

This study uses both image normalization and enhancement methods. Images are improved by using image filtering techniques, including kernel-based sharpening. Additionally, to guarantee uniformity across all photos, global normalization is used as the normalization step. To improve the sharpness and definition of an image's edges and details, a technique known as "kernel-based sharpening" is used. When the filter and the image are combined, the high-frequency elements of the image are amplified. Using a high-pass filter, it is possible to enhance high-frequency elements of an image. With the use of this filter, the low-frequency details in the image are carefully removed because they could contribute to blurriness or lack of sharpness, leaving only the high-frequency details. The image is thus crisper and more detailed. The high-pass filter or kernel is created using a mathematical model called the Laplacian. The second derivative of a function in a multidimensional space is described by the Laplacian operator. The Laplacian operator is discretized and approximated as a convolution kernel to create the kernel. The Laplacian equation serves as the basis for the kernel design. After that, the high-pass filter is implemented using this kernel. The generated kernel is then used with convolution to enhance high-frequency edges and features. Mathematically, the Laplacian model for FDG-PET and sMRI is shown in Equations (1) and (2).

$$FDG - PET'(n, m) = FDG - PET(n, m) - [H_{fdg-pet}(n, m) * FDG - PET(n, m)] \quad (1)$$

$$sMRI'(n, m) = sMRI(n, m) - [H_{smri}(n, m) * sMRI(n, m)] \quad (2)$$

where  $FDG - PET'(n, m)$  and  $sMRI'(n, m)$  indicate the target image of FDG-PET and the target image of sMRI, respectively, while  $FDG - PET(n, m)$  and  $sMRI(n, m)$  represents FDG-PET source image (enhanced image) and the sMRI source image (enhanced image), respectively. The high pass filter kernel for FDG-PET and sMRI is indicated by  $H_{fdg-pet}(n, m)$  and  $H_{smri}(n, m)$ , respectively, while  $*$  denotes a 2D convolution operation.  $H_{smri}(n, m)$ ,  $H_{fdg-pet}(n, m)$  is given by  $[-1, -1, -1]$ ,  $[-1, 9, -1]$ ,  $[-1, -1, -1]$ . The Laplacian kernel is a 3x3 filter with a center weight of 9 and surrounding weights of -1. High Pass Filters are applied to the FDG-PET and sMRI images, respectively, with the purpose of improving the high-frequency components while attenuating the low-frequency components. By convolving the original images with these high-pass filters, the fine details and edges within the images are emphasized, leading to improved image quality. Although there may indeed be visual similarities between kernel-based sharpening and standard windowing of contrasts, the underlying mechanisms and outcomes of these techniques diverge significantly. Kernel-based sharpening involves the application of advanced image processing algorithms that improve local details and edges by adjusting the pixel intensities based on neighboring information [55,56]. On the other hand, standard windowing primarily focuses on adjusting the image's brightness and contrast levels within specific intensity ranges to highlight specific features. Fig. 2 shows the original image and enhanced image for MRI and PET. Metrics such as the Peak Signal-to-Noise Ratio PSNR, SSIM, and MSE are employed to assess the improved and fused image quality. The combination of these parameters allowed us to fully evaluate the image quality. The quality of the images improves as the PSNR increases. The similarity between the two images is measured by SSIM. The quality of the image improves as the SSIM number approaches 1. The average squared difference between the base image and the enhanced image is what MSE calculates. The quality of the image improves with decreasing MSE.

### 4.2. Laplacian pyramid

LP functions at multiple scales and resolutions and is a very efficient image processing technique [57]. Images can be broken down into layers with varying resolutions by analyzing them at various scales, which allows for the view of details such as edges, textures, and other details. As the size decreases, these layers have more and more finer details. Consequently, this method enables a more thorough understanding of the content and structure of an image. The result of the reconstruction procedure can be obtained by fusing the matching layers of many source images (enhanced image) separately to create a fused pyramid. The existence of artifacts is one major flaw of LP. The quality of the fused images may suffer from these problems. Sumiya et al. [58] created a solution for the problem by introducing a simple and adaptable GLP. The GLP method applies a sequence of Gaussian filters to the image at every level, resulting in a reduced high-frequency content and eliminating any noise or artifacts. The adaptability of the method enables the user to balance the smoothness and detail of the image by customizing the blurring level and scale. When GLP is implemented, the image becomes clearer and more visually appealing. The image is converted into a Gaussian pyramid using Gaussian decomposition, and the Laplacian pyramid of the image is produced by a differential operation based on the Gaussian Pyramid (GP) [59,60].

### 4.3. SPCNN

A neural network known as SPCNN is used during the merging process of sMRI and FDG-PET images. This network is built to help neurons communicate with pulsed signals. The SPCNN can successfully blend data from numerous images to obtain a more complete and accurate output [61]. The network consists of neurons, each of which might represent a pixel of the input images. The pulsed signals of the neurons, which are delivered to the matching pixels, carry the information communicated between them. As a result, the network can analyze visual data by looking at patterns in pulsed signals. Pulsed neural coupling networks are frequently used in numerous tasks, including image segmentation, pattern recognition, target categorization, and image fusion [60]. The feedback network in the pulsed neural coupling network model is made up of several neurons coupled in a single layer [62]. Equations (3) and



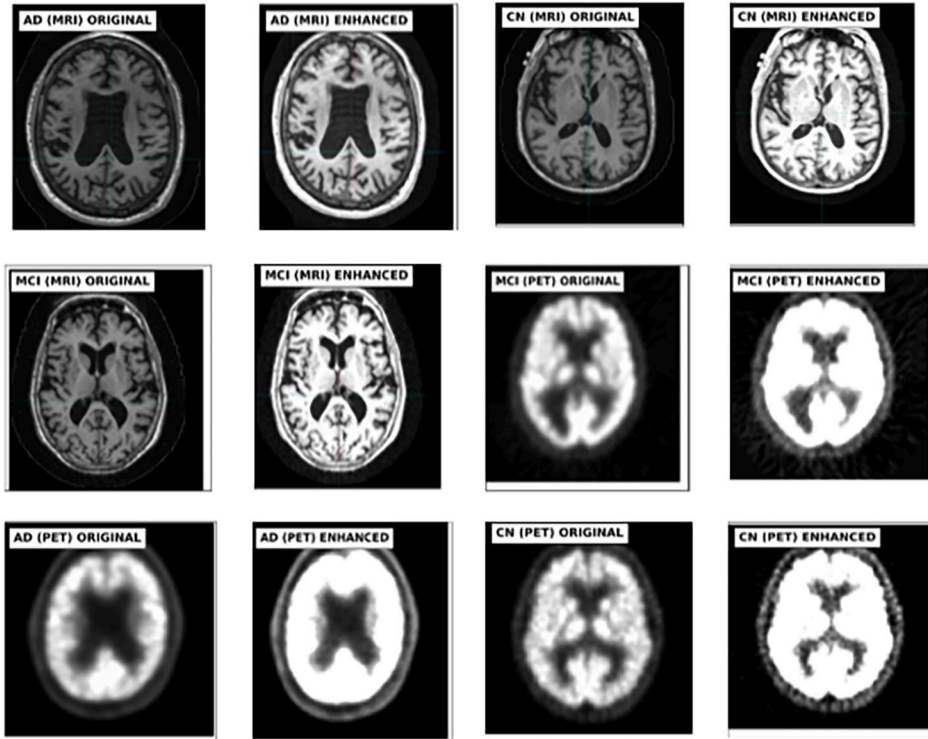


Fig. 2. Samples of sMRI and FDG-PET images before and after kernel-based sharpening. AD - Alzheimer Disease; MCI - Mild Cognitive Impairment; CN - Cognitive Normal; Original - Image before sharpening; Enhanced - Image after Sharpening.

(4) mathematically express the pulsed neural coupling network that consists of two matrices. This network incorporates the neural connection process and functions in a pulsed manner.

$$a(k, l) = m(k, l) + \beta * [b(k-1, y) + b(k+1, y) + b(k, l-1) + b(k, l+1)] \quad (3)$$

$$b_{\text{new}}(k, l) = b_{\text{old}}(k, l) * \exp(\theta) + a(k, l) * [1 - \exp(\theta)] \quad (4)$$

where:  $a(k, l)$  and  $b_{\text{new}}(k, l)$  = output images obtained after each iteration of the algorithm,  $b_{\text{old}}(k, l)$  = the current output image obtained before next iteration of the algorithm.,  $m(k, l)$  = source image from GLP,  $a$  and  $b$  = matrix which are initially zeros,  $k$  and  $l$  = spatial coordinates of the image  $\beta$  = decay factor  $\theta$  = threshold value.  $\exp$  = exponential function. The matrices  $a$  and  $b$  have the same size as the input image, that is, each element in these matrices corresponds to a pixel in the input image. These matrices are updated at each iteration based on equation (3) and (4) respectively. Equation (4) involves a recursive step in which  $b(k, l)$  appears on both sides of the equation because the new value of the function is calculated based on the initial value. Equation (4) is a recursive definition for updating the values of the output image  $b(k, l)$ . On the left-hand side,  $b(k, l)$  represents the current value of the pixel at coordinates  $k, l$  in the output image  $b$ . The right-hand side of the equation calculates a new value for  $b(k, l)$  based on its current value and other factors. Each pixel's value is recursively updated based on its current value, which helps refine and enhance features in the output image across multiple iterations of the algorithm. Equations (3) and (4) are applied iteratively to produce the final fused image.

#### 4.4. MRI and PET image fusion

To combine MRI and PET images across several modalities, GLP and SPCNN are used. By merging the data from the two different types of images, the target area is represented more accurately and comprehensively. To provide a multiscale representation of the MRI and PET input images, two sets of GP are built. This method enables a more thorough study of the photos at various scales. The multiscale depiction of GP makes it easier to spot patterns and characteristics that might not be discernible on a single scale. The Gaussian pyramids are used to create Laplacian pyramids, and between each level of the pyramid and its extended version, a subtraction operation is carried out. At each level of the pyramid, the left and right halves of the images are connected using the horizontal stack method. The concatenation of arrays along the horizontal axis results in the final merged image. After upsampling and integrating the merged images, the final image is reconstructed using the Laplacian pyramid, and then SPCNN is used. SPCNN further processes and enhances the fused image, potentially improving its accuracy and quality. It also provides the spatial relationships between the corresponding regions in the MRI and PET images. This network takes into account the local context and spatial

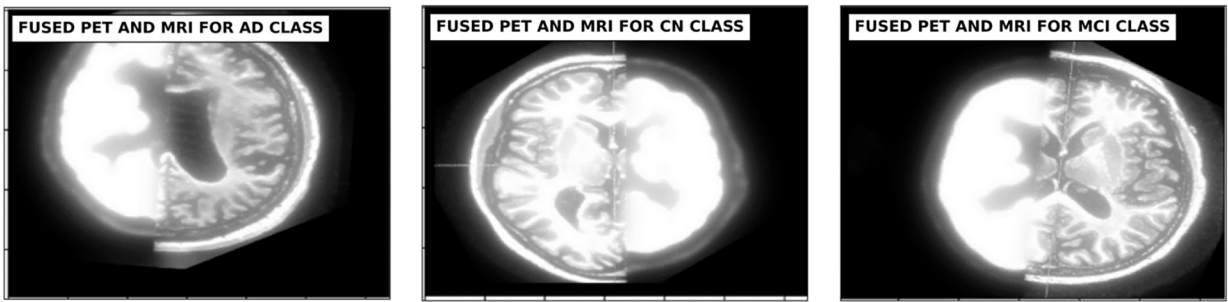


Fig. 3. Outcome of Gaussian Laplacian pyramid methods.

information. However, its effectiveness with the GLP is highly dependent on the quality of the input images. To obtain the final fused image seen in Fig. 3, the SPCNN output is normalized to the range [0, 255].

#### 4.5. Augmentation

A set of image transformations [63], was applied to the sMRI, FDG-PET, to increase the data set. The applied transformations were implemented using the PyTorch ‘transforms’ module, which provides a range of image manipulation. The following transformations were used: *RandomRotation*(45) rotates the image by a degree between -45 and +45 degrees, *RandomHorizontalFlip*( $p = 0.5$ ) flips the image horizontally at random with a probability of 0.5, and *RandomVerticalFlip*( $p = 0.5$ ) randomly flips the image vertically with a probability of 0.5.

#### 4.6. Application of deep learning using MViT

The MViT model is a variation of the ViT model created especially for use on mobile devices [64]. The architecture is designed with several key features:

- **Multiscale Feature Hierarchies.** There are many channel resolution scaling steps in MViT. The stages increase the channel capacity while decreasing the spatial resolution in a hierarchical manner, starting with the input resolution and a modest channel size. As a result, a multiscale pyramid of features is formed, with the early levels modeling basic low-level visual information at high spatial resolution and the deeper layers at spatially coarse but complex high-dimensional characteristics.
- **Decomposed Relative Positional Embeddings.** MViT incorporates decomposed relative positional embeddings. This technique encodes the relative position between two input elements, which only depends on the relative location distance between tokens in the pooled self-attention computation, important for maintaining shift invariance in vision tasks.
- **Residual Pooling Connections.** MViT uses residual pooling connections, which adds the pooled query tensor to the output sequence, increasing information flow and facilitating the training of pooling attention blocks in MViT.
- **Pooling Attention.** MViT uses a clustering attention mechanism to reduce the computation and memory complexity in attention blocks, which is particularly important for high-resolution object detection and space-time video understanding tasks, although it also significantly increases the computation complexity due to the quadratic complexity of self-attention operators in transformers.
- **Hybrid Window Attention.** MViT employs a hybrid window attention scheme that can complement pooling attention for better accuracy/compute trade-off, which is needed for object detection tasks that require high-resolution inputs and feature maps.

In addition to these features, MViT has the ability to effectively handle limited data scenarios, in contrast to traditional computational neural networks (CNNs). MViT’s efficiency lies in its ability to capture long-range dependencies and spatial relationships using self-attention mechanisms, while requiring fewer computational resources compared to CNNs [65]. Taking advantage of the transformer architecture, MViT can efficiently process and extract meaningful features from small datasets, making it an appropriate choice for our study, where limited data availability is a crucial factor.

The 256x256 input images were used in our suggested model, and a list of numbers (96, 120, 144) was used to define the hidden dimensions at various resolution levels. For each layer of the model, we also specified the number of channels using a list of integers (16, 32, 48, 48, 64, 64, 80, 80, 96, 96, 384). The output of the MViT model is produced by combining a conventional 3X3 convolutional layer and a 1X1 convolutional layer. The 1X1 convolutional layer converts the input tensor into a high-dimensional space by learning linear combinations of the input channels, while the 3X3 convolutional layer collects local spatial features. Our implementation of the proposed model was developed in Python using the PyTorch framework, an open source machine learning library built on top of Tensorflow. The suggested approach for AD image classification demonstrates efficient and accurate processing, which requires only 1.2 seconds per image on a computer with an Intel® Core™ i5-6400 processor, 8 GB of RAM and Windows 10. Compared to using only the CPU, using a dedicated GPU (we tested on Nvidia 1650) for greater parallel processing capabilities can boost the computation speed for AD picture classification jobs by tenfold or more, which makes the approach suitable for a wide range of computer settings.

**Table 2**  
Total Number of Slices Extracted from PET and MRI NIFTI Images for test data.

Group	MRI (ADNI)		
	AD(n=100)	CN (n=100)	MCI (n=100)
Sex(M/F)	6 /10	6 /10	6 /10
CDR Score	2	0	0.5, 1
MMSE Score	19 to 23	29 to 30	26 to 28
Age (mean)	65	65	65
Group	FDG-PET (ADNI)		
	AD(n=100)	CN(n=100)	MCI(n=100)
Sex(M/F)	8 /11	8 /11	8 /11
CDR Score	2	0	0.5, 1
MMSE Score	19 to 23	29 to 30	26 to 28
Age (mean)	63	63	63
Group	MRI (OASIS)		
	AD(n=10)	CN(n=10)	MCI(n=14)
Sex(M/F)	5 /5	5 /5	5 /5
CDR Score	2	0	0.5, 1
MMSE Score	19 to 23	29 to 30	26 to 28
Age (mean)	63	63	63

ADNI - AD Neuroimaging Initiative; OASIS - Open Access Series of Imaging Studies; CDR - Clinical dementia rating, CN - Cognitive Norm, AD - Alzheimer Disease in dementia stage, MCI - Mild Cognitive Impairment, n-Number of slices.

Before training, hyperparameters are established to assist in the learning process and the refinement of the model through the adjustment of various hyperparameters, including batch size (ranging from 8 to 32), learning rate (ranging from 0.01 to 0.0003), number of epochs (ranging from 10 to 30), optimators (Adam, AdamW, and stochastic gradient descent), and weight decay (0.5 and 0.05). The optimal parameters of our proposed model were determined to be a weight decrease of 0.05, an SGD optimizer, and 30 epochs, while the learning rate for each binary classification was set at 0.002. The training data set consisted of 60% of the fused images, and the remaining 40% was used for validation. Data augmentation, regularization (dropout=0.3), and learning rate scheduler was utilized to control over fitting. Additionally, an independent test data set was used comprising individuals from the ADNI and OASIS databases to validate the proposed model. Only demographic criteria was used to create the test data and slices were extracted from each subject and the selection of specific slices was performed using the MRICroGL software. The test data set contained no fused images and no splitting is required as the data are independent data from ADNI and OASIS. The total number of slices extracted for each of the classes is shown in Table 2.

Finally, the sensitivity, specificity, and accuracy of the model in predicting the test images were evaluated.

#### 4.7. Neural architecture search

The Neural Architecture Search (NAS) uses a search strategy to find the next architecture to be evaluated, and an evaluation strategy to assess its performance [66,67]. Formally, NAS problem can be defined as follows:

$$\begin{aligned}
 & \underset{A \in \mathcal{A}}{\text{minimize}} && L(A; D_{\text{train}}) \\
 & \text{subject to} && L(A; D_{\text{val}}) \leq \epsilon, \\
 & && \text{Cost}(A) \leq \kappa,
 \end{aligned} \tag{5}$$

Here:  $A$  represents the architecture of the neural network.  $\mathcal{A}$  is the search space for all possible architectures.  $L(A; D_{\text{train}})$  is the loss function that measures the performance of architecture  $A$  in the training data set  $D_{\text{train}}$ .  $L(A; D_{\text{val}})$  is the loss function that measures the performance of architecture  $A$  in the validation data set  $D_{\text{val}}$ .  $\epsilon$  is the maximum allowable validation loss.  $\text{Cost}(A)$  is a function that measures the computational cost of architecture  $A$ .  $\kappa$  is the maximum computational cost allowable. The goal of NAS is to find an architecture  $A$  that minimizes training loss, while also ensuring that validation loss and computational cost do not exceed their respective thresholds.

#### 4.8. Pareto-optimal QDO algorithm

Pareto-Optimal QDO Algorithm refers to an optimization algorithm that uses principles of quantum computing and Pareto optimality. Quantum computing leverages quantum bits (qubits) to perform computations, potentially offering exponential speed-up for certain tasks [68]. Pareto optimality, on the other hand, is a concept in multi-objective optimization where a solution is considered Pareto optimal if there is no other solution that can improve one objective without worsening at least one other objective [69].



The {(QDO) [70], is an iterative optimization algorithm based on a quantum dynamic process. The pseudocode of the algorithm is presented in Algorithm 1. Here, the population  $P$  consists of  $N$  individuals, each representing a potential solution. The fitness  $f_i$  of an individual  $i$  is evaluated and the best solution  $P_{\text{best}}$  is updated if  $f_i$  is better. The velocity  $v_i$  of each individual is updated using a quantum dynamic equation, which is not specified here because it would depend on the specific implementation of the QDO algorithm. The position  $x_i$  of each individual is updated by adding the velocity to the current position. Gaussian sampling is applied to  $x_i$  to introduce randomness and promote exploration of the search space. The position  $x_i$  is discretized using a threshold function to convert the continuous position into a discrete solution. The worst individual in the population is replaced by the mean of the population to maintain diversity and avoid premature convergence. The algorithm continues until a termination condition is met, such as reaching a maximum number of iterations or achieving a desired level of fitness. The best solution found is then returned.

---

**Algorithm 1** Quantum Dynamic Optimization Algorithm.

---

```

1: Initialize population  $P$  with  $N$  individuals
2: while termination condition not met do
3:   for each individual  $i$  in  $P$  do
4:     Evaluate fitness  $f_i$  of individual  $i$ 
5:     Update best solution  $P_{\text{best}}$  if  $f_i$  is better
6:     Update velocity  $v_i$  using quantum dynamic equation
7:     Update position  $x_i$  using  $x_i = x_i + v_i$ 
8:     Apply Gaussian sampling to  $x_i$ 
9:     Discretize  $x_i$  using a threshold function
10:  end for
11:  Replace worst individual with mean of population
12: end while
13: return best solution  $P_{\text{best}}$ 

```

---

A quantum dynamic equation used in the QDO algorithm is a simplified version of the Schrödinger equation [71], which is a fundamental equation in quantum mechanics:

$$z\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}, t) = \left[ -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}, t) \right] \Psi(\mathbf{r}, t) \quad (6)$$

here

- $z$  is the imaginary unit, which is defined as  $z = \sqrt{-1}$ .
- $\hbar$  is the reduced Planck constant, which is a fundamental constant in quantum mechanics. It is related to the Planck's constant  $h$  by  $\hbar = \frac{h}{2\pi}$ .
- $\frac{\partial}{\partial t}$  is the partial derivative with respect to time  $t$ . This operator is used to calculate the rate of change of the wave function with respect to time.
- $\Psi(\mathbf{r}, t)$  is the wave function of the system, which is a function of position  $\mathbf{r}$  and time  $t$ . The wave function contains all the information about the state of a quantum system.
- $-\frac{\hbar^2}{2m} \nabla^2$  is the kinetic energy operator, where  $m$  is the mass of the particle and  $\nabla^2$  is the Laplacian operator. The Laplacian operator is a differential operator that calculates the divergence of the gradient of a function.
- $V(\mathbf{r}, t)$  is the potential energy, which can depend on position  $\mathbf{r}$  and time  $t$ . The potential energy represents the energy caused by the position of the particle in a potential field.

Equation (6) describes how the wave function of a quantum system evolves over time. In the context of a QDO algorithm, the wave function could represent the state of the optimization problem, and the potential energy could represent the fitness function. The algorithm would then use this equation to update the state of the system in each iteration, with the goal of finding the state that minimizes the fitness function. In a QDO, these variables would be interpreted as follows: The wave function  $\Psi(\mathbf{r}, t)$  represents the state of the optimization problem. The potential energy  $V(\mathbf{r}, t)$  represents the fitness function of the optimization problem. The kinetic energy operator represents the exploration of the search space. The time evolution of the wave function represents the iterative process of the optimization algorithm.

#### 4.9. Pareto-optimality of QDO algorithm

In the context of the QDO Algorithm, Pareto-optimality could be used to guide the search process when there are multiple conflicting objectives to optimize. For example, in a neural architecture search problem, one might want to minimize both the prediction error and the complexity of the neural network architecture. These two objectives are typically conflicting, as reducing the prediction error often requires more complex architectures, while reducing the complexity often leads to higher prediction error. The QDO algorithm can be adapted to handle such multi-objective problems by maintaining a population of solutions and using Pareto dominance to guide the selection of solutions. Specifically, a solution for reproduction would be selected if it is not dominated by any other solution population in terms of all objectives.

A solution  $x^*$  is Pareto-optimal if there does not exist another solution  $x$  such that  $f_i(x) \leq f_i(x^*)$  for all objectives  $i$  and  $f_j(x) < f_j(x^*)$  for at least one objective  $j$ , where  $f_i(x)$  is the value of the  $i$ -th objective function at solution  $x$ . Here:

- $x^*$  is a potential Pareto-optimal solution. This is a solution that we are considering as a candidate for Pareto-optimality.
- $x$  is another solution. This is a solution that we compare with the candidate solution  $x^*$ .
- $f_i(x)$  is the value of the  $i$ -th objective function at solution  $x$ . This represents the performance of solution  $x$  with respect to the  $i$ -th objective.
- The inequality  $f_i(x) \leq f_i(x^*)$  means that the  $i$ -th objective value of solution  $x$  is not worse than that of solution  $x^*$ . In other words, solution  $x$  does not perform worse than solution  $x^*$  with respect to the  $i$ -th objective.
- The inequality  $f_j(x) < f_j(x^*)$  means that the  $j$ -th objective value of solution  $x$  is better than that of solution  $x^*$ . In other words, solution  $x$  performs better than solution  $x^*$  with respect to at least one objective.
- If both of the above conditions are met, then solution  $x^*$  is not Pareto-optimal, because there exists another solution  $x$  that performs equally well or better in all objectives and better in at least one objective. If no such solution  $x$  exists, then solution  $x^*$  is Pareto-optimal.

#### 4.10. Ablation study

To perform an ablation study on proposed approach for the classification of stages of AD using fused images that combine sMRI and FDG-PET, we systematically remove specific components of the approach to evaluate the contribution of each segment of the approaches to the overall performance. Firstly, MViT is trained on fused image without sharpening and SPCNN. Secondly, MViT is trained on images fused without sharpening. Lastly, MViT is trained on images fused without SPCNN. The advantage of conducting the ablation in this manner lies in its ability to determine whether the sharpening algorithm and SPCNN fusion technique are essential components for improving the quality and informativeness of the fused image in classification of AD stages.

#### 4.11. Visualization of MViT based multimodal fused images with gradient class activation map

All fused images were inputted to train the MViT model. For visualizing the region using the MViT model used for the decision in classification, we used the Gradient Class Activation Map (Grad-CAM). CAM of the target class is usually upsampled to the size of the input image to obtain fine-grained pixel-scale representations [72]. The idea of CAM is to generate the output map instead of a value that represents the probability of AD [73]. Grad-CAM is a class-discriminative localization technique that generates visual explanations for any CNN-based network without requiring architectural changes or re-training [74]. It weighs the 2D activations by the average gradient [75].

## 5. Results

The experimental environment used for this study is PyTorch deep learning framework with the Jupyter notebook based on the Ubuntu 16.04 operating system and the Windows 10 operating system. Preprocessing techniques were implemented on windows 10. The software used is Python 3.9.11 and the MRICroGL software version 1.2. Using three metrics, namely SSIM, PSNR, and MSE, the quality of the fused images was assessed. SSIM measure the similarity between two images, and a value close to 1 or exactly indicates a perfect similarity, PSNR compares two images in terms of their pixel values in which higher PSNR values indicate a higher quality image, while MSE provides a quantitative measure of the overall difference between the two images. The results of the experiments are reported at the subject level to ensure a comprehensive and cohesive presentation. The analysis involves the fusion of information from selected slices for each individual subject. The concatenation method is used to combine the relevant information from different slices within each subject. Specifically, the results from individual slices for a given subject are merged to create a comprehensive summary of that subject's data. Furthermore, how each value for each subject differs from each other is not significant, so the average of the values for all subjects in a class is considered as the final value. Table 3 shows the quantitative evaluation values for the fusion technique in different AD classes.

Our study involved putting into practice and contrasting the Simple ViT and Deep ViT models against the suggested MViT model with Pareto optimization. The hyperparameters used for the different models are depicted in Table 4. The hyperparameters were tuned in this manner: batch size = [8,16,32], learning rate = [0.01, 0.001, 0.0001, 0.02, 0.002, 0.0002, 0.03, 0.003, 0.0003], optimizers = [Adam, AdamW, stochastic gradient descent (SGD)], weight decay = [0.5,0.05]. The best parameters for our proposed model were found as optimizer = SGD and weight decay = 0.05, while the learning rate for each of the binary classifications is 0.002 for the binary classes.

The MViT model integrated the merged MRI-PET images and had a resolution of 256 x 256 pixels for its input image, as shown in the proposed fusion approach, to maintain the same resolution as the original images before fusion to ensure the fidelity of the fused result. The model architecture includes 11 ViT layers, each containing a series of transformer blocks resembling the original ViT design. These layers have different patch sizes, which range from 96 to 144 tokens. Spatial information is extracted using convolution layers and then combines the output of both the final convolutional layer and the final ViT layer to pass through an embedding layer. Two hidden layers of Multilayer Perception (MLP) and a final linear layer process the embedding, resulting in the production of output logits. The model is trained on fused MRI and PET images to classify the images into two distinct categories based on the MRI test images and the PET test images, which produces a shape tensor (batch size, 2) as output. To implement all variants of the

**Table 3**  
Assessment of Proposed Fusion Technique's Effectiveness on Various MRI and PET Image Categories.

Pyramid Level	Class	SIMM		
		Original/Enhanced/SD	PSNR(db) Original/Enhanced/SD	MSE Enhanced
Level 6	AD (PET)	0.56 / 0.64/0.01	30.5 / 35.60/0.003	0.21
	CN (PET)	0.55 / 0.60/0.008	30.5 / 34.10/0.002	0.22
	MCI (PET)	0.55 / 0.61/0.012	30.5 / 34.10/0.002	0.22
	AD (MRI)	0.50 / 0.60/0.011	30.5 / 35.50/0.003	0.22
	CN (MRI)	0.53 / 0.66 /0.001	31.2 / 35.30/0.001	0.20
Level 5	MCI (MRI)	0.52 / 0.64/0.001	30.5 / 34.2/0.003	0.21
	AD (PET)	0.44/ 0.48/0.010	20.4 / 22.20/0.002	0.30
	CN (PET)	0.40 / 0.50/0.001	21.4 / 22.20/0.001	0.30
	MCI (PET)	0.40 / 0.50/0.040	21.4 / 22.20/0.003	0.30
	AD (MRI)	0.39 / 0.48/0.010	20.9 / 22.20/0.001	0.30
Level 4	CN (MRI)	0.38 / 0.46/0.010	20.9 / 22.20 /0.001	0.30
	MCI (MRI)	0.40 / 0.44/0.010	20.4 / 22.20/0.001	0.30
	AD (PET)	0.38 / 0.40/0.011	20.9 / 20.10/0.001	0.35
	CN (PET)	0.38 / 0.42/0.010	20.9 / 20.10 / 0.001	0.35
	MCI (PET)	0.33 / 0.40/0.010	30.1 / 20.10/ 0.003	0.34
	AD (MRI)	0.32 / 0.45/0.012	30.1 / 20.20/0.001	0.35
	CN (MRI)	0.33 / 0.40/0.010	30.1 / 20.10/0.001	0.35
	MCI (MRI)	0.31 / 0.38/0.001	30.1 / 20.10/0.001	0.35

SIMM-Structured Similarity Indexing Method; PSNR-Peak Signal-to-noise Ratio; MSE-Mean Squared Error; AD - Alzheimer Disease; CN - Cognitive Normal; MCI - Mild Cognitive Impairment; SD-Standard Deviation.

**Table 4**  
Hyperparameters of various ViT models and the proposed model.

Hyperparameters	Simple ViT	Deep ViT	Mobile ViT
Learning rate	0.0002	0.0001	0.0002
Optimizer	SGD	SGD	SGD
Number of Layers in MLP	2	2	2
Batch size	16	16	16
Dropout	0.1	0.4	0.3
Epochs	100	100	100
Learning Rate Schedule	P = 2, M = min	P = 2, M = min	P = 2, M = min

ViT - Vision Transformer; MLP - Multilayer perceptron; SGD - Stochastic Gradient Descent, P - Patience, M - Mode.

ViT model, Adam, AdamW, and SGD optimizers were used, and several epochs ranging from 0 to 30 were adjusted accordingly. The models were trained from scratch on the task of classifying images generated from fused MRI-PET data. While transfer learning can be beneficial in certain contexts, we chose to train our models from scratch due to the unique characteristics of our dataset and the specific requirements of our classification task using the SGD optimizer and a learning rate of 0.0002, the validation accuracies for CN / MCI, AD / CN, and AD / MCI were 85.60%, 89.50%, and 78.40% respectively, which were achieved by the Simple ViT model, while the Deep ViT model attained validation accuracies of 80.60%, 86.50%, and 70.50% in CN / MCI, AD / CN, and AD / MCI respectively. Table 5 displays the validation precisions achieved by the Mobile ViT model proposed in our study, which were 97.5%, 99.16%, and 96.25% for CN / MCI, AD / CN, and AD / MCI respectively. These results were obtained using an SGD optimizer and a learning rate of 0.0002. The proposed MViT model validation accuracy and validation loss curves on fused images of sMRI and FDG-PET as shown in Fig. 4, and Fig. 5 respectively.

The method we proposed achieved excellent levels of precision, sensitivity, and specificity for the classification of AD / CN, AD / MCI, and CN / MCI in ADNI MRI test data. Our method demonstrated a precision of 89.36%, 94.73% and 92.98% for CN / MCI, AD / CN and AD / MCI, respectively. Similarly, sensitivity to 90.91%, 90.70%, and 90.70% and sensitivity to 85.71%, 100%, and 100% for the same binary classification, respectively.

Table 6 summarizes the performance of our proposed model on the test data.

The comparison of the proposed model with some currently used fusion approaches to classify AD / CN, AD / MCI, and CN / MCI is shown in Table 7.

Model based on Inception-ResNet and DWT classified AD / CN with 95.90% accuracy, 98.40% sensitivity and 92.20%. Multimodal MRI-PET investigation was carried out using ResNet 50 with DWT and provided 97.90% precision, 97.00% sensitivity, and 97.00% specificity for AD / CN classification. Similar results were obtained by our proposed model Mobile ViT with GLP, which provided 94.93%, 90.70%, 100% accuracy, specificity, and sensitivity for AD / CN. The accuracy of the Inception-ResNet and DWT model was 95.90%, which is higher than the accuracy of our proposed model, Mobile ViT with GLP, which received a score of 94.93%, but lower than the accuracy of ResNet 50 with the DWT model, which scored 97.00%. The Inception-ResNet and DWT models have lower sensitivity and specificity compared to the other two models. The DWT model combined with ResNet 50 achieved a sensitivity and

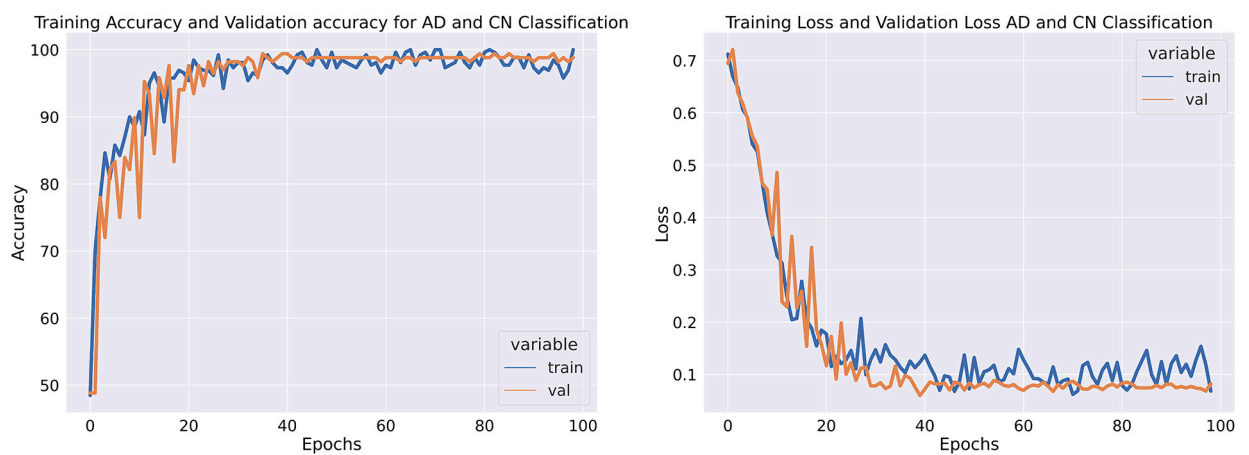
**Table 5**  
Comparison of the performance of various ViT models with the proposed model.

Model	Binary Class	Accuracy %
Simple ViT	AD / CN	89.50
	AD / MCI	78.40
	CN / MCI	85.60
Deep ViT	AD / CN	86.50
	AD / MCI	70.50
	CN / MCI	80.60
Mobile ViT	AD / CN	99.16
	AD / MCI	96.25
	CN / MCI	97.5

AD - Alzheimer Disease; CN - Cognitive Normal; MCI - Mild Cognitive Impairment; ViT - Vision Transformer.



**Fig. 4.** Training Accuracy, Validation Accuracy, Training Loss, and Validation Loss.



**Fig. 5.** Training Accuracy, Validation Accuracy, Training Loss, and Validation Loss with Regularization and Data Augmentation.

specificity of 97.00%, similar to the performance of other models. 3D convolutional neural networks were used to train multimodal magnetic resonance and PET data and achieved an accuracy of 93.21% in AD vs CN classification.

**Table 6**  
Description of Model Performance.

Test Data	Binary Class	Accuracy (%)	Sensitivity (%)	Specificity (%)
Test data ADNI (MRI)	AD vs CN	94.73	90.70	100.00
	AD vs MCI	92.98	90.70	100.00
	CN vs MCI	89.36	90.91	85.17
Test data OASIS (MRI)	AD vs CN	80.00	90.00	70.00
	AD vs MCI	87.50	90.00	85.71
	CN vs MCI	79.16	70.00	85.71
Test data ADNI (PET)	AD vs CN	95.00	90.00	100.00
	AD vs MCI	91.66	90.00	92.88
	CN vs MCI	95.80	100.00	92.86

AD - Alzheimer Disease; CN - Cognitive Normal; MCI - Mild Cognitive Impairment; ADNI - AD Neuroimaging Initiative; OASIS - Open Access Series of Imaging Studies.

**Table 7**  
Analysis of the Proposed Model Considering Various Deep Learning Models.

Model	Binary Class	Accuracy (%)	Specificity (%)	Sensitivity (%)
Hypergraph-based Regularization [35]	AD vs CN	92.51	90.44	94.08
	AD vs MCI	-	-	-
	CN vs MCI	-	-	-
3D-CNN with Sparse Autoencoder [36]	AD vs CN	93.21	95.42	91.43
	AD vs MCI	85.63	81.21	95.54
	CN vs MCI	86.52	94.34	81.64
Inception-ResNet with DWT [26]	AD vs CN	95.90	98.40	92.20
	AD vs MCI	94.10	97.30	92.10
	CN vs MCI	95.40	96.50	94.10
ResNet50 with DWT [46]	AD vs CN	97.00	97.00	97.00
	AD vs MCI	97.50	99.00	96.00
	CN vs MCI	94.00	97.00	97.00
Mobile ViT with GLP	AD vs CN	94.93	90.70	100
	AD vs MCI	92.98	90.70	100
	CN vs MCI	89.36	90.91	85.71

DWT-Discrete Wave Transform; ViT - Vision Transformer; GLP - Gaussian Laplacian.

**Table 8**  
Performance of MViT model based on ablation study.

Approach	Binary Class	Accuracy %
Fused image without sharpening and SPCNN	AD / CN	89.40
	AD / MCI	78.30
	CN / MCI	85.50
Fused without sharpening but with SPCNN	AD / CN	86.20
	AD / MCI	70.10
	CN / MCI	80.40
Fused image without SPCNN but with sharpening.	AD / CN	99.14
	AD / MCI	96.15
	CN / MCI	97.40

AD - Alzheimer Disease; CN - Cognitive Normal; MCI - Mild Cognitive Impairment; ViT - Vision Transformer.

### 5.1. Ablation experiments

To provide an accurate comparison, all experiments utilized the same settings for a fair comparison for the ablation study is shown in Table 8. Table 9, Table 10, and Table 11 depict ablation result on test data. Table 9 serves as the reference table that Table 6, Table 10, and Table 11 are compared with to ascertain the impact of sharpening and SPCNN on test data.

### 5.2. Visualization of regions used for classification

We visualized the regions related to AD dementia, CN and MCI using the MViT model. The brain metabolic region and the structural region related to AD, CN and MCI were localized on individual fused images. 4 sample images are visualized from AD, CN, and MCI class as shown in Fig. 6, Fig. 7, and Fig. 8, respectively.

The areas highlighted in Fig. 6 by the warmer colors on the Grad-CAM heat maps are consistent with regions known to be associated with Alzheimer's disease, particularly those involved with memory and spatial orientation, such as the hippocampus [78]



**Table 9**  
Performance of MViT model based on Fused image without sharpening and SPCNN on test data.

Test Data	Binary Class	Accuracy (%)	Sensitivity (%)	Specificity (%)
Test data ADNI (MRI)	AD vs CN	85.00	90.00	80.00
	AD vs MCI	82.50	90.00	75.00
	CN vs MCI	85.00	90.00	80.00
Test data OASIS (MRI)	AD vs CN	75.00	80.00	60.00
	AD vs MCI	82.50	85.00	75.00
	CN vs MCI	75.00	65.00	85.71
Test data ADNI (PET)	AD vs CN	85.00	80.00	90.00
	AD vs MCI	81.50	80.00	80.00
	CN vs MCI	85.80	90.00	80.16

AD - Alzheimer Disease; CN - Cognitive Normal; MCI - Mild Cognitive Impairment; ADNI - AD Neuroimaging Initiative; OASIS - Open Access Series of Imaging Studies.

**Table 10**  
Performance of MViT model based on Fused image without sharpening but with SPCNN on test data.

Test Data	Binary Class	Accuracy (%)	Sensitivity (%)	Specificity (%)
Test data ADNI (MRI)	AD vs CN	85.18	90.36	80.36
	AD vs MCI	82.68	90.36	75.6
	CN vs MCI	85.18	90.36	80.36
Test data OASIS (MRI)	AD vs CN	75.80	80.40	60.40
	AD vs MCI	82.90	85.40	75.40
	CN vs MCI	75.80	65.40	86.11
Test data ADNI (PET)	AD vs CN	85.20	80.40	90.40.00
	AD vs MCI	81.90	80.40	80.40
	CN vs MCI	85.90	90.40	80.56

AD - Alzheimer Disease; CN - Cognitive Normal; MCI - Mild Cognitive Impairment; ADNI - AD Neuroimaging Initiative; OASIS - Open Access Series of Imaging Studies.

**Table 11**  
Performance of MViT model based on Fused image without SPCNN but with sharpening on test data.

Test Data	Binary Class	Accuracy (%)	Sensitivity (%)	Specificity (%)
Test data ADNI (MRI)	AD vs CN	93.70	87.70	97.00
	AD vs MCI	91.98	87.70	97.00
	CN vs MCI	88.34	87.91	82.17
Test data OASIS (MRI)	AD vs CN	79.00	89.00	69.00
	AD vs MCI	86.50	89.00	84.71
	CN vs MCI	78.12	69.00	84.71
Test data ADNI (PET)	AD vs CN	94.00	89.00	99.00
	AD vs MCI	90.60	89.00	91.80
	CN vs MCI	94.81	99.00	91.80

AD - Alzheimer Disease; CN - Cognitive Normal; MCI - Mild Cognitive Impairment; ADNI - AD Neuroimaging Initiative; OASIS - Open Access Series of Imaging Studies.

and parietal regions [76]. The occipital focus in the first image is less typical for Alzheimer's disease, which may suggest that the model is identifying atypical patterns or that the image captures a different aspect of the disease. The analysis of the highlighted regions can be summarized as follows:

- In the first image on the left, there is a significant hotspot in the posterior part of the brain, which usually corresponds to the occipital lobe, which is primarily associated with visual processing, or potentially the posterior parietal region.
- The second image shows a pronounced focus of activity in the central to posterior region of the brain. This can potentially indicate attention towards the parietal lobe or the posterior cingulate cortex, which are regions involved in various cognitive functions and can be affected in Alzheimer's disease [76,77].
- The third image also shows activity in a central region, but with an extension towards the anterior part. This pattern often implies a focus on the limbic structures such as the hippocampus and surrounding areas of the medial temporal lobe, which are critical in memory and are often affected in Alzheimer's disease [78].
- In the fourth image, there is a concentration of warmer colors towards the posterior and somewhat towards the central region, which potentially indicate the posterior cingulate [77] or the precuneus [79], both of which are involved in memory and spatial orientation and are often areas of interest in Alzheimer's research.

The highlighted regions in Fig. 7 can be connected to areas known to be involved in cognitive processes and could be areas of interest in conditions such as cognitive impairment or Alzheimer's disease:

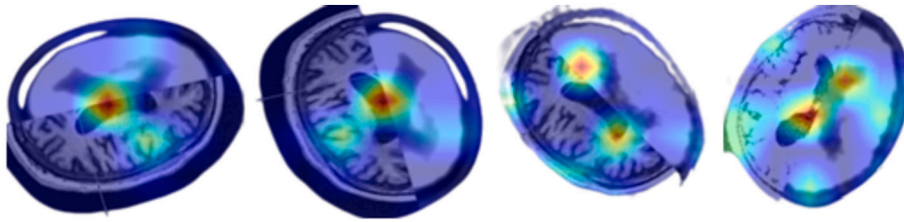


Fig. 6. Grad-CAM Visualization for AD Class.

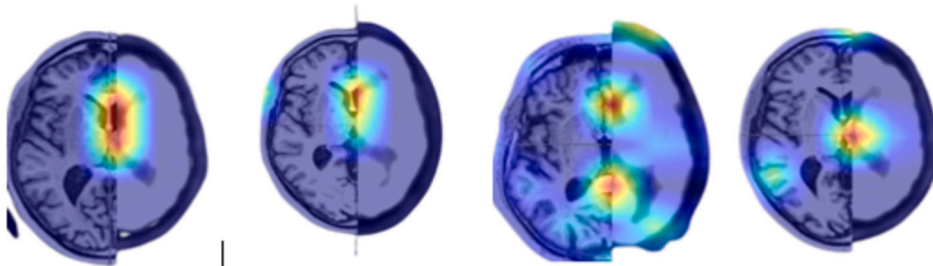


Fig. 7. Grad-CAM Visualization for CN Class.

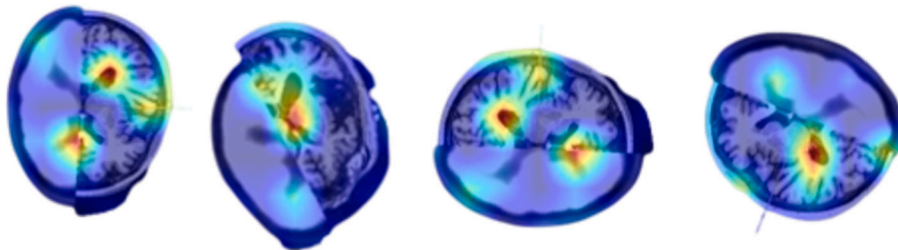


Fig. 8. Grad-CAM Visualization for MCI Class.

- The first image from the left shows a concentration of warmer colors in the central to slightly posterior part of the brain, which potentially indicate the parietal lobe or posterior cingulate cortex [77].
- The second image has a hotspot in the central region but more towards the top of the brain, which suggest the model is focusing on the superior part of the brain possibly related to the parietal lobe or the precuneus.
- The third image has warmer colors in the central region, similar to the first image, indicating the same brain regions like the parietal lobe or posterior cingulate cortex.
- The fourth image displays a concentration of warmer colors toward the anterior and middle portions of the brain, indicating regions like the frontal lobe, which is associated with executive functions, and parts of the limbic system, potentially including the hippocampus, which is central to memory processing.

The regions shown in Fig. 8 correspond to areas typically implicated in Alzheimer's disease and MCI research:

- The middle sections of the brain correspond to areas like the hippocampus and surrounding medial temporal lobe structures, are important for memory and are often the focus in studies related to cognitive impairments and Alzheimer's disease [78].
- The areas toward the back of the brain, potentially indicative of the parietal lobe or posterior cingulate cortex, also known to be involved in cognitive processes and are affected in Alzheimer's disease and MCI.
- Activity shown around the frontal lobe plays a role in cognitive functions and executive decision-making.

## 6. Discussion

From a medical point of view, the use of AI algorithms for the evaluation of fused MRI and PET images presents several distinct advantages over manual evaluations conducted by specialists, particularly in certain contexts. To begin with, the manual evaluation of MRI and PET images is a labor-intensive process that requires specialized training and is susceptible to high measurement variability arising from diverse protocols [80]. In contrast, AI has a remarkable ability to quickly process large volumes of medical imaging data, outpacing human experts in efficiency. This is particularly noteworthy in conditions such as Alzheimer's, where timely detection is

of paramount importance. AI's proficiency in scrutinizing intricate patterns within fused MRI and PET scans contributes significantly to the early diagnosis of the disease. AI's forte in pattern recognition empowers it to discern subtle changes or patterns in medical images that may pose challenges for human experts. In addition, algorithms are adept at seamlessly integrating and analyzing diverse datasets, ensuring a high level of consistency in evaluations. This attribute facilitates a more comprehensive assessment of the patient's condition [81]. This consistency is of particular significance in the longitudinal monitoring of disease progression over time. Current diagnostic approaches for AD begin to focus primarily on modern imaging modalities, with a distinct focus on detecting either amyloid deposition or neurodegeneration [80].

The study's findings affirm the capacity of AI to make meaningful contributions to early diagnosis, demonstrating its accuracy in distinguishing between various stages of the disease. This reinforces the notion that AI, with its efficiency, pattern recognition capabilities, and consistent evaluations, holds significant promise in enhancing medical diagnostics and patient care. The results (see Table 3) indicate that a highly effective multiscale fusion technique can be achieved when images are fused at the sixth level of the LP. This finding is significant, as it suggests that image fusion can be optimized by selecting the appropriate level of the LP. The fused images produced using this technique were found to be sufficiently detailed replicas of the original images, as confirmed by the SSIM readings at level 6 of the LP. Specifically, 0.64 for AD on PET images and 0.60 for AD on MRI images. The average PSNR was approximately 34.80 dB, and the average MSE was approximately 0.213.

The quality of the fusion method suggested that our method maintained the spatial information of the magnetic resonance image and the spectral characteristics of the PET images [82], and this provides further evidence that our proposed method achieved a fused image of good quality. After training different variants of ViT (Simple and Deep) and Mobile ViT on our fused data, we discovered that the Pareto-optimized approach of MViT achieved higher accuracy than the other two ViT variants. As shown in Table 5, our proposed method, which uses Pareto optimization on a mobile ViT and MRI / PET fusion data, performed better than traditional image fusion techniques and deep learning algorithms used with the two ViT variants were selected. The ablation study result depicted in Table 8 shows that there are no significant changes in training accuracy across different variations of the fused image (with or without sharpening and SPCNN) when compared to Table 5, and this suggests that the model might be effectively learning from the training data regardless of the specific preprocessing techniques applied to the input images. The consistency in training accuracy across different variations of the fused image shows that the model is capable of learning from the provided training data regardless of whether sharpening or SPCNN is applied. This could indicate that other factors, such as the inherent quality and informativeness of the training data, are more influential in determining training accuracy. Table 9, Table 10, and Table 11 depicts the ablation study results when the model is tested with test data, the model trained with Fused image without sharpening and SPCNN, Fused image without sharpening but with SPCNN, and Fused image without SPCNN but with sharpening respectively. The model trained with fused image with sharpening and SPCNN generalized well as shown in Table 6 with 9.73% increment in accuracy for AD vs CN on ADNI (MRI). Although the impact of SPCNN is not so significant as seen in Table 10, in which 0.18% increment was achieved in AD vs CN on ADNI (MRI) test data. However, the fact that the model trained with the fused image with sharpening and SPCNN generalized well on the test data indicates that these preprocessing techniques help to improve the model's ability to perform well on unseen data. Sharpening enhances the structural details in the images, while SPCNN fusion integrates both structural and functional information effectively, leading to a more informative representation for the model to learn from. Overfitting was able to be controlled by augmenting the data and with the use of weight decay regularization during optimization, which encourages the weights of the model to be smaller. GLP and SPCNN provide several benefits compared to the DWT. To begin with, both methods excel at maintaining image details and reducing noise, leading to superior image representations. Additionally, SPCNN exhibits greater resilience to image distortions and fluctuations compared to DWT, making it a more efficient option for practical image processing tasks. The combination of SPCNN and GLP has been shown to deliver better results with respect to low feature dimensionality, as DWT breaks images into sub-bands of predetermined bandwidths, while GLP allows greater flexibility in selecting the bandwidth for each subband [83]. The GLP-SPCNN combination outperforms DWT and other approaches, offering advantages like maintaining image details, reducing noise, and being more resistant to distortions. GLP-enabled mobile ViT achieved 100% specificity in identifying CN and MCI without mis-classification. It surpasses other ViT variants and image fusion techniques in AD classification accuracy, addressing overfitting through data augmentation, regularization (dropout), epochs and learning rate scheduler. Fig. 5 shows that the utilization of 0.3 dropout, learning rate scheduler (mode = min, patience = 2), and epochs of 100) controlled the overfitting of the model. The proposed model enables better spatial connections between AD class features, leading to precise AD stage classification. It achieved a low false positive rate, ensuring a reliable identification of individuals without AD, which makes it promising for enhancing AD imaging across multiple modalities. Brain metabolic region and structural region related to AD, CN, and MCI were localized on individual fused images. The related regions of individuals were partly similar to each other. As shown in Fig. 6, Fig. 7, and Fig. 8. This similarity implies indicates that there are similar regions of interest across the different classes.

The combination of MRI and PET data, in particular, can provide the following advantages:

- Magnetic resonance imaging records structural brain integrity, while PET indicates functional alterations related to glucose metabolism or neuronal activity.
  - The fusion of both of these modalities can build a better association between anatomical brain abnormalities and functional impairments by merging these complementary data, potentially offering a more thorough understanding of the neurodegenerative processes of AD.
  - The fusion-based approach may provide extra discriminatory information for more accurate differential discrimination and reducing the rate of misdiagnoses.

- The integration of combined MRI and PET data holds the potential to uncover neuroanatomical biomarkers associated with Alzheimer's disease, offering a promising avenue for advancing our understanding of the diagnostic capabilities of computer vision-based approaches.

## 7. Limitations of this study

The methods applied in this paper have inherent limitations. The initial manifestations of AD-specific topographic patterns emerge in structures of the medial temporal lobe, including the entorhinal and perirhinal cortex, as well as the hippocampus. Although decreased hippocampus volume, assessed in magnetic resonance images, is commonly used to evaluate AD pathology, it is important to note that hippocampal atrophy is not exclusively indicative of AD and is rarely adequate for a definitive diagnosis. Instead, it is considered a biomarker for neurodegeneration [84]. Several cross-sectional studies have indicated that entorhinal cortex measurements do not provide significantly greater benefits over hippocampal volumetry to distinguish AD patients from the control group, despite the potentially higher accuracy in volumetric evaluations. As a result, MRI findings, while valuable, should be interpreted in conjunction with other relevant findings to avoid potentially misleading conclusions [85]. In PET tomography, hypometabolism in the temporoparietal regions is recognized as a diagnostic criterion for AD. However, it is important to recognize that the glucose uptake measured in this context reflects astrocyte function rather than neuronal function. PET can also estimate the accumulation of Amyloid- $\beta$  ( $A\beta$ ), the first hallmark of AD. However,  $A\beta$  accumulation tends to plateau as the disease progresses. Furthermore, different subtypes of AD exhibit diverse changes in metabolic patterns, which decreases the diagnostic accuracy of FDG-PET. Consequently, FDG-PET is currently not recommended for the preclinical diagnosis of AD due to the challenge of verifying whether hypometabolism is directly related to AD pathology [86]. A synergistic approach involving both structural and functional imaging, assessed through AI, holds promise as a potential solution to the challenging task of differentiating Alzheimer's disease from other conditions. This integrated methodology leverages the strengths of structural and functional data, offering a more comprehensive understanding of the brain's status. AI becomes a valuable tool designed to complement healthcare professionals, offering them additional insight and improving overall efficiency in the diagnostic process. However, it is imperative to recognize that the effectiveness of AI is dependent on the quality of the algorithms used and the specificity of their application. Rigorous validation in real-life conditions and continuous refinement of these algorithms are still essential to ensure their reliability and relevance in clinical settings. While the use of external test data, such as the small sample extracted from the OASIS database, offers a valuable opportunity for validation, it also introduces limitations. The small size of the external test dataset, driven by the limited number of subjects who meet specific inclusion criteria, particularly for AD dementia, underscores the challenge of generalizing the findings to broader populations. This limitation highlights the need for caution when extrapolating results and emphasizes the importance of further validation efforts, preferably with larger and more diverse datasets, to enhance the robustness and applicability of AI-based approaches in clinical practice.

## 8. Conclusion and future works

The study presented an algorithm for classifying different stages of AD using fused images obtained from sMRI and FDG-PET imaging. The results showed promise, showing a good level of accuracy, sensitivity, and specificity in distinguishing Alzheimer's disease dementia from normal cognitive impairment, Alzheimer's disease dementia from mild cognitive impairment, and normal cognitive from mild cognitive impairment. This innovative approach, which uses a Pareto-optimized MViT model with GLP and SPCNN for image fusion, could overcome the limitations of error-prone feature selection techniques in multimodal imaging. By applying vision-transformer models for image categorization and fusion, more accurate and reliable AD analysis might be achievable, providing complementary data from multiple imaging modalities. The results of the OASIS and ADNI datasets suggest that this method could aid in early diagnosis, potentially improving patient outcomes and quality of life. With further validation and refinement, this technique could become a valuable tool for early identification of AD.

Promising avenue for the future research lies in enhancing the discriminatory capability of our diagnostic method to distinguish between amnesic and atypical presentations of AD. These distinct presentations may entail different patterns of brain involvement, and further optimization of our technique could yield insights into the specific brain areas affected in these varied AD subtypes. Such refinements hold the potential to contribute to more personalized approaches to diagnosis and treatment. Another intriguing direction for future exploration involves investigating the diagnostic value of our novel technique in distinguishing AD from other neurodegenerative disorders, such as frontotemporal lobar degenerations, Lewy body disorders, tauopathies, TDP-43 proteinopathies, and synucleinopathies. Given the potential overlap in affected brain regions among these conditions, our method could play a crucial role in providing clarity during the challenging process of making differential diagnoses, having the potential to significantly enhance our ability to differentiate AD from a spectrum of neurodegenerative disorders, thereby facilitating more accurate and timely clinical decisions. Finally, the integration of additional imaging modalities or clinical data could help to further enhance the precision and reliability of the approach, as would the further analysis of different high-pass filters, which could provide a more comprehensive understanding of the various factors that influence multimodal fusion in medical imaging analysis.

## CRedit authorship contribution statement

**Modupe Odusami:** Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation. **Robertas Damaševičius:** Writing – review & editing, Validation, Formal analysis. **Egle Milieškaitė-Belousovienė:** Validation, Formal analysis. **Rytis Maskeliūnas:** Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database is available from <http://adni.loni.usc.edu/> (accessed on March 18, 2024).

## Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and the DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California.

## References

- [1] G. Gopalakrishna, P. Joshi, P.-H. Tsai, N. Patel, Advances in Alzheimer's dementia: an update from banner Alzheimer's institute, *Am. J. Geriatr. Psychiatry* 31 (3) (2023) S18–S19, <https://doi.org/10.1016/j.jagp.2022.12.300>.
- [2] L. Liu, Y. Zhang, L. Tang, H. Zhong, D. Danzeng, C. Liang, S. Liu, The neuroprotective effect of byu d mar 25 in lps-induced Alzheimer's disease mice model, *Evidence-Based Complementary and Alternative Medicine* 2021 (2021) 1–12, <https://doi.org/10.1155/2021/8879014>.
- [3] O. Grigas, R. Maskeliunas, R. Damaševičius, Early detection of dementia using artificial intelligence and multimodal features with a focus on neuroimaging: a systematic literature review, *Health Technol.* 14 (2) (2024) 201–237, <https://doi.org/10.1007/s12553-024-00823-0>.
- [4] G. Deuschl, E. Beghi, F. Fazekas, T. Varga, K.A. Christoforidi, E. Sipido, C.L. Bassetti, T. Vos, V.L. Feigin, The burden of neurological diseases in Europe: an analysis for the global burden of disease study 2017, *Lancet Public Health* 5 (10) (2020) e551–e567, [https://doi.org/10.1016/s2468-2667\(20\)30190-0](https://doi.org/10.1016/s2468-2667(20)30190-0).
- [5] Y. Zhang, X. Li, Y. Ji, H. Ding, X. Suo, X. He, Y. Xie, M. Liang, S. Zhang, C. Yu, et al., Mraβ: a multimodal mri-derived amyloid-β biomarker for Alzheimer's disease, *Hum. Brain Mapp.* (2023).
- [6] J. Graff-Radford, K.X.X. Yong, L.G. Apostolova, F.H. Bouwman, M. Carrillo, B.C. Dickerson, G.D. Rabinovici, J.M. Schott, D.T. Jones, M.E. Murray, New insights into atypical Alzheimer's disease in the era of biomarkers, *Lancet Neurol.* 20 (3) (2021) 222–234, [https://doi.org/10.1016/s1474-4422\(20\)30440-3](https://doi.org/10.1016/s1474-4422(20)30440-3).
- [7] P.A. Rowley, A.A. Samsonov, T.J. Betthausen, A. Pirasteh, S.C. Johnson, L.B. Eisenmenger, Amyloid and tau PET imaging of Alzheimer disease and other neurodegenerative conditions, *Semin. Ultrasound CT MR* 41 (6) (2020) 572–583, <https://doi.org/10.1053/j.sult.2020.08.011>.
- [8] F. Gao, Integrated positron emission tomography/magnetic resonance imaging in clinical diagnosis of Alzheimer's disease, *Eur. J. Radiol.* 145 (2021) 110017, <https://doi.org/10.1016/j.ejrad.2021.110017>.
- [9] Y.-D. Zhang, G. Zhao, J. Sun, X. Wu, Z.-H. Wang, H.-M. Liu, V.V. Govindaraj, T. Zhan, J. Li, Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm, *Multimed. Tools Appl.* 77 (17) (2017) 22629–22648, <https://doi.org/10.1007/s11042-017-5023-0>.
- [10] C. Bouter, P. Henniges, T.N. Franke, C. Irwin, C.O. Sahlmann, M.E. Sichter, N. Beindorff, T.A. Bayer, Y. Bouter, 18F-FDG-PET detects drastic changes in brain metabolism in the tg4–42 model of Alzheimer's disease, *Front. Aging Neurosci.* 10 (Jan. 2019), <https://doi.org/10.3389/fnagi.2018.00425>.
- [11] H.W. Kim, H.E. Lee, K. Oh, S. Lee, M. Yun, S.K. Yoo, Multi-slice representational learning of convolutional neural network for Alzheimer's disease classification using positron emission tomography, *Biomed. Eng. Online* 19 (1) (Sep. 2020), <https://doi.org/10.1186/s12938-020-00813-z>.
- [12] G. Chételat, J. Arbizu, H. Barthel, V. Garibotto, I. Law, S. Morbelli, E. van de Giessen, F. Agosta, F. Barkhof, D.J. Brooks, M.C. Carrillo, B. Dubois, A.M. Fjell, G.B. Frisoni, O. Hansson, K. Herholz, B.F. Hutton, C.R. Jack, A.A. Lammertsma, S.M. Landau, S. Minoshima, F. Nobili, A. Nordberg, R. Ossenkoppele, W.J.G. Oyen, D. Perani, G.D. Rabinovici, P. Scheltens, V.L. Villemagne, H. Zetterberg, A. Drzezga, Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias, *Lancet Neurol.* 19 (11) (2020) 951–962, [https://doi.org/10.1016/s1474-4422\(20\)30314-8](https://doi.org/10.1016/s1474-4422(20)30314-8).
- [13] J. Wang, C. Jin, J. Zhou, R. Zhou, M. Tian, H.J. Lee, H. Zhang, PET molecular imaging for pathophysiological visualization in Alzheimer's disease, *Eur. J. Nucl. Med. Mol. Imaging* 50 (3) (2022) 765–783, <https://doi.org/10.1007/s00259-022-05999-z>.
- [14] S. Nugent, et al., Selection of the optimal intensity normalization region for fdg-pet studies of normal aging and Alzheimer's disease, *Sci. Rep.* 10 (1) (2020), <https://doi.org/10.1038/s41598-020-65957-3>.
- [15] V. Senthilvel, V. Govindaraj, Y.-D. Zhang, P.R. Murugan, A.P. Thyagarajan, A smartly designed automated map based clustering algorithm for the enhanced diagnosis of pathologies in brain mr images, *Expert Syst.* 38 (2) (2021) e12625.
- [16] S. Rathore, M. Habes, M.A. Ifikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages, *NeuroImage* 155 (2017) 530–548, <https://doi.org/10.1016/j.neuroimage.2017.03.057>.
- [17] N. Sharma, Exploring biomarkers for Alzheimer's disease, *J. Clin. Diagn. Res.* (2016), <https://doi.org/10.7860/jcdr/2016/18828.8166>.
- [18] Y.-D. Zhang, et al., Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation, *Inf. Fusion* 64 (2020) 149–187, <https://doi.org/10.1016/j.inffus.2020.07.006>.
- [19] H. Wei, M. Kong, C. Zhang, L. Guan, M. Ba, The structural mri markers and cognitive decline in prodromal Alzheimer's disease: a 2-year longitudinal study, *Quant. Imaging Med. Surg.* 8 (10) (2018) 1004–1019, <https://doi.org/10.21037/qims.2018.10.08>.
- [20] J. Alber, et al., Developing retinal biomarkers for the earliest stages of Alzheimer's disease: what we know, what we don't, and how to move forward, *Alzheimer's Dement.* 16 (1) (2020) 229–243, <https://doi.org/10.1002/alz.12006>.



- [21] M. Odusami, R. Maskeliūnas, R. Damaševičius, S. Misra, Machine learning with multimodal neuroimaging data to classify stages of Alzheimer's disease: a systematic review and meta-analysis, *Cogn. Neurodyn.* (2023), <https://doi.org/10.1007/s11571-023-09993-5>.
- [22] H.-M. Zeng, H.-B. Han, Q.-F. Zhang, H. Bai, Application of modern neuroimaging technology in the diagnosis and study of Alzheimer's disease, *Neural Regen. Res.* 16 (1) (2020) 73–79, <https://doi.org/10.4103/1673-5374.286957>.
- [23] J. Song, J. Zheng, P. Li, X. Lu, G. Zhu, P. Shen, An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis, *Frontiers in Digital Health* 3 (Feb. 2021), <https://doi.org/10.3389/fdgth.2021.637386>.
- [24] M. Odusami, R. Maskeliūnas, R. Damaševičius, Optimized convolutional fusion for multimodal neuroimaging in Alzheimer's disease diagnosis: enhancing data integration and feature extraction, *Journal of Personalized Medicine* 13 (10) (2023), <https://doi.org/10.3390/jpm13101496>.
- [25] J. Yang, V.V. Govindaraj, M. Yang, S.-H. Wang, Hearing loss detection by discrete wavelet transform and multi-layer perceptron trained by nature-inspired algorithms, *Multimed. Tools Appl.* 79 (21–22) (2020) 15717–15745, <https://doi.org/10.1007/s11042-019-08344-z>.
- [26] S.V.P. Rallabandi, K. Seetharaman, Deep learning-based classification of healthy aging controls, mild cognitive impairment and Alzheimer's disease using fusion of mri-pet imaging, *Biomed. Signal Process. Control* 80 (2023) 104312, <https://doi.org/10.1016/j.bspc.2022.104312>.
- [27] S. Liu, et al., Two-scale multimodal medical image fusion based on structure preservation, *Front. Comput. Neurosci.* 15 (2022) 803724, <https://doi.org/10.3389/fncom.2021.803724>.
- [28] H.-I. Suk, S.-W. Lee, D. Shen, T.A.D.N. Initiative, Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis, *Brain Struct. Funct.* 221 (5) (2016) 2569–2587, <https://doi.org/10.1007/s00429-015-1059-y>.
- [29] J. Giorgio, S. Landau, W. Jagust, P. Tino, Z. Kourtzi, Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease, *NeuroImage Clin.* 26 (2020) 102199, <https://doi.org/10.1016/j.nicl.2020.102199>.
- [30] Y. Shi, et al., Asmfs: adaptive-similarity-based multi-modality feature selection for classification of Alzheimer's disease, *Pattern Recognit.* 126 (2022) 108566, <https://doi.org/10.1016/j.patcog.2022.108566>.
- [31] C. Zu, Y. Wang, L. Zhou, L. Wang, D. Zhang, Multi-modality feature selection with adaptive similarity learning for classification of Alzheimer's disease, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 1542–1545.
- [32] X.-H. Hao, et al., Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease, *Med. Image Anal.* 60 (2020) 101625, <https://doi.org/10.1016/j.media.2019.101625>.
- [33] V. Yousofzadeh, B. McGuinness, L.P. Maguire, K. Wong-Lin, Multi-kernel learning with darts improves combined mri-pet classification of Alzheimer's disease in aibl data: group and individual analyses, *Front. Human Neurosci.* 11 (2017) 2017, <https://doi.org/10.3389/fnhum.2017.00380>.
- [34] J. Song, et al., An effective multimodal image fusion method using mri and pet for Alzheimer's disease diagnosis, *Frontiers in Digital Health* 3 (2021) 2021, <https://doi.org/10.3389/fdgth.2021.637386>.
- [35] W. Shao, Y. Peng, C. Zu, M. Wang, D. Zhang, A.D.N. Initiative, et al., Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease, *Comput. Med. Imaging Graph.* 80 (2020) 101663.
- [36] Z. Kong, M. Zhang, W. Zhu, Y. Yi, T. Wang, B. Zhang, Multi-modal data Alzheimer's disease detection based on 3d convolution, *Biomed. Signal Process. Control* 75 (2022) 103565.
- [37] S. Sharma, P.K. Mandal, A comprehensive report on machine learning-based early detection of Alzheimer's disease using multi-modal neuroimaging data, *ACM Comput. Surv.* 55 (2) (2023) 1–44, <https://doi.org/10.1145/3492865>.
- [38] W. Dang, L. Xiang, S. Liu, B. Yang, M. Liu, Z. Yin, L. Yin, W. Zheng, A feature matching method based on the convolutional neural network, *J. Imaging Sci. Technol.* 67 (3) (2023) 030402, <https://doi.org/10.2352/j.imagingsci.technol.2023.67.3.030402>.
- [39] K.P. Muhammed Niyas, P. Thiagarajan, Feature selection using efficient fusion of Fisher score and greedy searching for Alzheimer's classification, *J. King Saud Univ. Comput. Inf. Sci.* 34 (8) (2022) 4993–5006, <https://doi.org/10.1016/j.jksuci.2020.12.009>.
- [40] T. Zhou, K.-H. Thung, X. Zhu, D. Shen, Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis, *Hum. Brain Mapp.* 40 (3) (2019) 1001–1016, <https://doi.org/10.1002/hbm.24428>.
- [41] Z. Ning, et al., Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis, *IEEE Trans. Med. Imaging* 40 (6) (2021) 1632–1645, <https://doi.org/10.1109/TMI.2021.3063150>.
- [42] J. Ramya, B.U. Maheswari, M.P. Rajakumar, R. Sonia, Alzheimer's disease segmentation and classification on mri brain images using enhanced expectation maximization adaptive histogram (eem-ah) and machine learning, *Inf. Technol. Control* 51 (4) (2022) 786–800.
- [43] Y. Gupta, J.-I. Kim, B.C. Kim, G.-R. Kwon, Classification and graphical analysis of Alzheimer's disease and its prodromal stage using multimodal features from structural, diffusion, and functional neuroimaging data and the apoe genotype, *Front. Aging Neurosci.* 12 (2020) 2020, <https://doi.org/10.3389/fnagi.2020.00238>.
- [44] X. Deng, E. Liu, S. Li, Y. Duan, M. Xu, Interpretable multi-modal image registration network based on disentangled convolutional sparse coding, *IEEE Trans. Image Process.* 32 (2023) 1078–1091, <https://doi.org/10.1109/tip.2023.3240024>.
- [45] N. Alseelawi, H. Hazim, H. Alrikabi, A novel method of multimodal medical image fusion based on hybrid approach of nscf and dtcwt 18 (2018) 28011, <https://doi.org/10.3991/ijoe.v18i03.28011>.
- [46] S. Dwivedi, et al., Multi-modal fusion based deep learning network for effective diagnosis of Alzheimer's disease, *IEEE Multimed.* (2022) 1, <https://doi.org/10.1109/MMUL.2022.3156471>.
- [47] Y. Huang, J. Xu, Y. Zhou, T. Tong, X. Zhuang, The Alzheimer's disease neuroimaging initiative (ADNI), diagnosis of Alzheimer's disease via multi-modality 3d convolutional neural network, *Front. Neurosci.* 13 (2019) 2019, <https://doi.org/10.3389/fnins.2019.00509>.
- [48] M. Odusami, R. Maskeliūnas, R. Damaševičius, Pixel-level fusion approach with vision transformer for early detection of Alzheimer's disease, *Electronics* 12 (5) (2023) 1218, <https://doi.org/10.3390/electronics12051218>.
- [49] T. Tirupal, et al., Medical image fusion using undecimated discrete wavelet transform for analysis and detection of Alzheimer's disease 137 (2019) 53905–53910, <https://doi.org/10.1049/ijoe.v18i03.28011>.
- [50] Y. Si, Lppcnn: a Laplacian pyramid-based pulse coupled neural network method for medical image fusion, *J. Appl. Sci. Eng.* 24 (3) (2021) 299–305, [https://doi.org/10.6180/jase.202106\\_24\(3\).0004](https://doi.org/10.6180/jase.202106_24(3).0004).
- [51] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: transformers for image recognition at scale, arXiv preprint, arXiv:2010.11929, 2020.
- [52] L. Beyer, X. Zhai, A. Kolesnikov, Better plain vit baselines for imagenet-1k, preprint, arXiv:2205.01580, 2022.
- [53] D. Zhou, B. Kang, X. Jin, L. Yang, X. Lian, Z. Jiang, Q. Hou, J. Feng, Deepvit: towards deeper vision transformer, preprint, arXiv:2103.11886, 2021.
- [54] S. Mehta, M. Rastegari, Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer, preprint, arXiv:2110.02178, 2021.
- [55] J. Choi, G. Kim, N. Park, H. Park, S. Choi, A hybrid pansharpening algorithm of vhr satellite images that employs injection gains based on ndvi to reduce computational costs, *Remote Sens.* 9 (10) (2017) 976, <https://doi.org/10.3390/rs9100976>.
- [56] F. Shi, J. Wang, V. Govindaraj, Sgs: squeeze-net-guided Gaussian-kernel svm for covid-19 diagnosis, *Mob. Netw. Appl.* (2024) 1–14.
- [57] J. Liang, L. Zhang, L. Zhang, Z. Kourtzi, P. Tino, High-resolution photorealistic image translation in real-time: a Laplacian pyramid translation network, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9392–9400.
- [58] Y. Sumiya, T. Otsuka, Y. Maeda, N. Fukushima, Gaussian Fourier pyramid for local Laplacian filter, *IEEE Signal Process. Lett.* 29 (2022) 11–15, <https://doi.org/10.1109/LSP.2021.3121198>.
- [59] H. Li, J. Wang, C. Han, Image mosaic and hybrid fusion algorithm based on pyramid decomposition, in: *2020 International Conference on Virtual Reality and Visualization (ICVRV)*, 2020, pp. 205–208.

- [60] N. Rafi, P. Rivas, A review of pulse-coupled neural network applications in computer vision and image processing, preprint, arXiv:1907.06334, 2019.
- [61] X. Li, H. Yan, W. Xie, L. Kang, Y. Tian, An improved pulse-coupled neural network model for pansharpening, *Sensors (Basel)* 20 (10) (2020) 2764, <https://doi.org/10.3390/s20102764>.
- [62] H. Ouerghi, O. Mourali, E. Zagrouba, Non-sampled shearlet transform based mri and pet brain image fusion using simplified pulse coupled neural network and weight local features in yiq colour space, *IET Image Process.* 12 (10) (2018) 1873–1880, <https://doi.org/10.1049/iet-ipr.2017.1298>.
- [63] Y.-D. Zhang, V. Govindaraj, Z. Zhu, Fecnet: a neural network and a mobile app for covid-19 recognition, *Mob. Netw. Appl.* (2023) 1–14.
- [64] Y. Li, C.-Y. Wu, H. Fan, K. Mangalam, B. Xiong, J. Malik, C. Feichtenhofer, Mvitv2: improved multiscale vision transformers for classification and detection, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 4794–4804.
- [65] Q. Wang, X. Jin, Q. Jiang, L. Wu, Y. Zhang, W. Zhou, Dbct-net: a dual branch hybrid cnn-transformer network for remote sensing image fusion, *Expert Syst. Appl.* 233 (2023) 120829, <https://doi.org/10.1016/j.eswa.2023.120829>.
- [66] M. Zhang, W. Jing, J. Lin, N. Fang, W. Wei, M. Woźniak, R. Damaševičius, Nas-hris: automatic design and architecture search of neural network for semantic segmentation in remote sensing images, *Sensors (Switzerland)* 20 (18) (2020) 1–15.
- [67] X. Liu, J. Zhao, J. Li, B. Cao, Z. Lv, Federated neural architecture search for medical data security, *IEEE Trans. Ind. Inform.* 18 (8) (2022) 5628–5636, <https://doi.org/10.1109/tii.2022.3144016>.
- [68] L. Gyongyosi, S. Imre, A survey on quantum computing technology, *Comput. Sci. Rev.* 31 (2019) 51–71.
- [69] A. Chinchuluun, P.M. Pardalos, A survey of recent developments in multiobjective optimization, *Ann. Oper. Res.* 154 (1) (2007) 29–50.
- [70] J. Jin, Q. Zhang, J. He, H. Yu, Quantum dynamic optimization algorithm for neural architecture search on image classification, *Electronics* 11 (23) (2022) 3969, <https://doi.org/10.3390/electronics11233969>.
- [71] S. Otake, R. Sasaki, Discrete quantum mechanics, *J. Phys. A, Math. Theor.* 44 (35) (2011).
- [72] S. Poppi, M. Cornia, L. Baraldi, R. Cucchiara, Revisiting the evaluation of class activation mapping for explainability: a novel metric and experimental analysis, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021*, pp. 2299–2304.
- [73] H. Choi, Y.K. Kim, E.J. Yoon, J.-Y. Lee, D.S. Lee, A.D.N. Initiative, Cognitive signature of brain fdg pet based on deep learning: domain transfer from Alzheimer's disease to Parkinson's disease, *Eur. J. Nucl. Med. Mol. Imaging* 47 (2020) 403–412.
- [74] A. Chattopadhyay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018, pp. 839–847.
- [75] M. Lucas, M. Lerma, J. Furst, D. Raicu, Rsi-grad-cam: visual explanations from deep networks via Riemann-Stieltjes integrated gradient-based localization, in: *International Symposium on Visual Computing, Springer, 2022*, pp. 262–274.
- [76] H.I. Jacobs, M.P. Van Boxtel, J. Jolles, F.R. Verhey, H.B. Uylings, Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings, *Neurosci. Biobehav. Rev.* 36 (1) (2012) 297–309, <https://doi.org/10.1016/j.neubiorev.2011.06.009>.
- [77] P.-L. Lee, K.-H. Chou, C.-P. Chung, T.-H. Lai, J.H. Zhou, P.-N. Wang, C.-P. Lin, Posterior cingulate cortex network predicts Alzheimer's disease progression, *Front. Aging Neurosci.* 12 (Dec. 2020), <https://doi.org/10.3389/fnagi.2020.608667>.
- [78] Y.L. Rao, B. Ganaraja, B.V. Murlimanju, T. Joy, A. Krishnamurthy, A. Agrawal, Hippocampus and its involvement in Alzheimer's disease: a review, *3 Biotech* 12 (2) (Feb. 2022), <https://doi.org/10.1007/s13205-022-03123-4>.
- [79] F. Ahmad, M. Javed, M. Athar, S. Shahzadi, Determination of affected brain regions at various stages of Alzheimer's disease, *Neurosci. Res.* 192 (2023) 77–82, <https://doi.org/10.1016/j.neures.2023.01.010>.
- [80] W.M. van Oostveen, E.C.M. de Lange, Imaging techniques in Alzheimer's disease: a review of applications in early diagnosis and longitudinal monitoring, *Int. J. Mol. Sci.* 22 (4) (2021) 2110, <https://doi.org/10.3390/ijms22042110>.
- [81] Z. Yao, H. Wang, W. Yan, Z. Wang, W. Zhang, Z. Wang, G. Zhang, Artificial intelligence-based diagnosis of Alzheimer's disease with brain mri images, *Eur. J. Radiol.* 165 (2023) 110934, <https://doi.org/10.1016/j.ejrad.2023.110934>.
- [82] K. Padmavathi, C.S. Asha, V.K. Maya, A novel medical image fusion by combining tv-l1 decomposed textures based on adaptive weighting scheme, *Int. J. Eng. Sci. Technol.* 23 (1) (2020) 225–239, <https://doi.org/10.1016/j.jestch.2019.03.008>.
- [83] M.A. Saleh, A.A. Ali, K. Ahmed, A.M. Sarhan, A comprehensive report on machine learning-based early detection of Alzheimer's disease using multi-modal neuroimaging data, *Electronics* 12 (1) (2022) 97, <https://doi.org/10.3390/electronics12010097>.
- [84] L. Mosconi, Glucose metabolism in normal aging and Alzheimer's disease: methodological and physiological considerations for pet studies, *Clinical and Translational Imaging* 1 (4) (2013) 217–233, <https://doi.org/10.1007/s40336-013-0026-y>.
- [85] A. Porsteinsson, R. Isaacson, S. Knox, M. Sabbagh, I. Rubino, Diagnosis of early Alzheimer's disease: clinical practice in 2021, *The Journal of Prevention of Alzheimer's Disease* (2021) 1–16, <https://doi.org/10.14283/jpad.2021.23>.
- [86] B. Dubois, H. Hampel, et al., Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria, *Alzheimer's Dement.* 12 (3) (2016) 292–323, <https://doi.org/10.1016/j.jalz.2016.02.002>.