Neuroimaging Fusion in Nonsubsampled Shearlet Domain by Maximizing the High-Frequency Subband Energy and Classification of Alzheimer's Disease using Local and Global Contextual CNN Features of Neuroimaging Data

Emimal Jabason

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

April 2024

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Emimal Jabason

Entitled: Neuroimaging Fusion in Nonsubsampled Shearlet Domain by Maximizing the High-Frequency Subband Energy and Classification of Alzheimer's Disease using Local and Global Contextual CNN Features of Neuroimaging Data

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Electrical and Computer Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Adel M. Hanna

_____ External Examiner
Dr. Sridhar Krishnan

_____ Arm's Length Examiner
Dr. Yong Zeng

_____ Examiner
Dr. Hassan Rivaz

_____ Examiner
Dr. Wei-Ping Zhu

_____Thesis Supervisor
Dr. M. Omair Ahmad

_____Thesis Supervisor
Dr. M.N.S. Swamy

Approved by _____
Dr. Jun Cai, Graduate Program Director

April 25, 2024 _____
Dr. Mourad Debbabi, Dean, Gina Cody School of Engineering and Computer Science

Abstract

Neuroimaging Fusion in Nonsubsampled Shearlet Domain by Maximizing the High-Frequency Subband Energy and Classification of Alzheimer's Disease using Local and Global Contextual CNN Features of Neuroimaging Data

Emimal Jabason, Ph.D.

Concordia University, 2024

Neuroimaging fusion is the process of combining brain imaging data from multiple imaging modalities to create a composite image containing complementary information such as structural and functional changes in the brain. Recent advancements in transform domain fusion are promising, but challenges remain in accurately representing empirical distributions and maximizing energy in fused images. The most common neurodegenerative disease is Alzheimer's disease, which demands accurate detection and classification for the care of the patient. Recent advancements in convolutional neural networks (CNNs)-based methods often overlook local features and do not pay attention to the discriminability of extracted features for the classification of Alzheimer's disease. Moreover, existing architectures often end up using numerous parameters to enhance feature richness.

The objective of this thesis has two parts. In the first part, a novel statistically driven approach for fusing multimodal neuroimaging data is developed. In the second part, a lightweight deep CNN capable of extracting both local and global contextual features for the classification of Alzheimer's disease is proposed.

In the first part of the thesis, a novel multimodal fusion algorithm using statistical properties of nonsubsampled shearlet transform coefficients and an energy maximization fusion

rule is developed. The Student's t probability density function is used to model heavy-tailed non-Gaussian statistics of empirical coefficients. This model is then employed to develop a maximum a posteriori estimator to obtain noise-free coefficients. Finally, a novel fusion rule is proposed for obtaining fused coefficients by maximizing the energy in the high-frequency subbands.

In the second part of the thesis, a novel lightweight deep CNN that extracts local and global contextual features for Alzheimer's disease classification is proposed. The network is designed to process local and global features separately using specialized modules that enhance feature extraction relevant to the disease. Finally, the impact of fused images, obtained using the fusion approach of the first part, on the classification accuracy of Alzheimer's disease is investigated.

Extensive experiments are carried out to validate the effectiveness of the various ideas and strategies proposed in this thesis for developing multimodal neuroimaging fusion and Alzheimer's disease classification schemes.

Acknowledgments

First and foremost, I extend my deepest thanks and all praises and glory to my GOD Almighty for His boundless grace, which has been the cornerstone of my strength in completing this Ph.D journey. My heartfelt thanks also go to the Beginning Pentecostal Truth Church, Malavilai, India, for their steadfast prayers.

I would like to express my sincere gratitude and a very special thank you to my thesis supervisors, Dr. M. Omair Ahmad and Dr. M.N.S. Swamy. Their constant support, guidance, and understanding have not only ensured the successful completion of this Ph.D thesis but have also empowered me with skills and knowledge that will benefit throughout my academic and professional career. I feel so grateful to my supervisors for the financial support that I received from the Supervisor(s) research grant(s) awarded by Natural Sciences and Engineering Research Council of Canada and Regroupement Stratégique en Microsystèmes du Québec. I am also thankful to Concordia University for the monetary awards and Fonds de recherche du Québec - Nature et technologies for the FRQ scholarship/fellowship award.

I wish to thank my cherished grandmother, H. Retna Bai, whose passing has left a profound void within me. While her physical presence may be absent, her enduring prayers, love, care, and sacrifices continue to strengthen me. I extend my heartfelt gratitude to my parents and all the family members for their support throughout my life.

Last but not least, I wish to convey my heartfelt gratitude to my dear husband, Mr. Jabason Stanley, and my two beloved children, Silvan Jabason and Shawn Stanley. Words cannot fully capture the depth of their love, patience, and the countless sacrifices. I am blessed to have them as an inseparable part of my life's voyage.

Contents

List of Figures

List of Tables

List of Symbols

H_p	Height of pooled feature tensor
K_{ij}	Key map
P	Probabilities of classes obtained after softmax activation
$Q^{\frac{f_1 f_2}{F}}$	Normalized weighted performance metric.
Q_{ij}	Query map
V_{ij}	Value map
W_p	Width of pooled feature tensor
W_{ij}	Weight matrix
X_{ij}^a	Attention-weighted feature tensor
X_{pi}	Pooled feature tensor
Y_f	One-dimensional feature vector
Y_i	Set of features obtained from pooled features, X_{pi}
Y_p	Average-pooled feature tensor
Y_{14}	Fused feature tensor obtained by concatenating Y_1 and Y_4

Y_1	Weighted local feature tensor
Y_4	Weighted global feature tensor
Z	Output of the second dense layer
d_i	Depth of pooled feature tensor
i	Indexes used for different feature levels
$D(k)$	Number of directions at scale k.
$G_i(u, v)$	NSST coefficients of the observed neuroimaging data.
K	Number of scales.
M	Number of source images.
N	Size of the image, I .
$O(i, j)$	Value of the output feature map at position (i, j).
$P\hat{f}_d^l$	Discrete Fourier transform of the high-pass image H_j^l.
X_1	Tensor resulting from the first convolutional layer, consisting of 16 feature maps of size $\frac{H}{2} \times \frac{W}{2}$
X_2	Tensor resulting from the second convolutional layer, consisting of 16 feature maps of size $\frac{H}{2} \times \frac{W}{2}$
X_3	Tensor resulting from the third convolutional layer, consisting of 32 feature maps of size $\frac{H}{2} \times \frac{W}{2}$
X_4	Tensor resulting from the fourth convolutional layer, consisting of 32 feature maps of size $\frac{H}{2} \times \frac{W}{2}$, representing the highest level of feature extraction

α	Parameter representing the relative importance of luminance in SSIM.
β	Parameter representing the relative importance of contrast in SSIM.
$\frac{H}{2} \times \frac{W}{2}$	Size of feature maps after processing with stride 2
γ	Parameter representing the relative importance of structure in SSIM.
$\sigma^2_{N_i(u,v)}$	Noise variance.
c	Contrast.
$f_1(x, y)$	Input image 1.
$f_2(x, y)$	Input image 2.
j	Index for columns in the output feature map.
l	Luminance.
m	Index for rows within the pooling window.
n	Index for columns within the pooling window.
s	Structure in SSIM equation.
$H \times W$	Dimension of the input image, where H is the height and W is the width of X
HH	Subband capturing high-frequency details in both horizontal and vertical directions

HL	Subband capturing high-frequency horizontal details and low-frequency vertical details
I	Input image for 2D Discrete Wavelet Transform
LH	Subband capturing low-frequency horizontal details and high-frequency vertical details
LL	Lowest-frequency subband, representing coarse approximation
X	Input image of size $H \times W$, representing a single slice of MRI data

List of Abbreviations

18F-FDG	Fluorine-18 Fluorodeoxyglucose
AD	Alzheimer's disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
CB	Contourlet transform-based
CNN	Convolutional neural network
CNNs	Convolutional neural networks
CT	Computed tomography
CTB	Curvelet transform-based
DA-MIDL	Dual attention multi-instance deep learning
EM	Expectation maximization
GGD	Generalised Gaussian distribution
GIFB	Guided image filter-based
GM	Gray matter

H-FCN	Hierarchical fully convolutional network
HC	Healthy control
IFT	Inverse Fourier transform
JSD	Jensen-Shannon divergence
KLD	Kullback-Leibler divergence
LLDB	Local Laplacian filtering domain-based
MAD	Median absolute deviation.
MCI	Mild cognitive impairment
MI	Mutual information
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MSD	Multiscale decomposition
NSLP	Nonsubsampled Laplacian pyramid
NSP	Nonsubsampled pyramid
PDF	Probability density function
PET	Positron emission tomography
pMCI	Progressive mild cognitive impairment
PSNR	Peak signal to noise ratio

RNN	Recurrent neural networks
SF	Shearing filters
sMCI	Stable mild cognitive impairment
SSIM	Structural similarity index
TV	Total variation
WBA	Whole brain atlas

Chapter 1

Introduction

1.1 Neurodegenerative Diseases: A Global Challenge

The brain is the most complex organ in the human body, consisting of an intricate network of billions of neurons and trillions of interconnections between them. Neurodegenerative diseases refer to conditions involving the gradual degeneration and loss of neurons in the nervous system. This degeneration leads to a progressive decline in neurological function and often results in irreversible damage and cell death. Common examples of neurodegenerative diseases include Alzheimer's disease, Parkinson's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis, and prion diseases.

Over the past few decades, there has been a notable increase in the prevalence of neurodegenerative diseases, particularly Alzheimer's disease, accounting for an estimated 60% to 80% of all cases. In the United States alone, it is estimated that the number of people living with Alzheimer's disease is 6.7 million, and it is a huge concern that this number could grow to 13.8 million by 2060 [1]. Moreover, this surge in neurodegenerative diseases places an increasing burden on healthcare systems globally. As the demand for specialized care rises, it results in substantial costs and presents a significant global challenge.

The complexity of the brain and the devastating impact of neurodegenerative diseases

demand enhanced research and innovative approaches to diagnosis, treatment, and prevention of these diseases.

1.2 Advances in Neuroimaging Techniques

Neuroimaging techniques play a crucial role in the understanding and diagnosis of neurodegenerative diseases such as Alzheimer's disease. Structural magnetic resonance imaging (MRI), computed tomography (CT), functional MRI (fMRI), and positron emission tomography (PET) are some of the common neuroimaging techniques used to capture changes in the brain.

Structural MRI is one of the most widely used neuroimaging techniques due to its ability to provide detailed images of the structure of the brain. It utilizes powerful magnetic fields and radio waves to generate images of the brain, allowing clinicians and researchers to detect abnormalities such as shrinkage of hippocampus [2], enlarged verticles [3], cortical atrophy [4], the presence of lesions [5], and changes in white matter hyperintensities [6], all of which are indicative of neurodegenerative processes. CT scans also provide detailed images of the structure of the brain by combining multiple X-ray images taken from different angles. While not as commonly used for neurodegenerative diseases as structural MRI or PET, CT scans are valuable for ruling out diseases by detecting structural abnormalities such as tumors, hemorrhages, or atrophy in the brain.

Functional MRI (fMRI) is a specialized form of MRI that measures changes in blood flow and oxygenation levels in the brain, providing insights into neural activity and connectivity. By mapping patterns of brain activation during specific tasks or at rest, fMRI can help identify alterations in brain function associated with neurodegenerative diseases [7]. PET imaging enables the visualization of metabolic and molecular processes in the brain by detecting radioactive tracers injected into the body. This allows for the assessment of various aspects of neurodegeneration, including the buildup of protein aggregates such as

amyloid-beta and tau, which are the hallmarks of Alzheimer's disease [1].

Despite the availability of various neuroimaging techniques, diagnosing neurodegenerative diseases can still be challenging for several reasons:

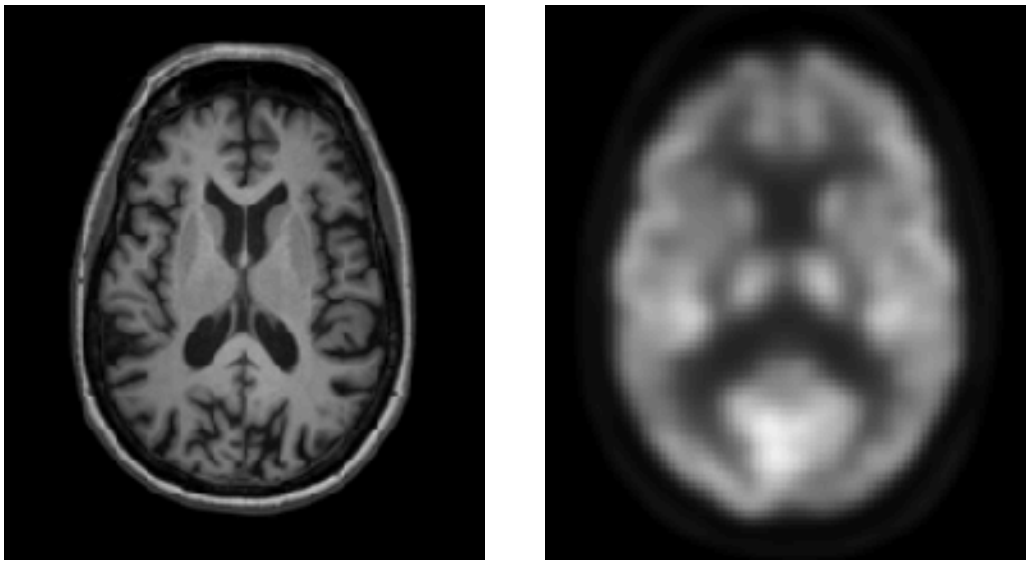
Complexity of Diseases: Neurodegenerative diseases are inherently complex and heterogeneous in nature. They often involve multiple pathological processes occurring simultaneously, such as protein aggregation, inflammation, and neuronal loss [8]. These processes can present differently in each individual, making it difficult to establish a definitive diagnosis based solely on information obtained from single neuroimaging technique.

Incomplete Information: Neuroimaging techniques, such as structural MR or PET imaging, provide valuable information about specific aspects of neurodegenerative diseases, such as brain structure or metabolic activity. However, each technique alone does not provide a comprehensive view of the disease process, leading to gaps in understanding and potentially missing important diagnostic markers.

Overlap of Symptoms: Many neurodegenerative diseases present with overlapping clinical symptoms, making it challenging to differentiate between them based solely on single neuroimaging technique. For example, both Alzheimer's disease and frontotemporal dementia can exhibit similar patterns of brain atrophy on structural MRI [9], [10]. Moreover, neurodegenerative diseases often occur in older adults who may also have other age-related conditions [11].

Interpretation Challenges: Without complementary insights from various imaging modalities, there is always a significant risk of misinterpreting the data or incompletely assessing the disease process. However, interpreting and correlating information obtained from single or multiple neuroimaging techniques can be challenging as it not only requires specialized expertise but is also subject to interpretation variability among clinicians.

While neuroimaging techniques have greatly enhanced our ability to diagnose and understand neurodegenerative diseases, diagnosing these conditions remains complex and



(a) (b)

Figure 1.1: Difference in structural and functional imaging of an Alzheimer's disease patient from ADNI database. (a) MR image using magnetization-prepared rapid gradient-echo (MP-RAGE) (b) Fludeoxyglucose (FDG) PET.

challenging due to the several reasons discussed above. Hence, it is essential to have a technique that combines the complementary information provided by each of the neuroimaging techniques for the diagnosis of neurodegenerative diseases.

1.3 The Role of Neuroimaging Fusion

Neuroimaging fusion is a technique that combines data obtained from multiple imaging modalities to improve diagnostic accuracy and provide a more comprehensive understanding of neurodegenerative diseases. By integrating information such as fusion of images obtained from various neuroimaging techniques offers several advantages over relying on images obtained from a single neuroimaging technique.

One of the primary benefits of neuroimaging fusion is the ability to complement the strengths of each imaging modality while mitigating their individual limitations. For example, structural MRI deals only with the valuable anatomical information of the brain.

As a result, the brain imaging data obtained from one of the above two techniques is not sufficient, in general, to carry out a proper diagnosis, and hence, data from multiple sources are acquired. As an example, structural MRI can detect changes in different regional structures such as lateral ventricle size with quite detailed resolution as shown in the Fig. 1.1(a). On the other hand, the functional neuroimaging technique deals only with the functional changes such as metabolic activity of neurons, local changes in blood flow, regional composition, and absorption of brain [12]. Moreover, it usually produces images of low resolution and fail to convey spatial information as shown in the Fig.1.1(b). In such situations, it is important to be able to see the relation between the anatomical details visible at one location in one modality and the functional details that are more visible at the corresponding location in the other modality for better diagnosis and surgical planning. It is not easy to see such a relation by observing multiple modalities and to interpret and correlate the information provided by each modality. The desire to visualize such complementary information in a single image has led to investigations in multimodal image fusion.

1.4 Brief Literature Review on Neuroimaging Fusion and Alzheimer's Disease Classification

A fused image obtained through a fusion of two or more different modalities is more suitable for human and machine perception or for further image-processing tasks such as segmentation, feature extraction and object recognition. However, the existing fusion algorithms face a number of challenges such as loss of significant attributes, staircase effects, blocking artifacts, noisy patches, and contrast reduction. For multimodal neuroimaging fusion, it is useful to have a precise comparison of brain functional details with its underlying structure, avoiding possible mistakes resulting from alternate viewing of two or more separate images. However, the manual way of integrating the information obtained from

multiple techniques is an extremely time-consuming and expensive process, which also demands years of experience to reduce human error. Hence, automated multimodal neuroimaging fusion is recognized as a promising solution to understand the different aspects of neurodegenarative diseases at an early stage [13].

In recent years, several techniques have been developed to handle the multimodal fusion problem for neuroimaging data [13]–[30] based on spatial, or transform domain approaches [31]. Despite these efforts, it still remains a challenging task and requires more research to overcome some of the drawbacks such as staircase effects, loss of significant attributes, blocking artifacts, and contrast reduction of the existing automated fusion algorithms.

In [16], the authors have developed a fusion method which is based on edge-preserving guided image filter. The idea is to first find two sharpened images by considering one as an input image and the other as a guidance image and vice versa. This is found by subtracting the input images from the blurred images which is obtained using guided image filtering. Then, the sharpened images are combined based on the weights obtained through image statistics. Although it provides better gradient information, it is clear that the intensity information could get lost. In addition, choosing a guidance image in case of fusing three or more images is a challenging task. Zhao *et al*. [18] have proposed an adaptive fractional order total variation (TV) technique to estimate the fused image in the spatial domain. It can suppress the noise including the staircase effects present in the original TV models [32], [33]; but it still suffers from contrast reduction, just as other spatial domain techniques do [13]. Given improved data representation, energy compaction, and reduced complexity, transform domain approaches are known to be more effective in handling multimodal image fusion problem [31]. The transform domain methods are generally categorized into sparse representation-based and multiscale decomposition-based methods. Liu *et al*. [21] have proposed a convolutional sparse representation-based fusion framework in which the

source images are modelled as the sum over a set of convolutions between sparse coefficient maps and dictionary filters. In [19], the source image patches have been classified based on geometrical features and coded sparsely via online multiclass dictionary learning algorithm. The sparse representation-based methods learn dictionary elements from the source images. However, in case of noisy input, a highly redundant dictionary may cause unwanted visual artifacts in the reconstructed output [34]. Recently, multiscale transforms such as local Laplacian pyramid [15], wavelet [17], [35], curvelet [30], [31], contourlet [13], [29], shearlet [27], or cascade of such transforms [25], [36] have also been employed to develop fusion algorithms for neuroimaging and other medical images. In local Laplacian pyramid-based method [15], the source images are first decomposed using local Laplacian filtering [37] into approximate and residual images at different scales. Then, the approximate and residual images are fused using a local energy maximum scheme and an information of interest-based scheme, respectively. However, it is known that the widely-used Laplacian pyramid and wavelet transform is not effective in representing anisotropic features such as edges of multidimensional data. To overcome this limitation, Candès et al. [38] proposed the simpler, faster, and less redundant curvelet representation to capture the geometric information in images. In [30], a curvelet based fusion is implemented based on the application of the additive wavelet transform on the source images. The maximum fusion rule is applied on the ridgelet transform of high-frequency bands. However, the implementation of curvelet in discrete domain is very challenging as it requires a rotation operation, which does not preserve the digital lattice [39]. To deal with this issue, two other multiresolution tools, contourlet [40] and shearlet [41] were introduced. Although these two transforms have good directional sensitivity, it is found that shearlet transform is more advantageous as there are no constraints on choosing the number of directions and the size of the supports for the shearing. It is also computationally more efficient unlike the usage of the directional filter banks in contourlet transform [41]. However, the downsamplers

present in the Laplacian pyramid makes the shearlet transform, which is constructed by combining the Laplacian pyramid and shearing filters, shift invariant resulting in pseudo-Gibbs phenomena around the singularities such as edges [42]. Nonsubampled shearlet transform (NSST) [26], [41], [43] resolves this artifact by using nonsubampled filters in shearlet transform and it is proved as an excellent multiscale decomposition tool for medical image fusion [14]. Since the distribution of the NSST domain coefficients of images has large peaks around zero and tails heavier than that of a Gaussian PDF, an appropriate choice of the distribution to model the statistics of the NSST domain coefficients would be the one having a heavy-tailed PDF. For this purpose, in [27], [35], a model-based fusion approach has been made to model the subbands coefficients using the generalised Gaussian distribution (GGD). However, the GGD model often exhibits exaggerated tails [44], and hence this PDF still cannot provide an accurate fit to the empirical density function of the images in the transform domain. Moreover, as the coefficients of the high-frequency subbands are complex numbers, one of the most commonly used procedures for finding the decision map is the absolute maximum selection rule [45]. Using this rule, a transform domain coefficient of the fused image is selected as the corresponding coefficient of the transform domain representation of one of the constituent images that has the largest magnitude. However, this approach leads to a fused image that may not necessarily have the highest possible energy.

Moreover, in recent years, convolutional neural networks (CNNs) have found widespread use in a variety of applications [46]. This is because CNN have several unique properties that make them well-suited to automatically learn hierarchical representations of relevant image features, and their robustness to translation, scaling, and noise. Furthermore, the advancements in hardware and computing power have enabled the training of deep and complex CNNs, leading to their improved performance for various applications. The research in detecting and classifying Alzheimer's disease using the image data obtained from

different neuroimaging techniques has greatly benefited from these advantages of the CNNs [47]–[55].

Gupta et al. [47] first train a sparse autoencoder [56] to learn a set of bases from randomly selected 2D patches of natural images to represent MRI data. Then, a set of feature maps is obtained using the slices of the MRI data and the bases learned from the sparse autoencoder as the weights for the filters of the 2D convolution layer. The feature maps extracted by the convolutional layer are then fed to a pooling layer for dimensionality reduction of the feature maps thus obtained. Finally, the pooled feature maps are fed to train a neural network for classifying a sample of the MRI data. To the best of our knowledge, this is the first work for the classification of Alzheimer's disease using CNN. However, as mentioned in [48], the features extracted from the unrelated source domain in this scheme degrade its classification performance. Moreover, fine-tuning models on the target data may lead to overfitting if there is a significant domain difference between the source and the target domains [57]. The work in [49] is similar to that in [47], but the authors use 3D patches of MRI data instead of 2D patches of natural images to learn the weights for the filters of the 3D convolutional layer. Hosseini-Asl et al. [50] have proposed a 3D CNN that combines pre-trained encoding layers of a 3D convolutional autoencoder as its lower layers with fully-connected upper layers. The autoencoder part of the network extracts generic features from the source domain and then the fully-connected upper layers are fine-tuned in the target domain for the classification of Alzheimer's disease. Sergey et al. [51] have proposed plain and residual 3D CNN architectures without relying on unsupervised autoencoder learning, as utilized in [47], [49], and [50]. Although the MR images of the whole brain are used to train the proposed CNNs in [50] and [51], the choice of the pooling layers in the network limits the capture of global features that are important for an accurate classification of Alzheimer's disease as concluded by the work in [58]. The authors in [58] have quantified the global and local gray matter (GM) volume to examine the differences in

MCI subjects compared to subjects with HC and AD. The global GM volume was assessed using segmentation and the local GM volume was assessed using a technique called voxel-based morphometry [59]. They have found that AD subjects had a significantly (12.3%) lower mean global GM volume when compared to that of HC subjects. Global GM volume in the MCI subjects was 6.2% lower than that of AD subjects and 6.5% higher than that of HC subjects. GM loss in the medial temporal region characterizes MCI, while GM loss in the parietal and cingulate cortices characterizes AD. Hence, for an accurate classification of Alzheimer's disease, it is not only crucial to have both the local and global features, but it is also important to have the features from different key regions of the brain. In the scheme [52], both types of features are confined only to the left and right hippocampi, and other brain regions are not considered for classification. The authors in [53], developed a hierarchical fully convolutional network (H-FCN) in which the discriminative local patches from all the regions of the brain are automatically identified unlike in [52] in which it is assumed that the informative location is only in hippocampi. However, there are certain drawbacks to consider in this scheme. One drawback is that the template does not always perfectly align with each subject's brain structure due to anatomical variations among individuals and image artifacts. This results in inaccuracies in the generation of patch locations and can adversely affect the classification performance by introducing irrelevant patches. Moreover, a selective number of patches used for feeding into the patch-level subnetwork from the registered image loses the ability to leverage contextual local information from all parts of the brain. The processing of the selected number of patches and the pruning of patch-level subnetworks reduce the complexity of the scheme; however, on the other hand, the computational complexity associated with the overhead of registration is also a significant concern. The authors in [54] have proposed an attention-based scheme called dual attention multi-instance deep learning (DA-MIDL) network, which still loses the contextual local features in view of its using only a subset of all the patches in an image. In

the network of the scheme [55], the entire brain image is fed to the network to automatically learn the features from all parts of the brain and an attention layer is employed only on the feature map produced by the last convolutional layer in each of the residual blocks to enhance the classification performance. By paying attention only to the output of the last convolutional layer of a residual block is perhaps not the best way of improving the classification performance. On the other hand, adding more attention layers in each of the residual blocks to improve the performance further would result in increasing the overall computational complexity of the network. It is also noted that this scheme employs limited contextual information, since the attention map is obtained using point convolution and not exploiting the neighbouring pixels for the query and keys at each spatial location.

1.5 Motivation and Objectives

It is seen from the literature review in Section 1.4 that the nonsubsampled shearlet transform (NSST) offers several advantages over traditional multiscale transforms, particularly in the context of neuroimaging fusion and analysis. However, the generalized Gaussian distribution used in model-based fusion approaches to characterize the non-Gaussian properties of shearlet coefficients does not adequately represent the actual distribution of the NSST coefficients of neuroimaging data. Moreover, a transform domain coefficient of the fused image is obtained as the corresponding coefficient of the transform domain representation of one of the constituent images that has the largest magnitude. Even though the energy level of the fused image so obtained is higher than that of the constituent images used for fusion, other fusion rules that could provide higher energy levels for the fused images remain to be investigated.

It is seen from the literature review that most of the existing schemes for the task of classification of Alzheimer's disease develop models that use a select number of patches, and therefore, the performance of these models depends on the selected patches. If one is

not able to select appropriate patches, the models may not capture crucial local and global contextual features from the brain. On the other hand, if one chooses to feed the network with the entire brain image to automatically learn the features from all parts of the brain, pooling layers used for dimensionality reduction may limit the classification performance due to the loss of contextual information. Moreover, the existing architectures often end up using a large number of parameters to enhance the richness of the extracted features.

The objective of this thesis is twofold: (i) developing a novel model-based scheme for the fusion of neuroimaging data based on the statistical properties of NSST coefficients of neuroimaging data as well as a novel fusion rule that maximizes the energy in the high-frequency transform coefficients, and (ii) developing a novel lightweight convolutional neural network (CNN) capable of extracting both the local and global contextual features and studying the performance of the CNN model for the binary and multiclass classifications of Alzheimer's disease using images of a given modality or images obtained by fusing images from one or more modalities.

1.6 Organization of the Thesis

The organization of this thesis is as follows. In Chapter 2, a brief review of the background materials in the context of developing the neuroimaging fusion scheme and the lightweight convolutional neural network (CNN) for the classification of Alzheimer's disease is presented. In Chapter 3, a neuroimaging fusion scheme, utilizing the statistical properties of nonsubsampled shearlet transform coefficients and a novel energy maximization fusion rule, is developed. In Chapter 4, a lightweight deep CNN, which processes separately the local and global features by using modules that pay special attention to local and global contextual features pertinent to the classification of Alzheimer's disease, is developed. The performance of the CNN model for the binary and multiclass classifications of Alzheimer's disease using single modality images is studied. In Chapter 5, the performance of the CNN

model for the classification of Alzheimer's disease using fused multimodal images obtained using the proposed fusion scheme is also investigated. Finally, in Chapter 6, some concluding remarks are made on the work undertaken in this thesis. Some suggestions for further investigations on the problems undertaken in this thesis are also made.

Chapter 2

Background Material

In this chapter, we provide a brief review of background material useful for understanding the work presented in this thesis. We start with a generic multiscale decomposition based image fusion followed by an overview of commonly used multiscale transforms such as the pyramid transform, discrete wavelet transform, non-subsampled contourlet transform, and non-subsampled shearlet transform. Next, we discuss the commonly used fusion rules applied in image fusion within the transform domain. Further, we discuss convolution operations, including standard and separable convolution, as well as pooling operations in deep convolutional neural networks. These operations are fundamental processes in deep networks, enabling the extraction of features and reduction of dimensionality while preserving crucial information. Finally, we describe the metrics used for evaluating the performance of the fusion and classification schemes.

2.1 Commonly Used Multiscale Transforms

Various multiscale transforms have been developed for decomposing images into different scales and orientations. Each transform has its own characteristics and is suitable for specific types of images and applications. In this section, we provide an overview of some

commonly used multiscale transforms.

2.1.1 2D Discrete Wavelet Transform

The 2D Discrete wavelet transform (DWT) is a multiscale transform commonly used in image processing. It decomposes a signal into frequency subbands at different scales from which it can be perfectly reconstructed. Figure 2.1 illustrates the 2D DWT process, where the image I is decomposed into its four components at the first level. Subsequently, the LL subband resulting from the first-level decomposition is further decomposed into its own LL, LH, HL, and HH subbands at the second level. Here, the LL represents the approximation coefficients, LH represents the horizontal detail coefficients, HL represents the vertical detail coefficients, and HH represents the diagonal detail coefficients.

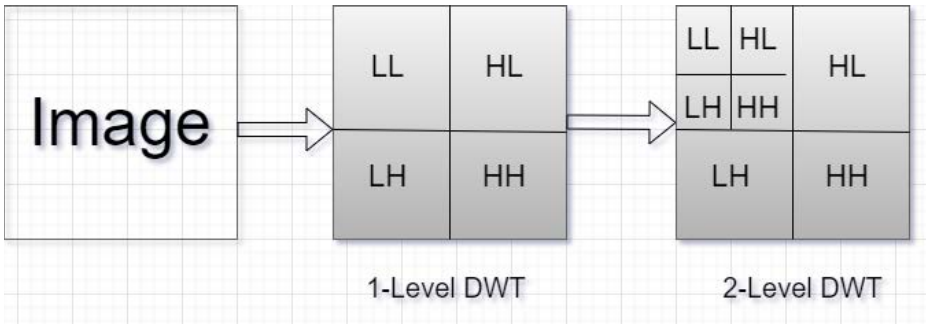


Figure 2.1: Illustration of the 2D Discrete Wavelet Transform.

2.1.2 Non-Subsampled Contourlet Transform

Figure 2.2 shows the decomposition framework of NSCT. Nonsubsampled pyramid (NSP) structure and nonsubsampled DFB are used in NSCT. The NSP is achieved by using two-channel nonsubsampled 2-D filter banks that ensure the multiscale property. The DFB is achieved by switching off the downsamplers/upsamplers in each two-channel filter bank in the DFB tree structure, which gives directionality. Although NSCT yields better frequency selectivity, regularity, and shift-invariance, it has restrictions on the number of directions,

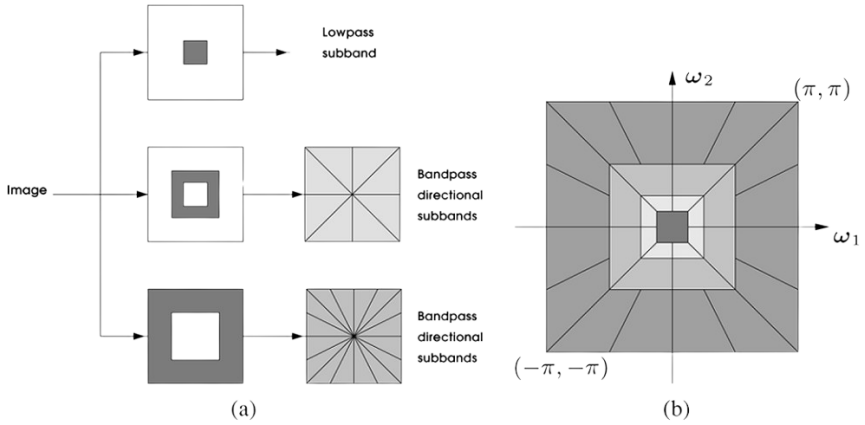


Figure 2.2: Nonsubsampled contourlet transform [60]. (a) NSFB structure that implements the NSCT. (b) Frequency partitioning.

and its computational complexity is exceedingly high. The implementation, applications, and other properties of NSCT are well explained in [60].

2.1.3 Nonsubsampled Shearlet Transform

The nonsubsampled shearlet transform, a multiscale, multidirectional, and shift invariant framework proposed in [41], is found to be a highly efficient multiresolution tool to provide an optimal sparse approximation for multidimensional signals with anisotropic features such as edges and other spatial discontinuities present in images. In this section, we briefly discuss the implementation of NSST and the readers can find the complete details in [41]. In fact, shearlet is an expansion of composite wavelets introduced in [61]. Consider the 2-D affine system generated by $\upsilon \in L^2(\mathbb{R}^2)$, a collection of functions obtained by applying dilation and translation of the form:

$$\{\upsilon_{a,s,t}(m) = |\det G_{a,s}|^{\frac{1}{2}} \upsilon(G_{a,s}^{-1} m - t) : \\ a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2\} \tag{2.1}$$

where $G_{a,s} = \begin{pmatrix} a & -\sqrt{a}s \\ 0 & \sqrt{a} \end{pmatrix}$ is the dilation group which is a composition of the anisotropic dilation matrix $\begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$ and the shear matrix $\begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix}$. The mapping between the original function and the transform domain in continuous shearlet transform (CST) is defined by

$$S\mathcal{H}_{\upsilon} : x \to S\mathcal{H}_{\upsilon}x(a, s, t) = \langle x, \upsilon_{a,s,t} \rangle \tag{2.2}$$

The generating function υ is a well-localized band-limited function and adequately meets the admissibility condition [62] associated with the shearlet approach so that $x \in L^2(\mathbb{R}^2)$ can be reconstructed by

$$x = \int_{\mathbb{R}^2} \int_{-\infty}^{\infty} \int_{0}^{\infty} \langle x, \upsilon_{a,s,t} \rangle \upsilon_{a,s,t} \, \frac{da}{a^3} \, ds \, dt \tag{2.3}$$

For any $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$, $\omega_1 \neq 0$, let Υ is chosen to be of the form:

$$\Upsilon(\omega) = \Upsilon_1(\omega_1)\Upsilon_2\left(\frac{\omega_2}{\omega_1}\right) \tag{2.4}$$

where Υ_1 and Υ_2 are smooth functions in which the supports are the proper subset of $[(-2, -\frac{1}{2})] \cup [\frac{1}{2}, 2]$ and $[-1, 1]$, respectively. In the Fourier domain,

$$\Upsilon_{a,s,t}(\omega_1, \omega_2) = a^{\frac{3}{4}}\Upsilon_1(a\omega_1)\Upsilon_2\left(a^{-}\frac{1}{2}\left(\frac{\omega_2}{\omega_1} - s\right)\right)e^{-2\pi i\omega t} \tag{2.5}$$

where $\Upsilon_{a,s,t}(\omega_1, \omega_2)$ has frequency support in which ω_1 is in the set $[(-\frac{2}{a}, -\frac{1}{2a})] \cup [(\frac{2}{a}, \frac{1}{2a})]$ and $|\frac{\omega_2}{\omega_1} - s| \leq \sqrt{a}$. Shearlet is a function of three parameters, namely, the scale a, the shear s, and, the translation t.

The discrete shearlet transform is obtained by dicretizing the CST $S\mathcal{H}_{\upsilon}x(a, s, t) = \langle x, \upsilon_{a,s,t} \rangle$ on suitable discrete set (scaling, shear, and translation parameters) [63]. NSST

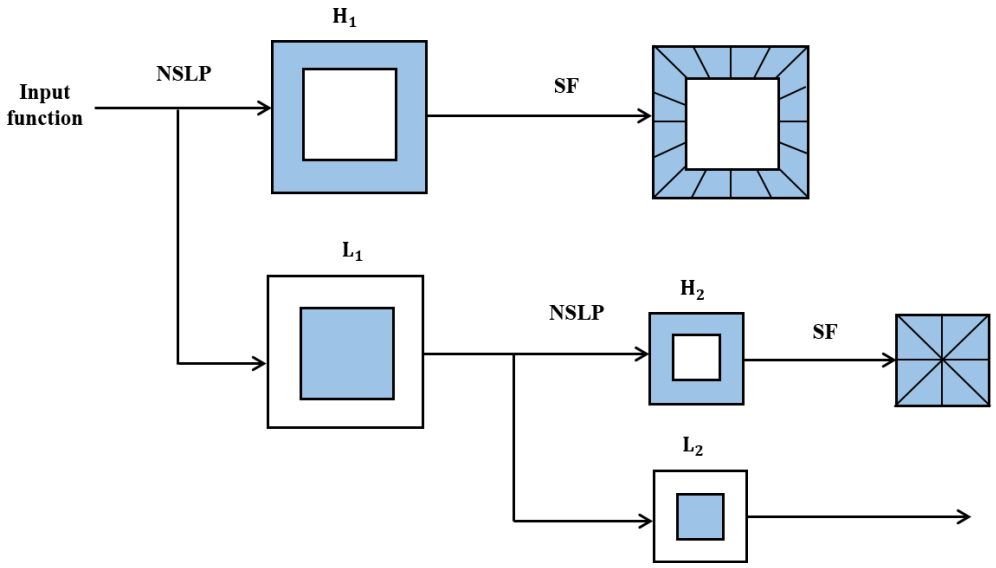


Figure 2.3: Block diagram of the NSST multiscale and multidirectional 2-scale decomposition. The nonsubsampled Laplacian pyramid (NSLP) is applied to the input function resulting in a low frequency subband, denoted by L, and a detailed subband, denoted by H. The detailed subband is fed into the shearing filter bank to obtain directional information.

is a special type of discrete shearlet transfrom and its decomposition is shown in Fig. 2.3. This figure depicts a two-scale ($a = 2$) decomposition, where the shear parameter s is chosen to be 8 and 4 at scales 1 and 2, respectively. The basic structure of NSST includes the nonsubsampled Laplacian pyramid (NSLP) transform, which establishes the multiscale property, in combination with shearing filters (SF) that offer a higher amount of directional information. Here, the use of nonsubsampled filters, without upsamplers or downsamplers, ensure the invariance property to the shift of the input functions.

2.2 Multiscale Decomposition based Fusion

Multiscale decomposition based fusion methods play a crucial role in combining information from multiple source images to generate a single fused image with enhanced features and details. As illustrated in Fig. 2.4, which depicts a typical flow diagram of image fusion based on multiscale decomposition (MSD), the process typically involves the following

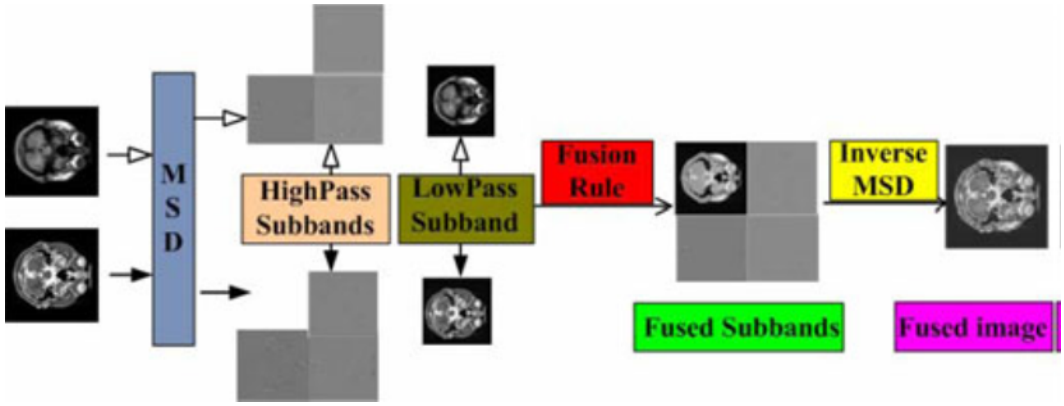


Figure 2.4: The generic framework of image fusion based on multiscale decomposition [64].

steps:

(1) **Decomposition:** The source images are decomposed into multiscale coefficients using a suitable multiscale transform. These transforms are designed to capture information at different scales and orientations, allowing for the extraction of both local and global features.

(2) **Fusion Rule Application:** Once the multiscale coefficients are obtained, fusion rules are applied to combine them. These fusion rules determine how information from different scales and orientations should be combined to produce the final fused image.

(3) **Inverse Transform:** After applying the fusion rule, the fused multiscale coefficients are inverse transformed to obtain the fused image in the spatial domain. This process reconstructs the fused image while preserving important spatial structures and details.

Thus, in multiscale decomposition, an input image can be represented in the transform domain by a sequence of detail images along with an approximation image at the coarsest scale. Hence, the multiscale decomposition of an input image I is represented as:

$$I = \{D^1, D^2, \ldots, D^L, A^L\} \tag{2.6}$$

where A^L represents the approximation image at the lowest scale L and $D^j, j = 1, 2, 3, \ldots, L$ represents the detail images at level j. The detail images include various directional bands depending on the type of multiscale transforms.

Multiscale decomposition-based fusion techniques offer several advantages, including the ability to capture and integrate information from different scales and orientations, which can lead to improved spatial resolution and enhanced feature representation. In the next section, we will discuss common fusion rules used in image fusion within the transform domain.

2.3 Commonly Used Fusion Rules in Image Fusion

In the transform domain, the fusion rule is used to obtain the fused coefficient by combining the corresponding coefficients from input images. In this section, some of the commonly used fusion rules are discussed.

2.3.1 Weighted Averaging

The weighted averaging fusion rule [65] combines the transformed coefficients of input images using weighted sums. Each transformed coefficient $I_{\text{fused}}(i, j)$ in the fused image is computed as follows:

$$I_{\text{fused}}(i, j) = \sum_{k=1}^{K} w_k \cdot I_k(i, j) \tag{2.7}$$

where:

- $I_{\text{fused}}(i, j)$ is the fused coefficient at position (i, j),
- $I_k(i, j)$ is the transformed coefficient of the k-th input image at position (i, j),
- w_k is the weight assigned to the k-th input image.

2.3.2 Local Standard Deviation

The fusion rule based on local standard deviation [66] selects coefficients by considering the standard deviation of the coefficients in a local neighborhood. Mathematically, the fused coefficient $I_{\text{fused}}(i, j)$ is determined as follows:

$$I_{\text{fused}}(i, j) = \arg\max_{I_k} \{\text{stddev}(C_k(i, j))\} \tag{2.8}$$

where stddev$(\cdot)$ denotes the standard deviation function and $\arg\max$ selects the coefficient yielding the maximum standard deviation in the local neighborhood.

2.3.3 Average Gradient

The fusion rule based on average gradient [66] selects coefficients by considering the average gradient magnitude in the local neighborhood. Mathematically, the fused coefficient $I_{\text{fused}}(i, j)$ is determined as follows:

$$I_{\text{fused}}(i, j) = \arg\max_{I_k} \{\text{avg\_gradient}(I_k(i, j))\} \tag{2.9}$$

where avg_gradient$(\cdot)$ denotes the average gradient magnitude function and $\arg\max$ selects the coefficient yielding the maximum average gradient magnitude in the local neighborhood.

2.3.4 Absolute Maximum Rule

The absolute maximum fusion rule [66] selects the coefficient with the maximum absolute value among the corresponding transformed coefficients of the input images at each location in the transformed domain. Mathematically, the fused coefficient $I_{\text{fused}}(i, j)$ is

computed as follows:

$$I_{\text{fused}}(i,j) = \max(|I_1(i,j)|, |I_2(i,j)|, \ldots, |I_K(i,j)|) \tag{2.10}$$

where $|\cdot|$ denotes the absolute value function.

2.4 Convolution and Pooling Operations in CNNs

In this section, we discuss briefly the commonly used convolution operations and the pooling operations.

2.4.1 Standard Convolution

Standard convolution is a fundamental operation in convolutional neural networks (CNNs) used for feature extraction. Given an input image I and a filter F, the standard convolution operation computes the output feature map O as follows:

$$O(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot F(m,n) \tag{2.11}$$

where:

- $O(i,j)$ is the value of the output feature map at position (i,j),
- $I(i+m, j+n)$ is the pixel value of the input image at position $(i+m, j+n)$,
- $F(m,n)$ is the value of the filter at position (m,n).

2.4.2 Separable Convolution

Separable convolution is a variant of standard convolution that factorizes a 2D convolution into two 1D convolutions. It decomposes the convolution operation into two simpler operations: a row-wise convolution and a column-wise convolution. This approach can significantly reduce the number of parameters and computations required compared to standard convolution.

Let F_r be a 1D filter applied along the rows of the input image and F_c be a 1D filter applied along the columns. The output of the separable convolution O can be computed as follows:

$$O(i, j) = \sum_{m} I(i + m, j) \cdot F_r(m) \cdot F_c(n) \tag{2.12}$$

where:

- $O(i, j)$ is the value of the output feature map at position (i, j),
- $I(i + m, j)$ is the pixel value of the input image at position $(i + m, j)$,
- $F_r(m)$ is the value of the row-wise filter at position m,
- $F_c(n)$ is the value of the column-wise filter at position n.

Separable convolution can offer computational advantages, especially in scenarios where reducing the number of parameters and computations is crucial, such as in deep neural networks deployed on resource-constrained devices.

2.4.3 Pooling Operations

Pooling operations in CNNs are used for dimensionality reduction and feature aggregation. Max pooling and average pooling are two common pooling operations.

Max Pooling

Max pooling divides the input feature map into non-overlapping regions and selects the maximum value from each region to form the output feature map. Given an input feature map I and a pooling window size $S \times S$, the max pooling operation computes the output feature map O as follows:

$$O(i,j) = \max_{m,n} I(S \cdot i + m, S \cdot j + n) \tag{2.13}$$

where:

- $O(i,j)$ is the value of the output feature map at position (i,j),
- $I(S \cdot i + m, S \cdot j + n)$ is the pixel value of the input feature map in the pooling window at position $(S \cdot i + m, S \cdot j + n)$.

Average Pooling

Average pooling computes the average value of each non-overlapping region in the input feature map to form the output feature map. Given an input feature map I and a pooling window size $S \times S$, the average pooling operation computes the output feature map O as follows:

$$O(i,j) = \frac{1}{S^2} \sum_{m,n} I(S \cdot i + m, S \cdot j + n) \tag{2.14}$$

where the symbols have the same meanings as in max pooling.

2.5 Commonly Used Evaluation Metrics

In this section, the commonly used evaluation metrics for image fusion and the classification are discussed.

2.5.1 Metrics for Fusion

For quantitative analysis, peak signal to noise ratio (PSNR), structural similarity (SSIM), mutual information (MI), entropy and $Q^{\frac{f_1 f_2}{F}}$ are the commonly used performance metrics for fusion, and these are briefly defined next.

The SSIM measure is a function of three independent components, namely, luminance (l), contrast (c), and structure (s).

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{2.15}$$

where α, β, and γ are the quantities used to calibrate the relative importance of these particular components. We choose $\alpha = \beta = \gamma = 1$; the readers are referred to [67] concerning the details on the choice of the parameters and the steps involved in the calculation of the SSIM metric.

In the context of image fusion, the MI metric measures the amount of information the fused image conveys about each of the source images. Considering two input images $f_1(x,y)$ and $f_2(x,y)$ and a fused image $F(x,y)$, we can calculate how much information $F(x,y)$ holds from $f_1(x,y)$ and $f_2(x,y)$ based on Kullback-Leibler measure.

$$I(F,f_k) = \sum_{i=1}^{L}\sum_{j=1}^{L} p_{Ff_k}(i,j)\log\frac{p_{Ff_k}(i,j)}{p_F(i)p_{f_k}(j)} \quad k = 1,2 \tag{2.16}$$

where $p_{Ff_k}(i,j)$ is the joint histogram of images $F(x,y)$ and $f_k(x,y)$, p_F and p_{f_k} are the normalized histograms of $F(x,y)$ and $f_k(x,y)$, respectively, and L is the number of intensity values. Thus, the image fusion MI measure is defined by simply taking the average of the composite image $F(x,y)$ and each of the input images $f_1(x,y)$, and $f_2(x,y)$, i.e.,

$$MI_F^{f_1 f_2} = \frac{I(F,f_1) + I(F,f_2)}{2} \tag{2.17}$$

Higher the value of MI better the quality of the fused image. Another metric that is used to measure the quality of fusion is normalized weighted performance metric $Q^{\frac{f_1 f_2}{F}}$. It calculates the amount of edge information transfers from each of the input images into the fused image. The normalized weighted performance measure $Q^{\frac{f_1 f_2}{F}}$ of a given fused image F that operates on input images f_1 and f_2 is defined as follows:

$$Q^{\frac{f_1 f_2}{F}} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} Q^{f_1 F}(x,y)w^{f_1}(x,y) + Q^{f_2 F}(x,y)w^{f_2}(x,y)}{\sum_{i=1}^{M}\sum_{j=1}^{N} w^{f_1}(i,j) + w^{f_2}(i,j)} \tag{2.18}$$

where $w^{f_1}(m,n)$ and $w^{f_2}(m,n)$ are the edge strength values of the input images f_1 and f_2, respectively, at each pixel location (x,y) and are used as weight factors in (2.18). $Q^{f_1 F}(x,y)$ and $Q^{f_2 F}(x,y)$ are the edge information preservation values of f_1 and f_2, respectively, in the fused image F at each pixel location (x,y). Each of these quantities is the product of the edge strength preservation values and the orientation preservation values calculated by using the Sobel operator [68], [69]. The value of the normalized weighted performance measure $Q^{\frac{f_1 f_2}{F}}$ ranges between 0 (complete loss of edge information) and 1 (no loss of edge information) [69]. As explained in [64], for color images, we first calculate the performance measures SSIM, MI, and $Q^{\frac{f_1 f_2}{F}}$ of an individual channel separately using (2.15), (2.17) and (2.19), respectively, followed by averaging out over the number of channels. As an example, for color images such as RGB, we calculate MI by considering the red, green and blue color components of each individual pixel so that (2.17) becomes

$$MI_F^{f_1 f_2} = \frac{1}{3}\sum_{i}^{R,G,B} \frac{I^i(F,f_1) + I^i(F,f_2)}{2} \tag{2.19}$$

The other common performance measures, PSNR and entropy, are also calculated in the same manner for the color images.

2.5.2 Metrics for Classification

Next, to evaluate the performance of the classification, we employ four different metrics, namely, accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PREC), F1-score (F1), and area under the ROC curve (AUC). For each class, these metrics are defined as:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2.20}$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.21}$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2.22}$$

$$\text{PREC} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.23}$$

$$\text{F1} = \frac{2 \times \text{PREC} \times \text{SEN}}{\text{PREC} + \text{SEN}} \tag{2.24}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative values, respectively. For a given class, TP refers to the number of samples being predicted correctly to belong to the class in question, TN refers to the number of samples being predicted correctly to belong to a class not in question. FP refers to the number of samples being predicted incorrectly to belong to the class in question, and FN refers to the number of samples being predicted incorrectly to belong to a class not in question. These quantities are calculated using a confusion matrix, which is defined as an $n \times n$ matrix whose (i, j)th entry is the number of samples that actually belong to the ith class but have been predicted to belong to the jth class. In addition to the above metrics, area under the curve (AUC) is a commonly used metric to measure the tradeoff between sensitivity and

specificity, and it is the total area under the TP vs. FP curve.

2.6 Summary

In this chapter, a brief review of background material pertinent to the work of this thesis has been provided. Multiscale transforms such as discrete wavelet, non-subsampled contourlet, and shearlet transforms, which are commonly used for image fusion based on multiscale decomposition, have been discussed. Then, the steps involved in the process of image fusion based on multiscale decomposition have been described. Various commonly used fusion rules in image fusion have also been presented. Moreover, the standard and separable convolution operations, along with commonly used pooling operations in deep convolutional neural networks, have been explained. Finally, the commonly used evaluation metrics to assess the performance of image fusion and classification algorithms have been presented in detail.

Chapter 3

Neuroimaging Fusion in Nonsubsampled Shearlet Domain using Location-scale Distribution by Maximizing the High Frequency Subband Energy

3.1 Introduction

In this chapter, we introduce a Bayesian inference-based method for fusing neuroimaging data obtained from an arbitrary number of modalities [70], [71]. This method relies on the statistical properties of nonsubsampled shearlet transform (NSST) coefficients and employs a novel energy maximization fusion rule. The marginal distributions of the high-frequency NSST coefficients exhibit heavier tails than the Gaussian distribution. As a consequence, we use a heavy-tailed probability density function, student's t location-scale distribution, to describe the highly non-Gaussian statistics of empirical NSST coefficients by estimating

the parameters using maximum likelihood estimation. We then employ this model to develop a maximum a posteriori estimator to obtain the noise free coefficients. In view of the fact that the high-frequency subbands contain highly discriminative features such as edges, corners, blobs, and ridges, a fusion rule for obtaining the fused NSST coefficients based on maximizing the energy in all such subbands is also developed. Experiments are carried out on fusing two or more neuroimages taken from the BrainWeb, Alzheimer's Disease Neuroimaging Initiative (ADNI) and Whole Brain Atlas databases.

3.2 Multiscale and Multidirectional Image Representation using Nonsubsampled Shearlet Transform

The nonsubsampled shearlet transform (NSST) as described in Section 2.1.3 equipped with the following properties:

(a) Efficient computation: It exhibits rich mathematical structure, and the shearing filters have smaller support sizes than that of the directional filters used in the contourlet transform so that they can be implemented much more efficiently.

(b) Better representation: It has the features of directionality, localization, shift invariance, anisotropy, and multiscale.

(c) It does not have any restrictions on the number of directions for the shearing, as well as the size of the supports.

The NSST transform is employed in the proposed fusion method to decompose a given image into multiscale and multidirectional subbands. Given an image I of size $N \times N$, the procedure for finding the NSST at fixed scale l is summarized below [41].

(1) Apply the NSLP to decompose I into a low-pass image L_j^l and a high-pass image H_j^l;

(2) Compute $P\hat{f}_d^l$, the discrete Fourier transform of H_j on the pseudo-polar grid;

(3) Apply band-pass filtering to $P\hat{f}_d^l$;

(4) Apply inverse fast Fourier transform (IFT) after re-assembling the samples on a Cartesian grid.

As an example, the shearing filters (wedge shaped) computed using the Meyer wavelet function are shown in Fig. 3.1. The shearing filters of sizes 64×64 and 128×128 are used in the first and second scale of decomposition, respectively, with the number of shearing directions chosen to be 4 and 6 from small scale to large scale. The two-scale NSST decomposition of an MR image using maxflat filters in the NSLP stage is shown in Figs. 3.2 (b)-(l). Fig. 3.2 (b) is the low frequency sub-band, followed by the high frequency subbands at scale 1 and scale 2 in Figs. 3.2 (c)-(f) and Figs. 3.2 (g)-(l), respectively.

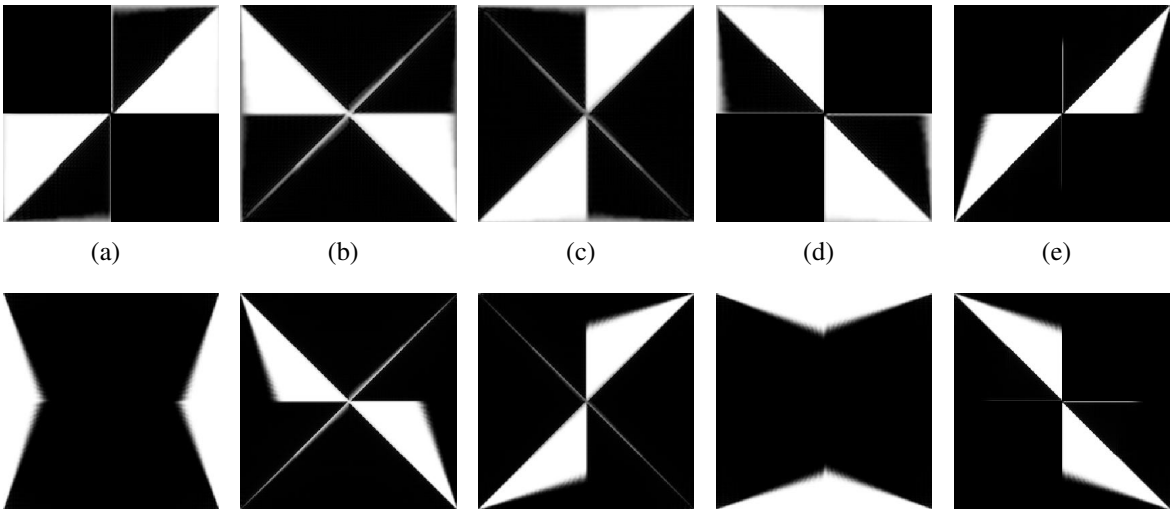


Figure 3.1: Examples of images of the shearing filters. Each detail subspace in NSST is chosen to be represented by four directions (a)-(d) in the first scale of decomposition and six directions (e)-(j) in the second scale of decomposition.

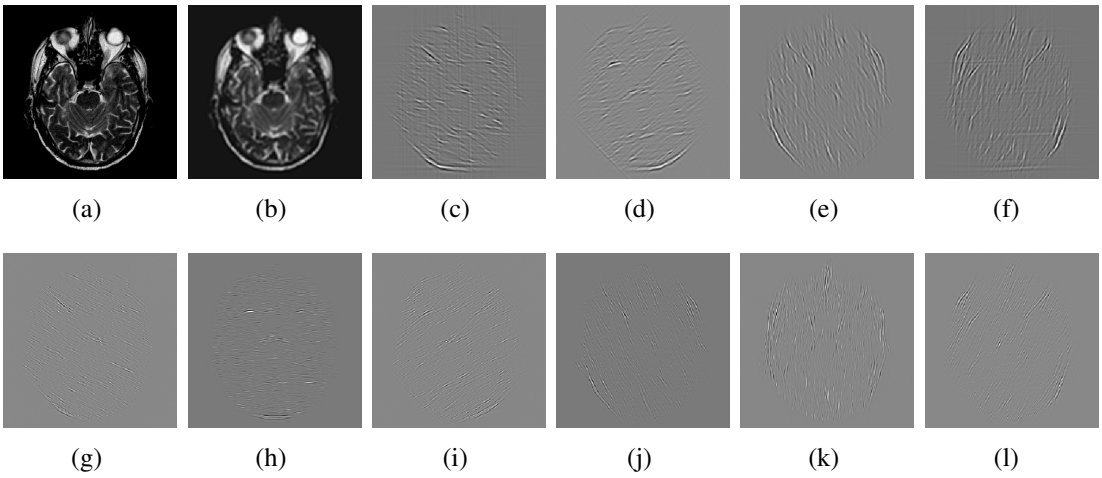


Figure 3.2: An example of two pyramidal scales nonsubsampled shearlet transform (NSST) decomposition. (a) MR image of an Alzheimer's disease patient. (b) The low-frequency subband. (c)-(f) The high-frequency subbands at scale 1. (g)-(l) The high-frequency subbands at scale 2.

3.3 The Forward Likelihood Model

Let $f_i(x, y)$ represent the unknown heterogeneous features such as structural and functional information of the brain, which are acquired by M different sensors and $g_i(x, y)$ the observed measurements. We consider neuroimaging fusion as the inverse of a linear forward model[32] that relates $f_i(x, y)$ and the observed brain images $g_i(x, y)$ by the following:

$$g_i(x, y) = f_i(x, y) + n_i(x, y) \qquad i = 1, 2, ..., M \tag{3.1}$$

where $n_i(x, y)$ is the noise associated with the i^{th} sensor at location (x, y). In general, signal fluctuations originate in the physical processes of imaging rather than in the tissue textures. In MRI, the acquired real and imaginary data in the frequency domain are known to be corrupted by white noise having Gaussian probability distribution. The frequency domain data is transformed into Cartesian domain via inverse Fourier transform (IFT). After IFT, the real and imaginary parts are still corrupted by Gaussian noise, since the transform is linear and orthogonal. However, in reconstructed PET images, unlike MRI,

the exact nature of the noise propagation is not well known. Hence, in this model, we assume that the target brain image is corrupted by additive white Gaussian noise with zero mean and a known standard deviation σ_n on each acquisition. Also, it is necessary that the input images are spatially aligned for developing image fusion algorithms [72]. Then, we apply non-subsampled shearlet transform (NSST) to all the observed images $g_i(x, y)$ to be fused. In general, if an image is decomposed into k scales and $D(k)$ denotes the number of directions chosen at scale k, the NSST transform provides a low frequency subband at the finest scale and $D(k)$ high frequency subbands at each scale k, in which the size of the subbands are similar to the size of the input image. We can write the model in (3.1) by an equivalent frequency domain representation:

$$G_i(u, v) = F_i(u, v) + N_i(u, v) \quad i = 1, 2, ..., M \tag{3.2}$$

where

$$G_i(u, v) = \begin{cases} G_i^L(u, v) & \text{for the low frequency subband} \\ G_i^{k,d}(u, v) & \text{for the high frequency subbands} \end{cases}$$

where

$$G_i^L(u, v) = F_i^L(u, v) + N_i^L(u, v)$$
$$G_i^{k,d}(u, v) = F_i^{k,d}(u, v) + N_i^{k,d}(u, v)$$

$$\begin{cases} i = 1, 2, ..., M \\ k = 1, 2, ..., K \\ d = 1, 2, ..., D(k). \end{cases}$$

where the terms in the upper case letters are the NSSTs of the corresponding terms in (3.2) and M, K and $D(k)$ denote the number of source images, number of scales and number of directions at each scale, respectively. The goal of the proposed neuroimaging fusion algorithm is to estimate the fused coefficients that combine the most significant information of $F_i^{k,d}(u,v)$ from the NSST coefficients of the given brain images, $g_i(x,y)$. Since we assume the probabilistic model associated with noisy NSST coefficients, $N_i^{k,d}(u,v)$, conditioned on $F_i^{k,d}(u,v)$ is Gaussian, the conditional probability is given by:

$$\begin{aligned} p(G_i(u,v)|F_i(u,v)) &= p_{N_i(u,v)}(G_i(u,v) - F_i(u,v)) \\ &= \mathcal{N}(0,\, \sigma^2_{N_i(u,v)}) \end{aligned}$$

which is proportional to

$$exp\frac{(G_i(u,v) - F_i(u,v))^2}{2\sigma^2_{N_i(u,v)})} \tag{3.3}$$

We use a robust median estimator [73] to estimate the noise variance $\sigma^2_{N_i(u,v)}$ which is given by:

$$\sigma^2_{N_i(u,v)} = MAD\frac{|Detail\ subbands\ of\ G_i(u,v)|}{0.6745} \tag{3.4}$$

where MAD is the median absolute deviation and $G_i(u,v)$ denotes the NSST coefficients of the observed neuroimaging data.

3.4 Proposed Neuroimaging Fusion Algorithm

In this section, we compute the fused NSST coefficients statistically using Bayes' rule with some knowledge of the prior distribution. Fig.3.3 shows a schematic view of different modules involved in the proposed Bayesian inference-based multimodal neuroimaging fusion algorithm.

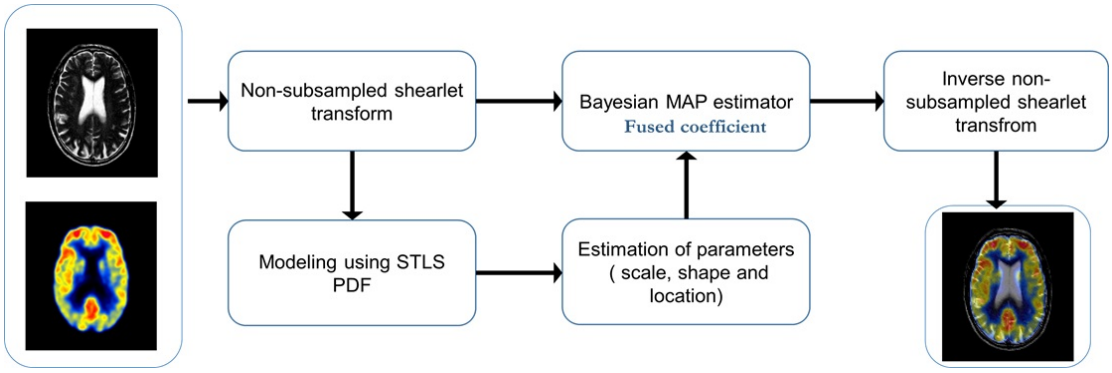


Figure 3.3: Block diagram of the proposed neuroimaging fusion method. Fusion takes place on individual channel for color images. STLS PDF stands for Student's t location-scale PDF.

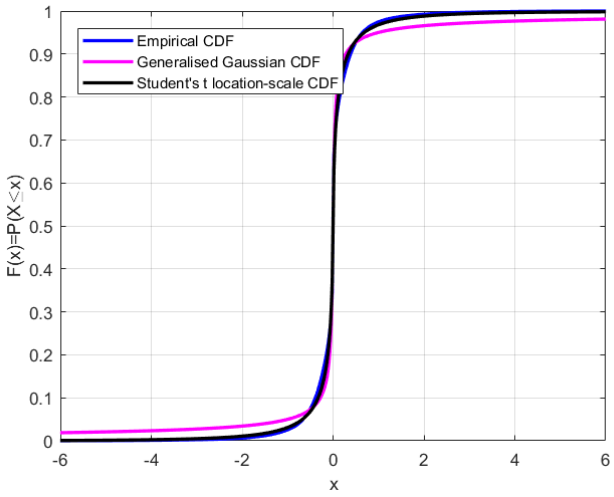


Figure 3.4: Empirical, generalized Gaussian and student's t location-scale CDFs for one of the NSST detail subbands of MR image of an AD patient.

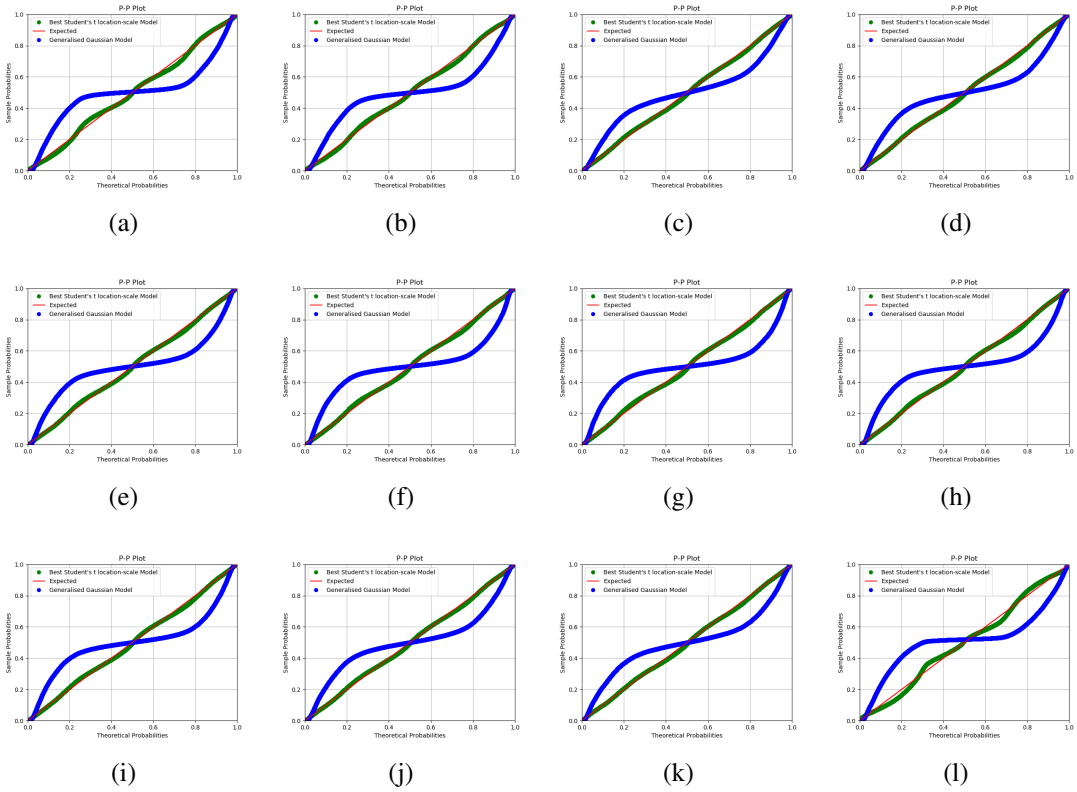


Figure 3.5: P-P plots for empirical, student's t location-scale and GG distribution for four NSST detail subbands at the finest scale. (a)-(d) for MR image, (e)-(h) for FDG-PET image and (i)-(l) for SPECT neuroimaging data.

3.4.1 Statistical Modeling of NSST Coefficients of Neuroimaging Data

At first, we propose to use the student's t location-scale PDF as a prior for modeling the heavy tailed nature of the NSST coefficients corresponding to the unknown target brain images under the assumption of independent identically distributed subbands [17]. Let x be a student's t location-scale distributed random variable; its PDF is described by

$$p(x|\nu,\mu_t,\sigma_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma_t\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})}\left[\frac{\nu+(\frac{x-\mu_t}{\sigma_t})^2}{\nu}\right] \tag{3.5}$$

where $\sigma_t > 0$ is the scale parameter, μ_t is the location parameter and $\nu > 1$ is the shape parameter, which determine the nature of the distribution. The shape parameter, ν is the most important parameter in controlling the shape of the distribution. When $\nu > 0$, the

three-parameter student's t location-scale distribution behaves like a traditional one parameter student's t distribution. When $\nu \to \infty$, it approaches the Gaussian distribution. The smaller values of the shape parameter yields a sharp peak around zero. To validate the behavior of the proposed prior PDF qualitatively, we examine the cumulative distribution function of the empirical real NSST coefficients with the student's t location-scale and the generalised Gaussian (GG) PDF's of the neuroimaging data. We examine the statistical properies of NSST coefficients of different multimodal neuoimaging data. Fig. 3.4 emphasizes the modeling performance of the real NSST coefficients of MR neuroimaging data of an AD patient. In addition, Fig.3.5 shows the qualitative fitting results in terms of the P-P plot of empirical, student's t location-scale and GG for four NSST detail subbands of multimodal neuroimaging data such as MR, FDG-PET, and SPECT images. The above two figures show that the student's t location-scale distribution provides a better fit in comparison with that provided by the GG distribution.

Parameter Estimation

In order to use student's t location scale distribution as a prior for estimating the fused coefficients, it is necessary to estimate the parameters from the NSST coefficients of the neuroimaging data. We employ the maximum likelihood estimation method [74] for learning the location, scale and shape parameters of the distribution. The maximum likelihood estimates are obtained through the expectation maximization (EM) algorithm [75], where the location and scale parameters can be estimated as

$$\mu_t^{(k+1)} = \frac{\sum_{i=1}^{n} W_i^{(k)} X_i}{\sum_{i=1}^{n} W_i^{(k)}} \tag{3.6}$$

$$\sigma_t^{(k+1)} = (1/n) \sum_{i=1}^{n} W_i^{(k)} (X_i - \mu_t^{(k)})(X_i - \mu_t^{(k)})^T \tag{3.7}$$

Table 3.1: Goodness-of-fit Measures in Terms of Jensen-Shannon Divergence

Subbands		MR		FDG-PET		SPECT	
		GGD	**STLSD***	**GGD**	**STLSD***	**GGD**	**STLSD***
I	**1**	0.0045	**0.0020**	0.0345	**0.0021**	0.0223	**0.0028**
	2	0.0049	**0.0029**	0.0228	**0.0022**	0.0342	**0.0023**
	3	0.0045	**0.0015**	0.0045	**0.0023**	0.0089	**0.0031**
	4	0.0058	**0.0023**	0.0082	**0.0081**	0.0065	**0.0034**
	5	0.0068	**0.0031**	0.0098	**0.0090**	0.0089	**0.0067**
	6	0.0023	**0.0019**	**0.0073**	**0.0073**	0.0045	**0.0034**
II	**1**	0.0065	**0.0021**	0.0213	**0.0032**	0.0185	**0.0027**
	2	0.0025	**0.0021**	0.0066	**0.0058**	0.0065	**0.0024**
	3	0.0048	**0.0024**	0.0088	**0.0076**	0.0056	**0.0045**
	4	0.0047	**0.0023**	0.0405	**0.0193**	0.0304	**0032**
	5	0.0043	**0.0021**	0.0078	**0.0076**	0.0256	**0.0087**
	6	0.0030	**0.0012**	0.0074	**0.0034**	0.0089	**0.0045**
	7	0.0045	**0.0019**	0.0078	**0.0065**	0.0065	**0.0034**
	8	0.0029	**0.0013**	0.0105	**0.0067**	0.0045	**0.0023**
	9	0.0065	**0.0025**	0.0387	**0.0145**	0.0082	**0.0076**
	10	0.0034	**0.0015**	0.0278	**0.0256**	0.0067	**0.0056**
	11	0.0045	**0.0022**	0.0069	**0.0025**	0.0077	**0.0024**

* Student's t location-scale distribution

where $W_i = \frac{\nu+1}{\nu+S_i}$ and $S_i = (X_i - \mu_t^{(k)})^T \sigma_t^{-1} (X_i - \mu_t^{(k)})$. Using the values of μ and σ, the following equation

$$\sum_{i=1}^{n} \Big(-\psi\Big(\frac{\nu^{(k+1)}}{2}\Big) + \log\Big(\frac{\nu^{(k+1)}}{2}\Big) + 1 + R_i^{(k)} - r_i^{(k)} \Big) = 0 \tag{3.8}$$

with

$$r_i^{(k)} = \psi\big(\frac{\hat{\nu}^{(k)}+1}{2}\big) - log\big(\frac{1}{2}\big(\nu^{(k)} + (x_i - \hat{\mu_t}^{(k)})^T \hat{\sum}^{(k)^{-1}} (x_i - \hat{\mu_t}^{(k)})\big)\big) \tag{3.9}$$

can be solved numerically to estimate the shape parameter. In summary, the EM algorithm for the maximum likelihood estimator is as follows: Given the estimates at iteration, k, the $k + 1^{th}$ iteration of EM consists of the following two steps: expectation (E) step and maximization (M) step. The E step computes the expected value of the log likelihood, given the observed data and current estimate of the model parameters. The M step maximizes the resulting function with respect to the model parameters, yielding new estimates to be used in the next iteration. These steps are repeated until convergence criterion is satisfied. Furthermore, for quantitative analysis, we use the Jensen-Shannon divergence (JSD [76] test to calculate the goodness of fit between the empirical and fitted distributions. The JSD metric, a symmetric form of Kullback-Leibler divergence (KLD), is a measure of similarity between two probability distributions and is defined by

$$D_{KL}(p(x), q(x)) = \sum_{x} p(x) \ln \frac{p(x)}{q(x)} \tag{3.10}$$

$$D_{KL}(q(x), p(x)) = \sum_{x} q(x) \ln \frac{q(x)}{p(x)} \tag{3.11}$$

$$JSD = \frac{D_{KL}(p(x), q(x)) + D_{KL}(q(x), p(x))}{2} \tag{3.12}$$

where $p(x)$ is the PDF of the fitted random variable and $q(x)$ is the PDF of the empirical data. The JSD values corresponding to the student's t location-scale and GG distributions in 3.4 are 0.0030 and 0.0078, respectively. Table 3.1 shows the comparison between GG and student's t location-scale distributions by averaging the values of JSD metric computed over various subbands for different neuroimaging data. It can be seen that the student's t location-scale fits the empirical cumulative distribution function better than the popularly-used GG distribution does.

3.4.2 Fusion Decision Rule

In Section 3.4.1, we have shown the modeling ability of student's t location-scale distribution for the NSST coefficients of neuroimaging data. Therefore, after estimating the parameters, this well approximated distribution is used as a prior PDF to estimate the noise-free coefficients using MAP estimator in which we find $F_i^{k,d}(u,v)$ that maximizes the posterior PDF $p(\frac{F_i(u,v)}{G_i(u,v)})$. Since the approximation coefficients, $F_i^L(u,v)$ deal with low frequency information, i.e., the coarse representation of the image, we follow averaging operation to combine the coefficients of the low frequency subbands of the various images [77]. The detailed coefficients contain high frequency information including important details such as edges and corners present in the images in different directions.

As the coefficients of the high frequency subbands are complex numbers, one of the most commonly used procedures for finding the decision map is the absolute maximum selection rule [45]. According to this rule, when two images are fused, the fused coefficient $F_f^{k,d}(u,v)$ is $F_1^{k,d}(u,v)$ or $F_2^{k,d}(u,v)$ depending on whether $|F_1^{k,d}(u,v)| \geq |F_2^{k,d}(u,v)|$ or $|F_1^{k,d}(u,v) < |F_2^{k,d}(u,v)|$, where $F_1^{k,d}(u,v)$ and $F_2^{k,d}(u,v)$ are the coefficients located at (u,v) in the subband at the k^{th} scale and in the d^{th} direction of the images 1 and 2, respectively. Thus, in the absolute maximum selection rule, the energy of the fused image at the position (u,v) is that of image 1 or image 2 at that position depending on the one that is larger. Thus, using this rule, the energy level of the entire fused image is, in general, larger than that of either of the two images used for fusion. The improvement in the quality of the fused image over that of the individual images can be attributed to the increased level of its energy. In order to further improve the quality of the fused image, one should be investigating an approach that will increase the energy level at its individual pixel position even further. For this purpose, we now propose the following fusion rule.

Let $F_i^{k,d}(u,v) = F_{iR}^{k,d}(u,v) + jF_{iI}^{k,d}(u,v)$ be the transform domain coefficient located at (u,v) in the high frequency subband at the k^{th} scale and in the d^{th} direction of the

image $i = 1, 2, 3, ..., M$ and $F_f^{k,d}(u,v) = F_{fR}^{k,d}(u,v) + jF_{fI}^{k,d}(u,v)$ be the corresponding coefficient at (u,v) in the subband at the k^{th} scale and in the d^{th} direction of the fused image. The fusion rule is defined as

$$F_{fR}^{k,d}(u,v) = \sum_i W_{iR}(u,v) F_{iR}^{k,d}(u,v) \tag{18a}$$

and

$$F_{fI}^{k,d}(u,v) = \sum_i W_{iI}(u,v) F_{iI}^{k,d}(u,v) \tag{18b}$$

where the decision maps, $W_{iR}(u,v)$ and $W_{iI}(u,v)$ are defined by

$$W_{iR}(u,v) = \begin{cases} 1, & i = \mu, \quad \text{if} \quad |F_{\mu R}^{k,d}(u,v)| \geq |F_{1R}^{k,d}(u,v)|, |F_{2R}^{k,d}(u,v)|....|F_{MR}^{k,d}(u,v)| \\ 0, & i \neq \mu \end{cases} \tag{19a}$$

$$W_{iI}(u,v) = \begin{cases} 1, & i = \gamma, \quad \text{if} \quad |F_{\gamma I}^{k,d}(u,v)| \geq |F_{1I}^{k,d}(u,v)|, |F_{2I}^{k,d}(u,v)|....|F_{MI}^{k,d}(u,v)| \\ 0, & i \neq \gamma \end{cases} \tag{19b}$$

It is to be mentioned that if there is more than one value of i that satisfies $i = \mu$ (or $i = \gamma$), then we choose any one of the i values as μ (or γ) and let the corresponding weight to '1', and the rest of the weights to '0'.

Thus, in the proposed fusion rule, the real and imaginary parts of the fused coefficient $F_f^{k,d}(u,v)$ in a given high frequency subband are obtained separately, as opposed to that in the absolute maximum selection rule. The real (imaginary) part of the fused coefficient $F_{fR}^{k,d}(u,v)\left(F_{fI}^{k,d}(u,v)\right)$ is chosen to be the one for which its absolute value is the largest amongst the absolute values of the real (imaginary) parts of $F_i^{k,d}(u,v)$ of the various images involved in the fusion.

It can easily be seen that, in general,

$$|F_f^{k,d}(u,v)| > |F_i^{k,d}(u,v)| \tag{20}$$

for all values of i. Thus, the total energy in the high frequency subbands using the proposed fusion rule will always be greater than that obtained using absolute maximum selection rule. We will refer to the proposed fusion rule as the energy maximizing fusion rule.

3.4.3 Implementation Steps

In this subsection, we summarize the main steps involved in the proposed medical image fusion algorithm.

Step 1: Apply the NSST transform on the individual source image.

Step 2: Estimate the parameters μ_t, σ_t, ν from the NSST coefficients of each input image using the method described in Section 2.4.1.

Step 3: Apply the fusion rule on the NSST coefficients as explained in Section 2.4.2.

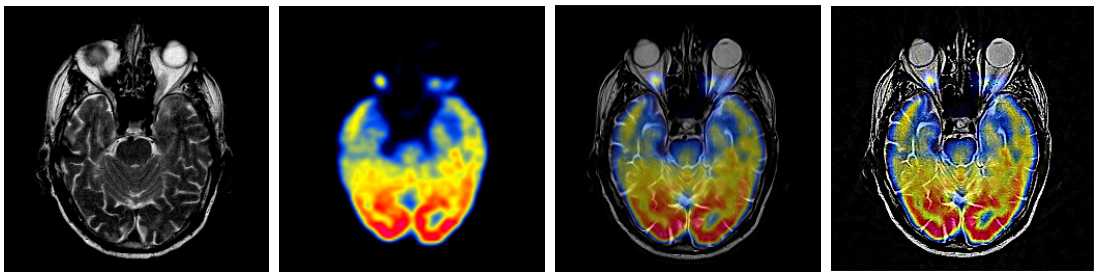
Step 4: Apply the inverse NSST to the fused coefficients obtained in Step 3.

3.5 Experimental Results

In this section, we study the performance of the proposed technique on different pairs of 2-D multimodal synthetic and real brain images. Also, we compare the performance of the proposed method with that of five other methods, namely, curvelet transform-based (CTB) method [30], contourlet transform-based (CB) method [29], guided image filter-based (GIFB) method [16], local Laplacian filtering domain-based (LLDB) method [78],

and parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform-based (PA-PCNN-NSSTB) method [14]. For a fair comparison, we choose a number of pairs of images from various databases (BrainWeb [79], Alzheimer's disease neuroimaging initiative (ADNI) [80], and Whole Brain Atlas (WBA) [81]), and obtain the fusion results for each of the methods mentioned above as well as for the proposed method. For the methods of [16], [78], we have used the codes published by the authors to obtain the fusion results, whereas for the other methods [14], [29], [30], we have written codes in Matlab. For the purpose of obtaining the experimental results, we select two scales for the contourlet and non-subsampled shearlet transforms to decompose the source images, where for the first scale six directions and for the second scale ten directions are chosen. Thus, in total, we have 16 high frequency subbands and one low frequency subband.

We first investigate the energy values obtained in each high frequency subband when the absolute maximum selection rule and the proposed energy maximizing fusion rule are applied. For this purpose, we consider Figs. 3.6(a) and 3.6(b) taken from the WBA database. It is found that the energy level of each subband is increased using the proposed fusion rule, with a total increase of 7.15% over that using the absolute maximum selection rule. We now compare the performance of our method that involves student's t location scale modeling using the proposed fusion rule with that using the conventional absolute maximum



(a) (b) (c) (d)

Figure 3.6: Fusion results using the conventional and proposed fusion rules. (a) and (b) are the pair of source images being fused. (c) and (d) are the fused images using conventional and proposed fusion rules, respectively.

selection rule. The objective quality of the fused images, using the two rules, as measured by the various metrics, is summarized in Table 3.2. It is seen from this table that there is an improvement in the values of every one of the metrics indicating the performance using the proposed fusion rule is superior to that using the absolute maximum selection rule. This improved performance is further reinforced by the fused images shown in Figs. 3.6(c) and 3.6(d) using the absolute maximum selection and proposed fusion rules, respectively. This improvement is clearly due to the use of the proposed fusion rule, which retains important high frequency details such as edges, corners, blobs and ridges better than the conventional absolute maximum selection rule does. Hence, in all our subsequent experiments, fusion is carried out using the novel energy maximizing fusion rule.

Table 3.2: Comparison of Conventional and Proposed Fusion Rules in Terms of Objective Fusion Measures

Source Image	**Fusion Rule**	**PSNR**	**SSIM**	**MI**	**Entropy**	$\mathbf{Q}^{\frac{f_1 f_2}{F}}$
Figs. 3.6(a) and 3.6(b)	**Absolute Maximum Selection**	16.0222	0.5946	1.5009	4.2751	0.7327
	Proposed	**17.1243**	**0.6334**	**1.5705**	**5.7234**	**0.8507**

3.5.1 Experimental Results Using Simulated Data

To assess the performance of our proposed method, we carry out extensive experiments on pairs of simulated brain MR images taken from the BrainWeb database. For this purpose, we consider slices along the axial plane from the longitudinal relaxation (T1)-weighted, transverse relaxation (T2)-weighted, and proton density (PD)-weighted MR brain volume with various slice thicknesses, intensity non-uniformity levels, and noise. Fig. 3.7 illustrates the results of fusion of T1-weighted MR (MR-T1) and PD-weighted MR (MR-PD) images of a normal brain. The objective quality of the fused images obtained for the various methods as measured by the different metrics is summarized in Table 3.3. It is seen from

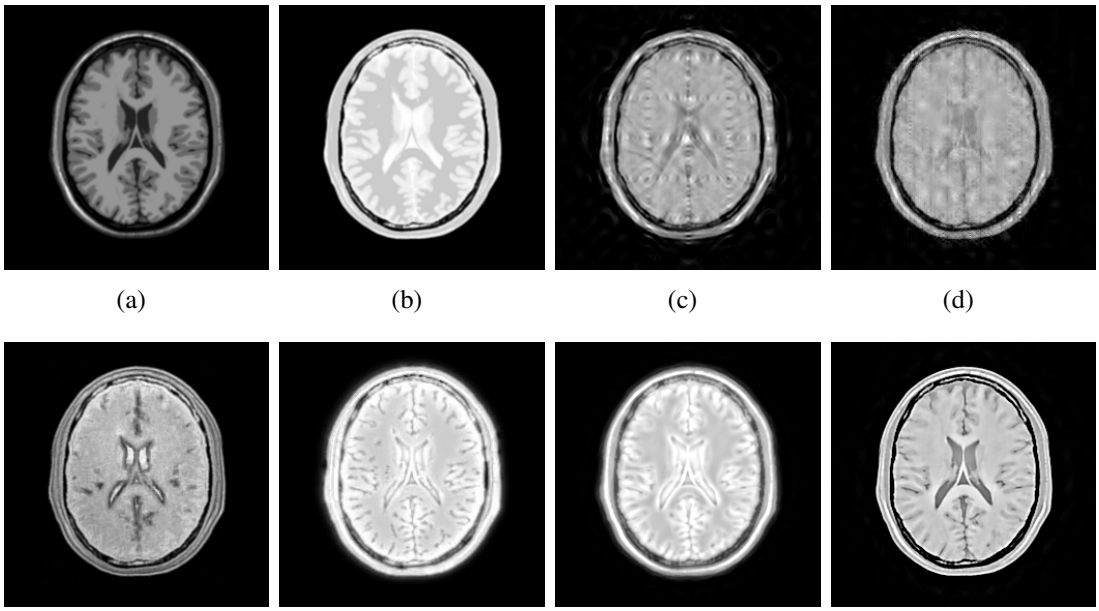
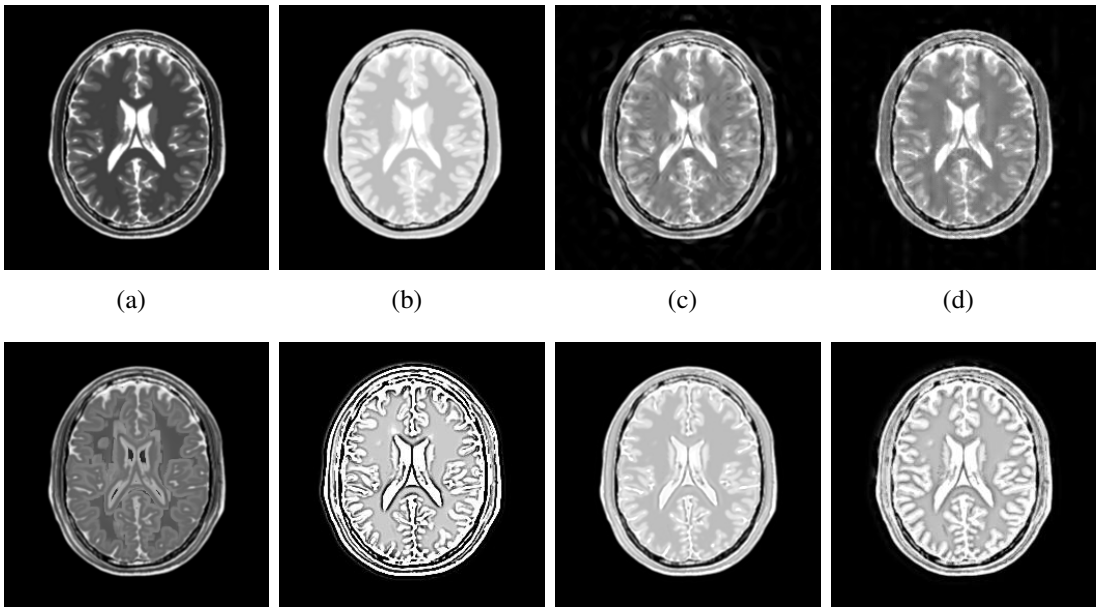


Figure 3.7: Multimodal fusion results of 2-D simulated neuroimages of a normal brain taken from the BrainWeb database. (a) and (b) correspond to MR-T1 and MR-PD images of a normal brain, respectively. (c), (d), (e), (f), (g), and (h) correspond to the fused images using curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively.

this table that the proposed method yields the highest values in all the measures considered. This is further reinforced by observing the visual quality of the fused images shown in Figs. 3.7 (c)–(h) using the curvelet-based, contourlet-based, guided image filter-based, local Laplacian filtering domain-based, parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform-based, and proposed methods, respectively. Fig. 3.8 illustrates the results of fusion of T2-weighted MR (MR-T2) and PD-weighted MR (MR-PD) images of a multiple sclerosis brain. The objective quality of the fused images obtained for the various methods as measured by the metrics is also summarized in Table 3.3. It is also seen from this table that the proposed method yields the highest values in all the measures considered. This is further reinforced by observing the visual quality of the fused images shown in Figs. 3.8 (c)–(h) using the curvelet-based, contourlet-based, guided



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.8: Multimodal fusion results of 2-D simulated neuroimages of a multiple sclerosis lesion brain taken from the BrainWeb database. (a) and (b) correspond to MR-T2 and MR-PD images of a multiple sclerosis lesion brain, respectively. (c), (d), (e), (f), (g), and (h) correspond to the fused images using curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively.

Table 3.3: Objective Fusion Results of 2D Neuroimages in Fig. 3.7 and Fig. 3.8

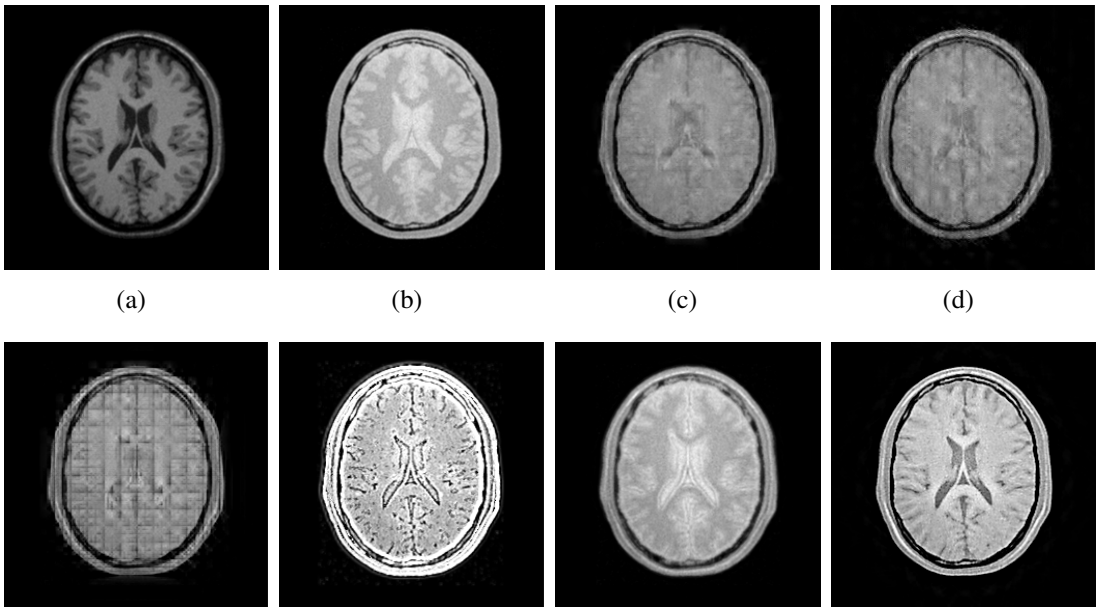
Source Images	**Method**	**PSNR**	**SSIM**	**MI**	**Entropy**	**Q$^{f_1 f_2}_{F}$**
Fig. 3.7 (a) and (b)	**CTB [30]**	16.0849	0.552	2.2389	5.398	0.683
	CB [29]	16.2785	0.6292	2.2553	5.0648	0.7204
	GIFB [16]	15.8754	0.6735	2.2145	4.4678	0.6689
	LLDB [78]	16.9352	0.7346	2.3267	5.3237	0.7864
	PA_PCN_NNSSTB [14]	17.0287	0.7425	2.2261	5.8458	0.8009
	Proposed	**17.2185**	**0.7948**	**2.5338**	**6.5722**	**0.8397**
Fig. 3.8 (a) and (b)	**CTB [30]**	16.9616	0.728	2.653	5.4947	0.8869
	CB [29]	17.1494	0.8042	2.7742	5.1992	0.9023
	GIFB [16]	16.9869	0.8465	2.6167	4.8734	0.8367
	LLDB [78]	17.2165	0.8498	2.6854	5.5841	0.8592
	PA_PCN_NNSSTB [14]	17.3674	0.8357	2.3268	5.5845	0.8612
	Proposed	**17.5287**	**0.8525**	**2.8461**	**5.7358**	**0.9309**

image filter-based, local Laplacian filtering domain-based, parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform-based, and proposed methods, respectively. From these two figures, it is clear that the proposed method is more efficient in transferring the structural details from the MR images of T1/T2 and PD weighting techniques compared to that obtained using other traditional methods.

In addition, we now consider the same two MR images shown in Figs. 3.7 (a) and (b), but with 40% intensity non-uniformity level and containing 1% or 5% or 9% noise. The noise in the simulated scans has Rayleigh statistics in the background and Rician statistics in the signal regions [79]. The fusion results as measured by the various metrics is summarized in Table 3.4. It is seen from this table that the proposed method has the highest values in all of the five performance metrics considered. This is further reinforced by the subjective fusion results shown in Fig. 3.9, when the noise is 5%. It is observed from the

Table 3.4: Objective Fusion Results of MR images with Noise

Noisy Data	Method	PSNR	SSIM	MI	Entropy	$Q^{\frac{f_1 f_2}{F}}$
40% intensity non-uniformity level and 9% noise	**CTB[30]**	20.1979	0.7005	2.4777	5.4376	0.7285
	CB [29]	20.3785	0.7489	2.5907	5.2155	0.7521
	GIFB [16]	19.9827	0.7139	2.5125	5.1003	0.7067
	LLDB [78]	19.8325	0.7540	2.6322	5.4289	0.7932
	PA-PCNN-NSSTB [14]	19.3576	0.7653	2.5632	5.4162	0.7843
	Proposed	**20.4189**	**0.781**	**2.6795**	**5.5379**	**0.819**
40% intensity non-uniformity level and 5% noise	**CTB [30]**	17.9937	0.6042	2.1455	5.254	0.6672
	CB[29]	18.1873	0.674	2.1728	4.9356	0.7009
	GIFB [16]	17.7914	0.6489	2.0689	4.7135	0.6589
	LLDB [78]	17.9842	0.7043	2.284	5.2165	0.7621
	PA-PCNN-NSSTB [14]	18.654	0.7216	2.3412	5.1685	0.7634
	Proposed	**18.733**	**0.7315**	**2.5779**	**5.4988**	**0.8157**
40% intensity non-uniformity level and 1% noise	**CTB [30]**	16.672	0.5713	2.193	5.3405	0.675
	CB [29]	16.8476	0.6473	2.2212	5.0154	0.711
	GIFB [16]	16.4853	0.6279	2.1098	4.7365	0.6664
	LLDB [78]	16.5731	0.687	2.2491	5.3727	0.7210
	PA-PCNN-NSSTB [14]	17.7462	0.6932	2.1560	5.1356	0.8127
	Proposed	**17.7885**	**0.7076**	**2.3567**	**5.5424**	**0.8396**



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.9: Multimodal fusion results of 2-D simulated neuroimages with 40% intensity non-uniformity level and 5% noise taken from the BrainWeb database. (a) MR-T1 image. (b) MR-PD image. (c), (d), (e), (f), (g), and (h) correspond to the fused images using curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively.

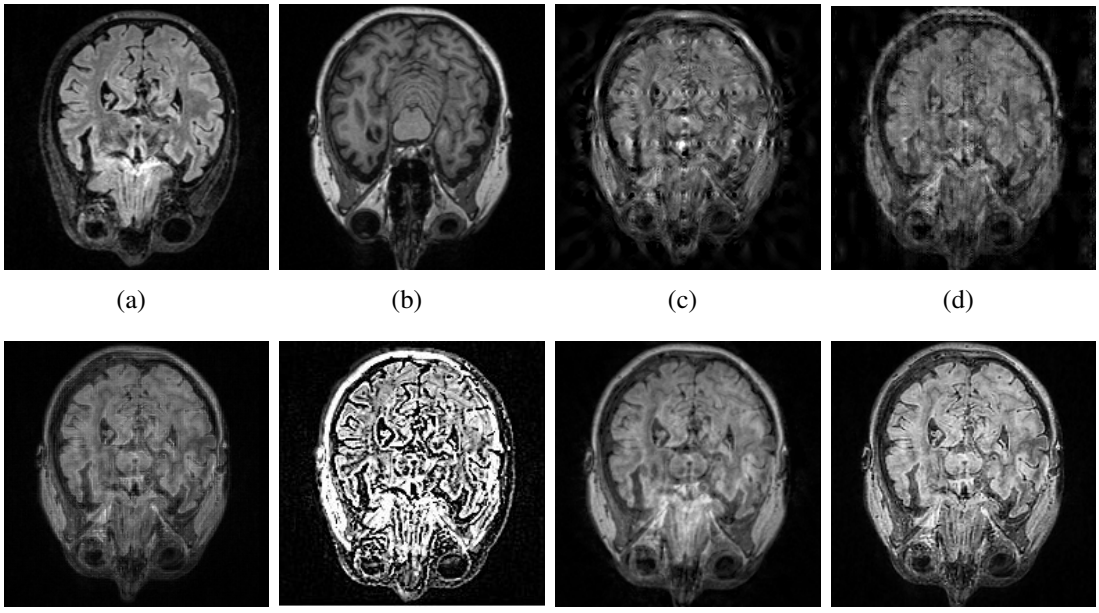
fused images that even in the noisy case, unlike the other traditional methods, the proposed method could retain most of the structural information of the source images in the fused image, and provide a better image quality. Thus, the proposed method is more robust than the other traditional methods.

3.5.2 Experimental Results Using Real Data

In this subsection, we carry out extensive experiments on the fusion of real neuroimaging data taken from Alzheimer's Disease Neuroimaging Initiative (ADNI) [80] and Whole Brain Atlas (WBA) [81] databases.

Fusion of Two MR Images taken from ADNI Database

We now consider the fusion of a pair of real MR images taken from the Alzheimer's disease neuroimaging initiative (ADNI) database. The images in 3.10 (a) and (b) are the slices along the axial plane from the MR-T1 and MR-PD volumes, respectively. Table 3.5 lists the objective fusion results for various methods. It is observed from this table that the proposed method outperforms all the other methods in terms of all the measures considered. In particular, it has a significantly higher entropy as well as a normalized weighted performance metric, $Q^{\frac{f_1 f_2}{F}}$. This is further reinforced by observing the visual quality of the fused images shown in Figs. 3.10 (c)–(h) using the curvelet-based, contourlet-based, guided image filter-based, local Laplacian filtering domain-based, parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform-based, and proposed methods,



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.10: Multimodal fusion results of 2-D real neuroimages of an Alzheimer's disease patient from the ADNI database. (a) MR-T2 image. (b) MR-PD image. (c), (d), (e), (f), and (g) correspond to the fused images using curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively.

Table 3.5: Objective Fusion Results of 2D Neuroimages in Fig. 3.10

Source Images	Method	PSNR	SSIM	MI	Entropy	$Q^{\frac{f_1 f_2}{F}}$
Fig. 3.10 (a) and (b)	**CTB [30]**	20.2011	0.7022	1.4183	5.2607	0.5756
	CB [29]	21.2807	0.8776	2.5294	4.9487	0.5842
	GIFB [16]	21.1421	0.8456	2.3876	4.6322	0.7920
	LLDB [78]	21.4567	0.8823	2.4762	5.9876	0.6237
	PA-PCNN-NSSTB [14]	21.4938	0.8923	2.3921	5.6854	0.7932
	Proposed	**21.5599**	**0.8977**	**2.8484**	**6.9367**	**0.8567**

respectively. It is seen from these fused images that the proposed technique preserves the structural details from both the T2 and PD weighting techniques better than the other traditional methods do.

Fusion of Two MR Images taken from WBA Database

We now consider the fusion of all possible pairs of MR-T1, MR-T2, and MR-PD images of a lyme encephalopathy patient and a mild Alzheimer's disease patient taken from the Whole Brain Atlas (WBA) database. Table 3.6 lists the objective performance measures for the three pairs of 2-D MR images of the lyme encephalopathy patient, while Table 3.7 lists the corresponding results for the mild Alzheimer's disease patient. It is seen from these tables that the proposed technique outperforms all the other methods in terms of all the measures considered. This is further reinforced by the subjective fusion results shown in Fig. 3.11 for the lyme encephalopathy patient and the corresponding results shown in Fig. 3.12 for the mild Alzheimer's disease patient. It is seen from these fused images that the proposed technique preserves the structural details of the source images for all the three pairs better than the other traditional methods do.

Fusion of an MR image with SPECT or PET image taken from WBA Database

We now consider the fusion of four pairs of multimodal brain images taken from the Whole Brain Atlas (WBA) database. The various pairs of images are shown in Fig. 3.13.

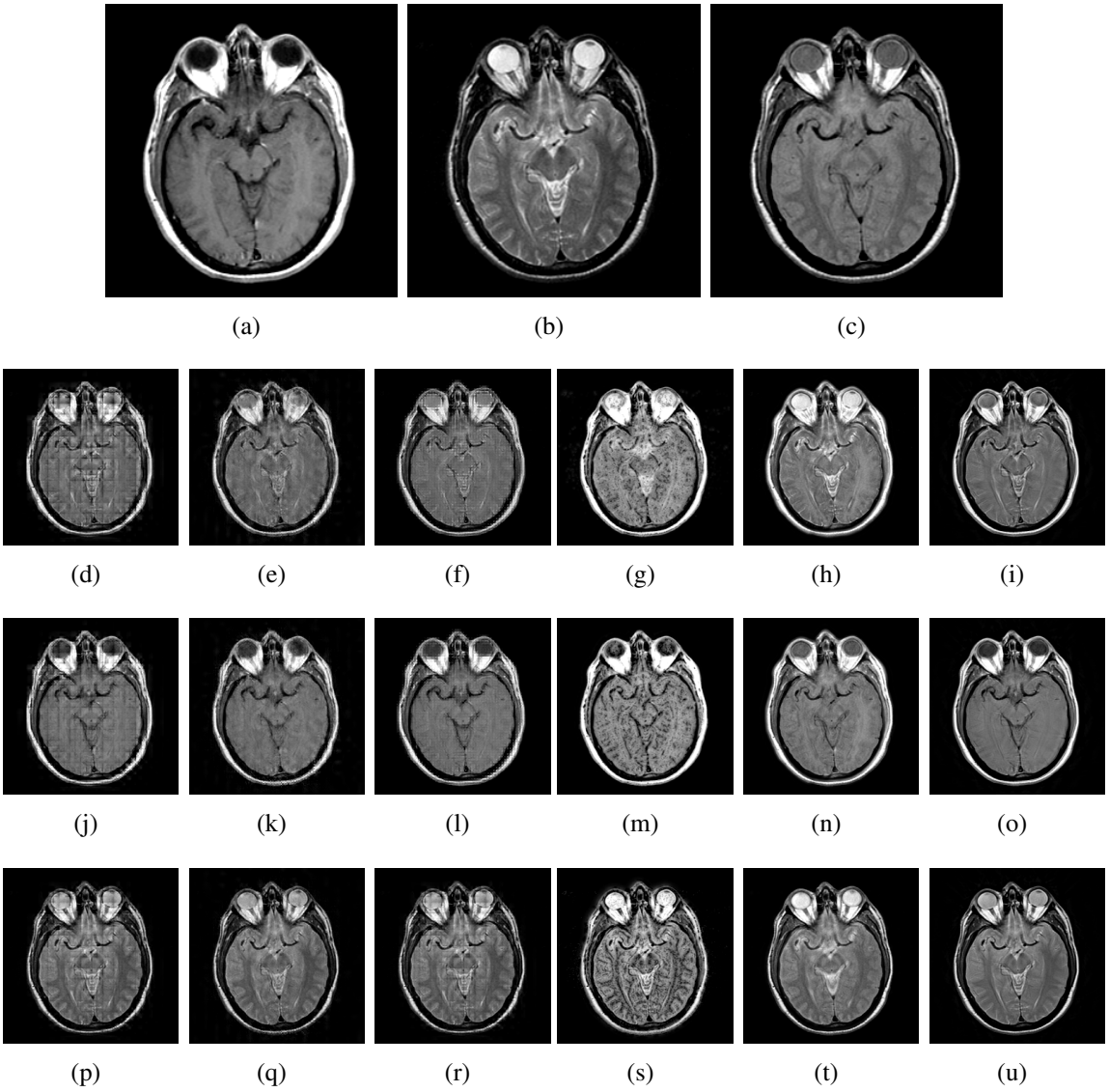


Figure 3.11: Multimodal fusion results of 2-D real neuroimages of a lyme encephalopathy patient from the Whole Brain Atlas database. (a) MRI-T1 image. (b) MR-T2 image. (c) MR-PD image. Rows 2 to 4 correspond to the fusion images of the input combinations (a) and (b), (a) and (c), and (b) and (c), respectively. Here, columns 1 to 6 correspond to the fused images of curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively.

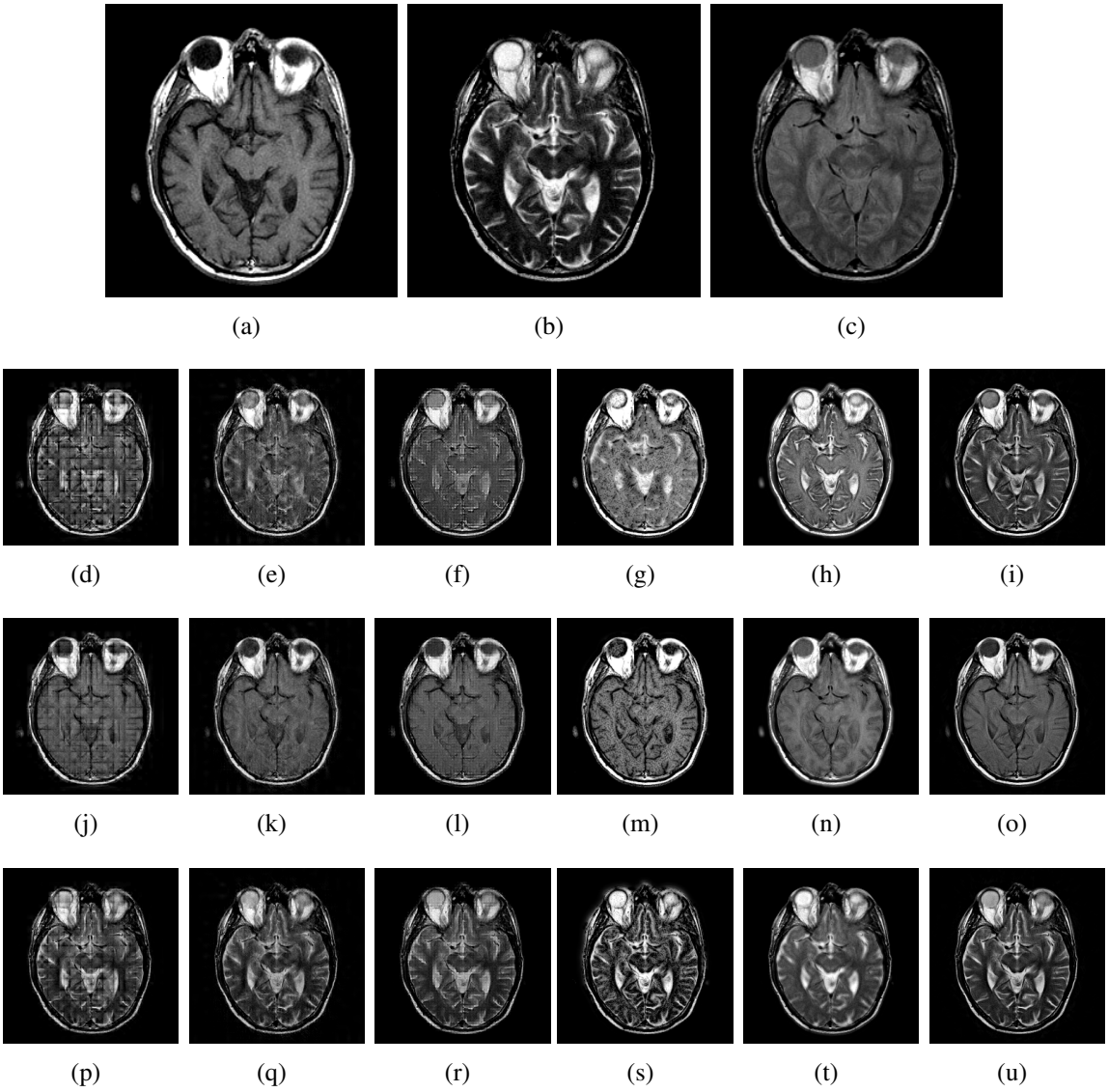


Figure 3.12: Multimodal fusion results of 2-D real neuroimages of a mild Alzheimer's disease patient from the Whole Brain Atlas database. (a) MR-T1 image. (b) MR-T2 image.(c) MR-PD image. Rows 2 to 4 correspond to the fusion images of the input combinations (a) and (b), (a) and (c), and (b) and (c), respectively. Here, columns 1 to 6 correspond to the fused images of curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively.

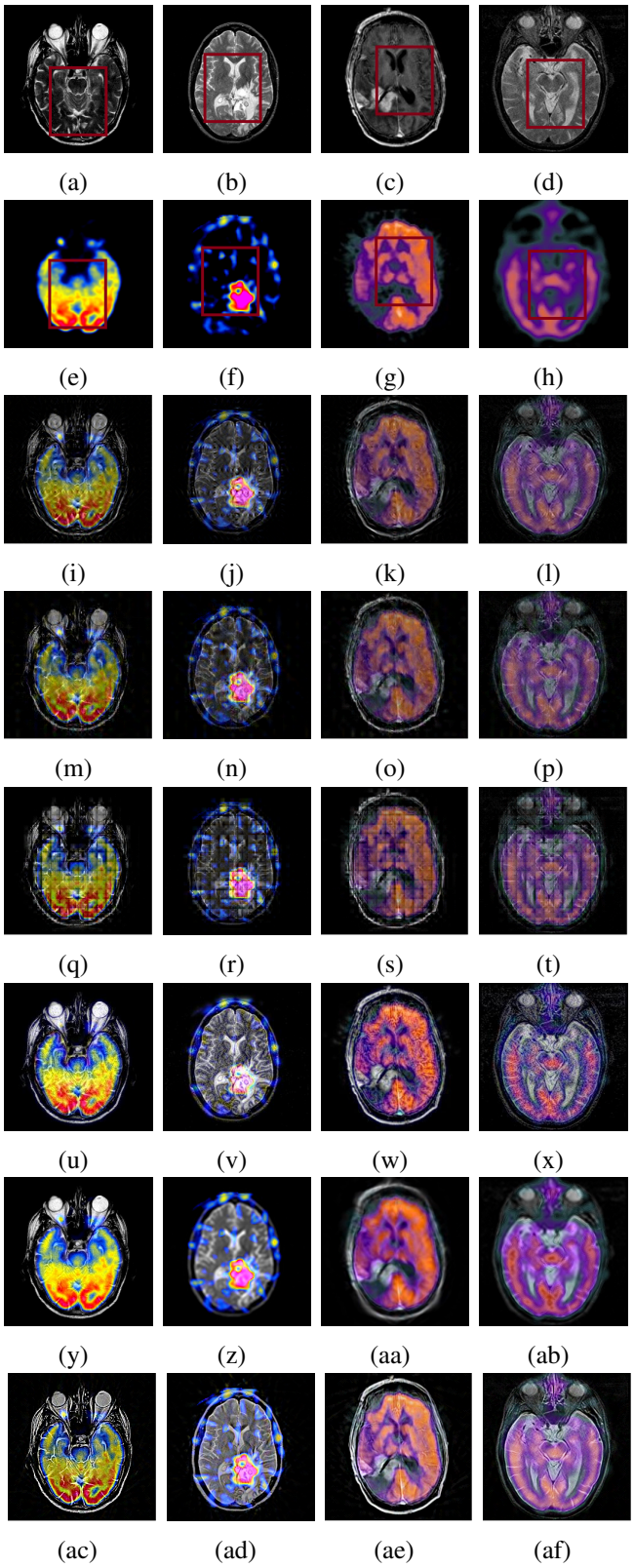


Figure 3.13: Multimodal fusion results of 2-D real neuroimages from the Whole Brain Atlas database. (a) and (e), (b) and (f), (c) and (g), and (d) and (h) are the pairs of source images being fused. Rows 3 to 8 correspond to the fusion results of curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively.

Table 3.6: Objective Fusion Results of 2D Neuroimages in Fig. 3.11

Source Images	Method	PSNR	SSIM	MI	Entropy	$Q^{f_1 f_2}_F$
Fig. 3.11 (a) and (b)	**CTB [30]**	15.1990	0.6660	1.8584	4.6795	0.5841
	CB [29]	16.2012	0.5455	1.6728	5.2058	0.6289
	GIFB [16]	16.2961	0.6582	1.7470	5.2219	0.6678
	LLDB [78]	14.1159	0.5800	1.5545	4.8573	0.6579
	PA_PCN_NNSSTB [14]	16.8567	0.6371	2.0316	5.1834	0.6531
	Proposed	**19.7274**	**0.7031**	**2.1381**	**5.2479**	**0.7299**
Fig. 3.11 (a) and (c)	**CTB [30]**	17.8508	0.6719	1.8183	4.6789	0.5748
	CB [29]	17.9119	0.5912	1.6636	5.5128	0.6558
	GIFB [16]	18.0121	0.6812	1.7471	4.9856	0.6943
	LLDB [78]	15.1264	0.5530	1.3744	3.8218	0.5250
	PA-PCNN-NSSTB [14]	16.2853	0.6398	1.9384	5.0508	0.5013
	Proposed	**21.6344**	**0.7927**	**2.2404**	**5.1395**	**0.7386**
Fig. 3.11 (b) and (c)	**CTB [30]**	17.9729	0.6822	1.8206	4.5653	0.5678
	CB [29]	18.0572	0.5868	1.7059	5.5985	0.6545
	GIFB [16]	18.1884	0.6830	1.7776	5.0653	0.6798
	LLDB [78]	15.5506	0.5732	1.4375	4.9315	0.6059
	PA-PCNN-NSSTB [14]	16.7662	0.6427	1.9194	5.0926	0.5331
	Proposed	**20.8525**	**0.8508**	**2.4986**	**5.1180**	**0.7204**

Table 3.7: Objective Fusion Results of 2D Neuroimages in Fig. 3.12

Source Images	Method	PSNR	SSIM	MI	Entropy	$Q^{f_1 f_2}_F$
Fig. 3.12 (a) and (b)	**CTB [30]**	17.4250	0.6996	1.9849	4.9650	0.7186
	CB [29]	16.4331	0.5997	1.8998	4.7807	0.7484
	GIFB [16]	16.6803	0.7106	2.0190	4.2428	0.7681
	LLDB [78]	15.6440	0.4904	1.5521	4.5880	0.5905
	PA_PCN_NNSSTB [14]	16.8253	0.7061	1.9313	4.9968	0.7577
	Proposed	**18.5626**	**0.8047**	**2.0743**	**5.3081**	**0.8357**
Fig. 3.12 (a) and (c)	**CTB [30]**	15.5317	0.8137	2.4057	4.8595	0.8507
	CB [29]	15.7643	0.7484	2.3738	5.0257	0.8652
	GIFB [16]	14.1377	0.8354	2.4763	5.0616	0.8741
	LLDB [78]	15.6020	0.5483	1.6156	4.4413	0.6012
	PA-PCNN-NSSTB [14]	17.8390	0.8093	2.5994	5.3449	0.7612
	Proposed	**18.2550**	**0.8929**	**2.6848**	**5.6313**	**0.8017**
Fig. 3.12 (b) and (c)	**CTB [30]**	16.8453	0.8324	2.4018	4.9674	0.8467
	CB [29]	16.8374	0.7725	2.2875	5.1198	0.8808
	GIFB [16]	17.0589	0.8417	2.3992	5.1825	0.8905
	LLDB [78]	15.3579	0.5174	1.4647	4.6997	0.5197
	PA-PCNN-NSSTB [14]	17.3163	0.8191	2.5185	5.0034	0.8191
	Proposed	**18.9036**	**0.8500**	**2.6662**	**5.2467**	**0.8155**

The pair of images in Figs. 3.13 (a) and (e) correspond to T2-weighted magnetic resonance imaging (MR-T2) and positron emission tomography using F-18-fluorodeoxy-glucose (FDG-PET) from a 70 year-old man with mild Alzheimer's disease, respectively. The

Table 3.8: Objective Fusion Results of 2D Neuroimages in Fig. 3.13

Source Images	Method	PSNR	SSIM	MI	Entropy	$Q^{\frac{f_1 f_2}{F}}$
Fig. 3.13 (a) and (e)	**CTB [30]**	15.6906	0.544	1.4931	5.1231	0.5949
	CB [29]	15.6585	0.5358	1.4557	5.1337	0.6118
	GIFB [16]	15.3278	0.534	1.465	5.5401	0.5934
	LLDB [78]	15.9276	0.5281	1.4972	5.6984	0.6038
	PA_PCN_NNSSTB [14]	16.7136	0.5982	1.4987	5.5872	0.7912
	Proposed	**16.7246**	**0.6334**	**1.5705**	**5.7234**	**0.8507**
Fig. 3.13 (b) and (f)	**CTB [30]**	16.9026	0.5997	1.3604	4.8770	0.6077
	CB [29]	16.8106	0.5795	1.2915	5.0118	0.6442
	GIFB [16]	16.1564	0.5876	1.1332	5.3483	0.6745
	LLDB[78]	16.9254	0.6237	1.392	5.4872	0.7238
	PA-PCNN-NSSTB[14]	16.3749	0.6357	1.2968	5.4845	0.7612
	Proposed	**16.9308**	**0.6573**	**1.4086**	**5.6932**	**0.8525**
Fig. 3.13 (c) and (g)	**CTB [30]**	19.1321	0.6348	1.9821	5.4808	0.6033
	CB [29]	19.1351	0.6233	1.9478	5.543	0.6372
	GIFB [16]	18.3678	0.6342	1.7654	6.1425	0.6859
	LLDB [78]	19.5923	0.6821	1.9943	6.1893	0.7815
	PA-PCNN-NSSTB [14]	20.0674	0.6324	1.9269	5.6956	0.7874
	Proposed	**20.2543**	**0.6926**	**2.0739**	**6.2802**	**0.8351**
Fig. 3.13 (d) and (h)	**CTB [30]**	20.326	0.5904	2.1641	6.1583	0.6251
	CB [29]	20.2649	0.5834	2.1123	6.1593	0.6302
	GIFB [16]	20.1453	0.5673	2.1432	6.5690	0.6003
	LLDB [78]	20.9246	0.56958	2.2871	6.9326	0.7152
	PA-PCNN-NSSTB [14]	20.9434	0.6053	2.1468	5.9865	0.7362
	Proposed	**21.5452**	**0.6547**	**2.3827**	**7.0571**	**0.8096**

pair of images in Figs. 3.13 (b) and (f) correspond to MR-T2 and single photon emission computed tomography with Thallium-201 (SPECT-T1) from a 51 year old woman with Anaplastic Astrocytoma, a type of rare malignant tumor, respectively. The pair of images in Figs. 3.13 (c) and (g) are that of the T1-weighted MR (MR-T1), after gadolinium-diethylenetriamine pentaacetic acid (Gd-DTPA)and FDG-PET from a 53 year old man with Astrocytoma, a type of cancer of the brain, respectively. Finally, the pair of images in Figs. 3.13 (d) and (h) are MR-T2 and SPECT with perfusion agent Tc99m-HM-PAO (SPECT-Tc) from a 76 year old man with Subdural Hygroma, accumumulation of cerebrospinal fluid in the subdural membrane. The objective performance measures obtained

for each of the four pairs of 2-D neuroimages using various fusion methods are summarized in Table 3.8. It is seen from this table that the proposed method consistently outperforms all the other methods in terms of all the objective measures considered and in particular, it has significantly higher values of the normalized weighted performance metric, $Q^{\frac{f_1 f_2}{F}}$, signifying that the proposed fusion algorithm conveys the gradient information without any significant loss. The fused images for each of the four pairs using the curvelet-based method are shown in Figs. 3.13 (i)-(l), the corresponding fused images using the contourlet-based method are shown in Figs. 3.13 (m)-(p), those using the guided image filter-based are shown in Figs. 3.13 (q)-(t), those using the local Laplacian filtering domain-based are shown in Figs. 3.13 (u)-(x), and those using the parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform-based are shown in Figs. 3.13 (y)-(ab). The fused images using the proposed method for each of the four pairs are shown in Figs. 3.13 (ac)-(af). In order to appreciate the effectiveness of the proposed method over that of the other methods in preserving both the structural and functional information of the two modalities of each pair of images, a small segment (red box as shown in Figs. 3.13 (a)-(h)) of each pair of the original images and the corresponding segments of the fused images obtained by various methods are all zoomed in and the resulting zoomed segments are shown in Fig. 3.14. It is clearly seen from Fig. 3.14 that the proposed technique preserves the structural features such as the edge information in MR and the functional features such as the color information in both PET and SPECT images better than the other existing techniques do. This is in view of the fact the NSST used in our approach provides optimal representation of neuroimaging data with edges and texture, and further, the student's t location-scale distribution, which acts a prior distribution, closely fits the empirical NSST coefficients of source images; more importantly, it is due to the proposed fusion rule, which increases the overall high frequency subband energy. Also, it is noticed that the guided image filter scheme produces blocking artifacts reflecting the lack of sufficient directional information.

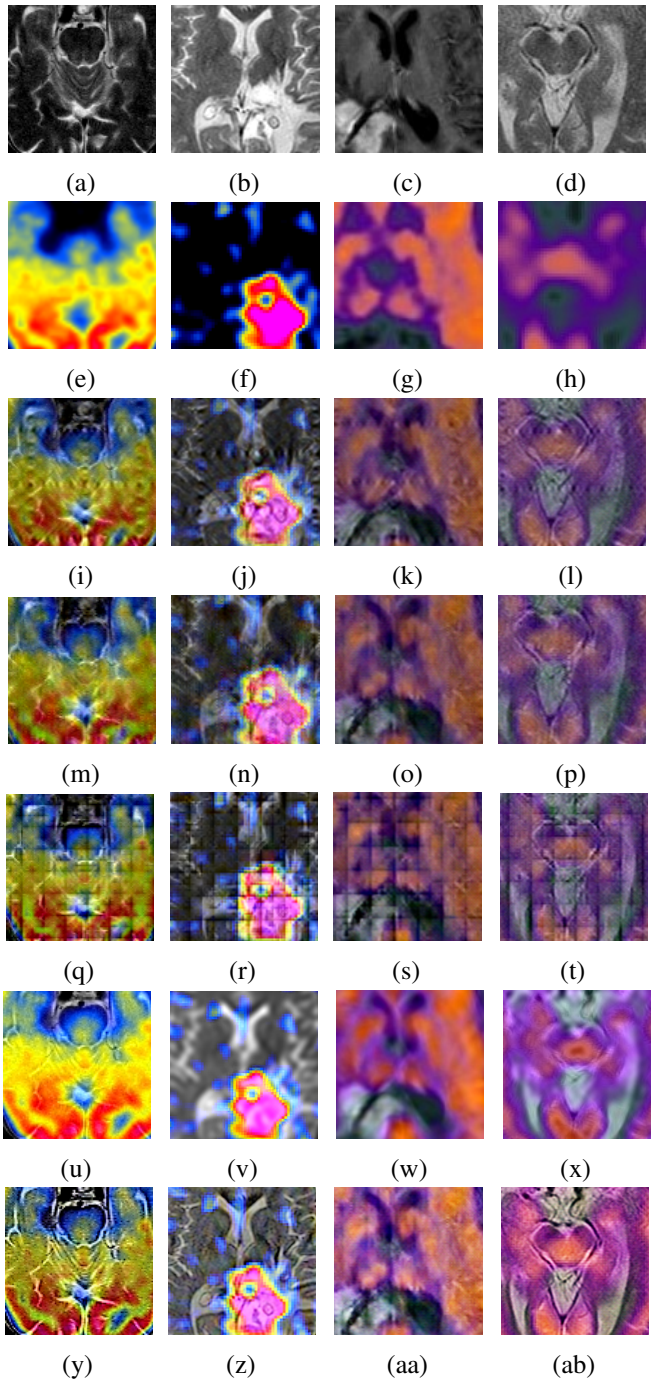


Figure 3.14: Multimodal fusion results of the zoom region [red box as shown in Figs. 3.13 (a) - (h)] of all the pairs of neuroimages in Fig. 3.11. The four images in each row correspond to the enlarged area of the images in Figs. 3.13.

Although the local Laplacian filtering scheme provides good contrast and no blocking artifacts, the functional information is not well preserved as observed from the color details in Fig. 3.14 (v) and some noise is introduced as seen from Fig. 3.14 (u). Also, it is noticed

from the fused image, shown in Fig. 3.14 (ad), that substantial structural information is lost in lieu of functional superimposition. However, only a subset of the large region of a mixed signal on MR-T2 corresponds to the active tumor which is difficult to localize. The thallium uptake of tumor region in SPECT has been shown to increase the sensitivity and specificity of tumor detection over MRI. Hence, fusing these two images provides clinicians with a single image that combines higher sensitivity of the color SPECT image with the improved localization of the suspected details in the MRI image.

Fig. 3.15 shows the bar graph of the average normalized weighted performance metric, $Q^{\frac{f_1 f_2}{F}}$ obtained using the proposed and other traditional methods for the multimodal pairs, namely, MR-T1/SPECT, MR-T2/SPECT, MR-PD/SPECT, MR-T1/PET, MR-T2/PET, and MR-PD/PET. It is seen from this figure that the proposed technique outperforms the other traditional methods in terms of the normalized weighted performance metric, $Q^{\frac{f_1 f_2}{F}}$ for all the multimodal pairs considered. Ideally, the $Q^{\frac{f_1 f_2}{F}}$ values of 0 and 1 indicate the loss of and complete recovery of edge information, respectively.

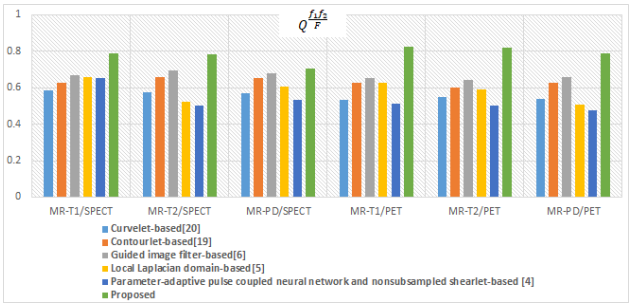


Figure 3.15: Objective multimodal fusion results comparison in terms of $Q^{\frac{f_1 f_2}{F}}$.

We next consider the fusion of an MR image and SPECT image of a lyme encephalopathy patient taken from the WBA database.

Fusion of an MR image with the SPECT image of a lyme encephalopathy patient

We now consider fusing each of the MR images of a lyme encephalopathy patient shown in Figs. 3.16 (a), (b), and (c) with the functional SPECT image shown in Figs. 3.16 (d).

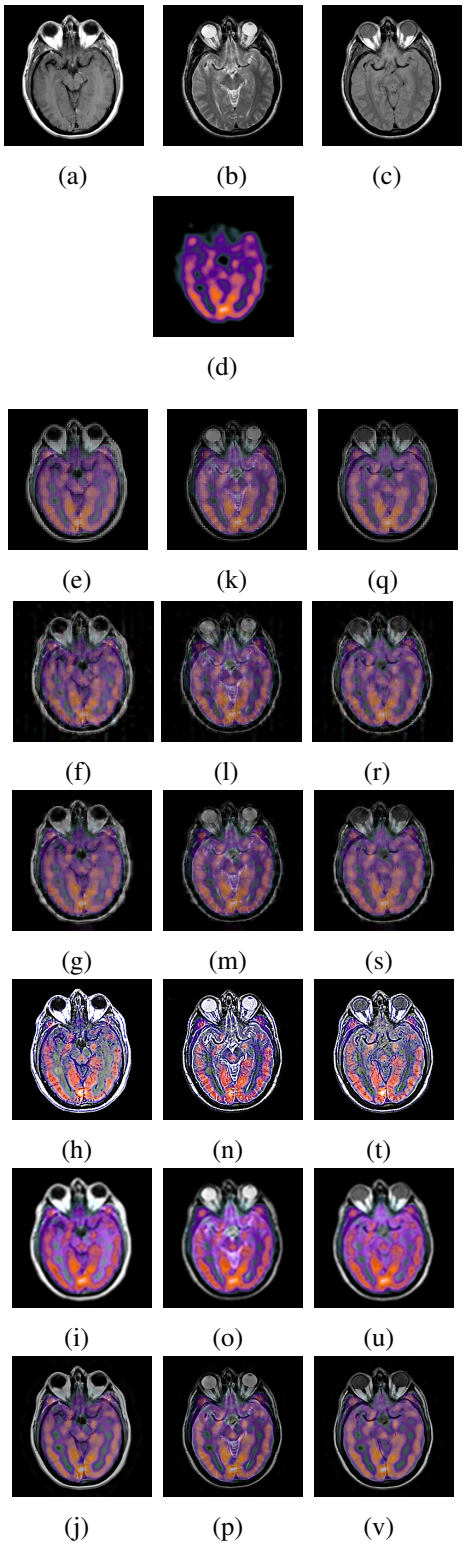


Figure 3.16: Multimodal fusion results of an MR image with the SPECT image of a lyme encephalopathy patient. (a) MR-T1 image (b) MR-T2 image (c) MR-PD image, and (d) SPECT image. (a) and (d), (b) and (d), and (c) and (d) are the pairs of source images being fused. (e)-(j), (k)-(p), and (q)-(v) are the fused images of (a) and (d), (b) and (d), and (c) and (d), respectively. Rows 3 to 8 correspond to the fused images obtained using curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively. NOTE: Figs. 3.16 (a), (b), and (c) are, respectively, the same as Figs. 3.11 (a), (b), and (c), and are repeated here for convenience.

Table 3.9: Objective Fusion Results of 2-D Neuroimages in Fig. 3.16

Source Images	Method	PSNR	SSIM	MI	Entropy	$Q^{f_1 f_2}_F$
Figs. 3.16 (a) and (d)	**CTB [30]**	15.8341	0.5929	2.3998	5.7207	0.6484
	CB [29]	15.5331	0.5947	2.8998	5.7817	0.6484
	GIFB [16]	16.6803	0.5136	2.0190	5.2438	0.6681
	LLDB[78]	16.5470	0.5904	2.5521	5.5840	0.7905
	PA_PCN_NNSSTB[14]	15.8253	0.6261	**2.9313**	6.9968	0.6977
	Proposed	**20.5426**	**0.6947**	2.8743	**7.3781**	**0.8357**
Figs. 3.16 (b) and (d)	**CTB[30]**	17.5348	0.6359	1.8456	4.6535	0.5968
	CB [29]	16.5120	0.5472	1.6932	5.3938	0.7328
	GIFB [16]	18.1522	0.6902	1.7491	4.9052	0.6693
	LLDB[78]	15.7244	0.5801	1.4794	3.5617	0.6307
	PA-PCNN-NSSTB [14]	16.9849	0.6532	**2.1038**	5.1954	0.5013
	Proposed	**18.6264**	**0.6976**	2.0968	**5.9629**	**0.7855**
Figs. 3.16 (c) and (d)	**CTB [30]**	17.9678	0.6127	1.8702	4.8238	0.5376
	CB [29]	18.2474	0.6438	1.7162	5.1034	0.6257
	GIFB [16]	18.2906	0.6746	1.5689	5.1263	0.6985
	LLDB [78]	15.5468	0.5762	1.5638	4.9336	0.6278
	PA-PCNN-NSSTB [14]	16.4669	0.6367	1.9267	5.0689	0.5731
	Proposed	**19.9033**	**0.6992**	**2.8728**	**5.1568**	**0.7025**

Figs. 3.16 (a), (b), and (c) are, respectively, the same as Figs. 3.11 (a), (b), and (c), and are repeated here for convenience. Table 3.9 lists the corresponding objective fusion results. The corresponding subjective fusion results are shown in Fig. 3.16. It is clear from these results that the proposed method performs better than the traditional methods do in terms of both the objective and subjective fusion results.

We next consider the fusion of an MR image and FDG-PET image of a mild Alzheimer's disease patient taken from the WBA database.

Fusion of an MR image with the FDG-PET image of a mild Alzheimer's disease patient

We now consider fusing each of the MR images of a mild Alzheimer's disease patient shown in Figs. 3.17 (a), (b), and (c) with the functional FDG-PET data shown in Fig. 3.17 (d). Figs. 3.17 (a), (b), and (c) are, respectively, the same as Figs. 3.12 (a), (b), and (c), and are repeated here for convenience. Table 3.10 lists the corresponding objective fusion

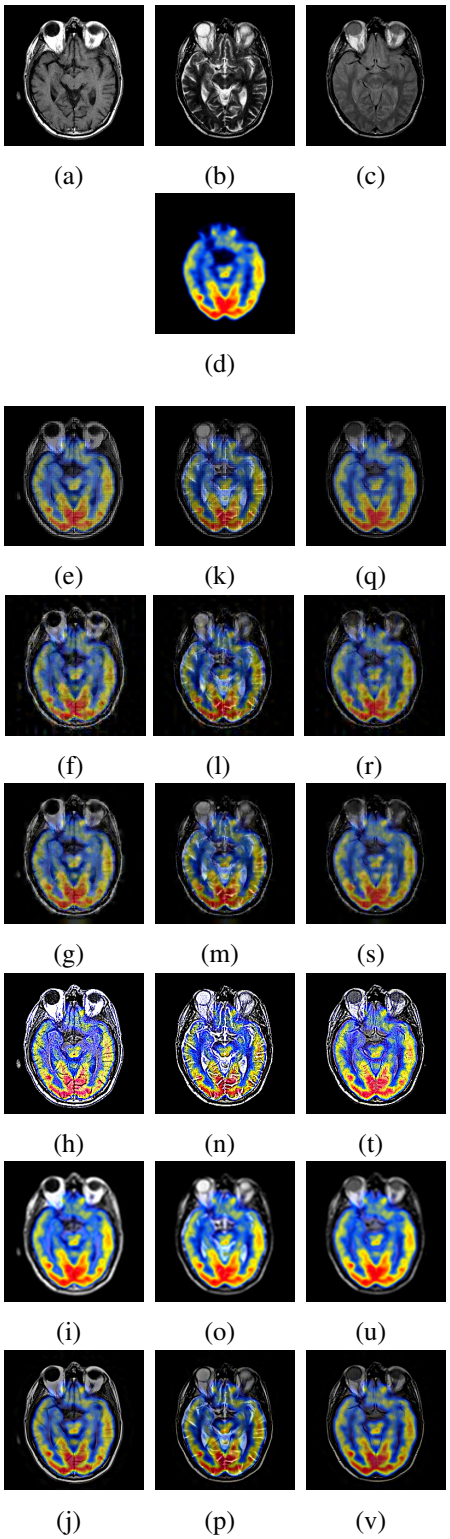


Figure 3.17: Multimodal fusion results of an MR image with the PET image of a mild Alzheimer's disease patient. (a) MR-T1 image (b) MR-T2 image (c) MR-PD image, and (d) FDG-PET image. (a) and (d), (b) and (d), and (c) and (d) are the pairs of source images being fused. (e)-(j), (k)-(p), and (q)-(v) are the fused images of (a) and (d), (b) and (d), and (c) and (d), respectively. Rows 3 to 8 correspond to the fused images obtained using curvelet-based [30], contourlet-based [29], guided image filter-based [16], local Laplacian filtering domain-based [78], parameter-adaptive pulse coupled neural network and nonsubsampled shearlet-based [14], and proposed methods, respectively. NOTE: Figs. 3.17 (a), (b), and (c) are, respectively, the same as Figs. 3.12 (a), (b), and (c), and are repeated here for convenience.

Table 3.10: Objective Fusion Results of 2D Neuroimages in Fig. 3.17

Source Images	Method	PSNR	SSIM	MI	Entropy	$Q^{\frac{f_1 f_2}{F}}$
Figs. 3.17 (a) and (d)	**CTB[30]**	15.6707	0.6369	1.8564	4.6118	0.5530
	CB[29]	15.6985	0.5274	1.6706	5.6544	0.6286
	GIFB[16]	15.7968	0.6325	1.7519	5.1185	0.6515
	LLDB[78]	14.3914	0.5737	1.4910	3.3657	0.6262
	PA_PCN_NNSSTB[14]	15.6722	0.5850	1.8903	4.9748	0.5102
	Proposed	**16.1480**	**0.7141**	**1.8925**	**5.7751**	**0.8255**
Figs. 3.17 (b) and (d)	**CTB[30]**	15.6616	0.6258	1.7394	4.2633	0.5484
	CB[29]	15.6993	0.5164	1.5765	5.2628	0.6013
	GIFB[16]	15.7978	0.6185	1.6470	4.6844	0.6423
	LLDB[78]	14.9290	0.5774	1.2805	3.8800	0.5908
	PA-PCNN-NSSTB[14]	15.6880	0.5816	1.7446	4.6772	0.5015
	Proposed	**16.3369**	**0.6313**	**1.8431**	**5.5082**	**0.8180**
Figs. 3.17 (c) and (d)	**CTB[30]**	17.4610	0.6582	1.8526	4.3291	0.5391
	CB[29]	17.4849	0.5683	1.7143	5.3242	0.6264
	GIFB[16]	17.5935	0.6585	1.8073	4.7825	0.6565
	LLDB[78]	16.8937	0.5726	1.4906	4.6048	0.5070
	PA-PCNN-NSSTB[14]	16.7705	0.5976	**1.8744**	4.7313	0.4751
	Proposed	**17.7378**	**0.7553**	1.8656	**5.6774**	**0.7874**

results. The corresponding subjective fusion results are shown in Fig. 3.17. It is clear from these results that the proposed method performs better than the traditional methods do in terms of both the objective and subjective fusion results.

Experiments on Fusing More than Two Images taken from WBA database

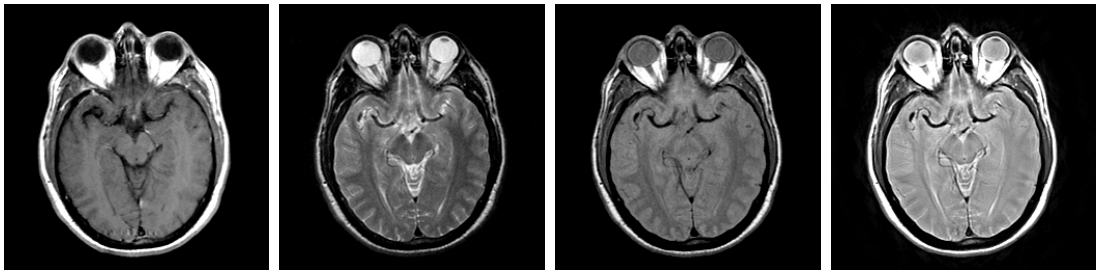
We now consider the fusion of three 2D brain images from the Whole Brain Atlas database (WBA) using the proposed method. Since data is not available for considering the fusion of more than two functional images, we have considered only the fusion of three MR images that reveal structural information.

To compare the performance of the proposed method on fusing three images with that of fusing two images, we choose the source images, MR-T1, MR-T2, and MR-PD shown

Table 3.11: Objective Fusion Results of three 2D MR images

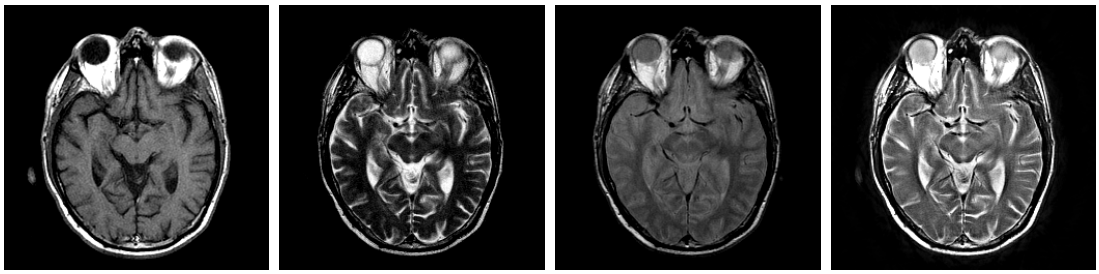
Source Images	**PSNR**	**SSIM**	**MI**	**Entropy**	**Q**$^{\frac{f_1 f_2 f_3}{F}}$
Fig. 3.18 (a),(b), and (c)	22.8285	0.9878	2.7578	5.5354	0.7417
Fig. 3.19 (a),(b), and (c)	19.1549	0.8948	2.8636	5.6831	0.8863

in Fig. 3.11 (a), (b), and (c), respectively, of the lyme encephalopathy patient. The objective quality of the fused image as measured by the various metrics is summarized in Table 3.11. By comparing these results with that shown in Table 3.6, it is observed that the values of the various metrics when all the three images are fused are higher than those when any two of the three images are fused. This is further confirmed by the subjective fusion results. Fig. 3.18 shows the fusion results of the three multimodal MR images. It is clearly seen from Fig. 3.18 (d) that the signal intensities are well enhanced when all the different weighting techniques, namely, T1, T2, and PD of MR images are considered for fusion unlike fusing only any two different weighting techniques of MR images as in Fig. 3.11. Similar experiments are carried out for the the source images, MR-T1, MR-T2, and MR-PD shown in Fig. 3.12 (a), (b), and (c), respectively, of the mild Alzheimer's disease patient. The corresponding objective fusion results are given in Table 3.11 and the subjective fusion results are shown in Fig. 3.19.



(a) (b) (c) (d)

Figure 3.18: Multimodal fusion of three 2-D real MR neuroimages of a lyme encephalopathy patient from the Whole Brain Atlas database. (a) MR-T1 image. (b) MR-T2 image. (c) MR-PD image. (d) Fused image of (a), (b) and (c)



(a) (b) (c) (d)

Figure 3.19: Multimodal fusion of three 2-D real MR neuroimages of a mild Alzheimer's disease patient from the Whole Brain Atlas database. (a) MR-T1 image. (b) MR-T2 image. (c) MR-PD image. (d) Fused image of (a), (b) and (c)

We next consider the fusion of two MR images and one PET/SPECT image. To compare the performance of the proposed method on fusing three images with that of fusing two images, we choose the source images, MR-T1, MR-T2, MR-PD, and SPECT shown in Fig. 3.16 (a), (b), (c), and (d), respectively, of the lyme encephalopathy patient. The combinations are MR-T1/MR-T2/SPECT, MR-T1/MR-PD/SPECT, MR-T2/MR-PD/SPECT. The objective quality of the fused images obtained using as measured by the various metrics is summarized in Table 3.12. By comparing these results with that shown in Table 3.9, it is observed that the values of four of the five metrics considered are higher than those when any one of the MR images and the SPECT image are fused at a time. The subjective fusion results are shown in Fig. 3.20. It is clear from the fused images that the features

Table 3.12: Objective Fusion Results of three images (two MR images and one PET/SPECT image)

Source Images	**PSNR**	**SSIM**	**MI**	**Entropy**	**Q** $^{\frac{f_1 f_2 f_3}{F}}$
Fig. 3.20 (a),(b), and (d)	23.1210	0.9629	2.9646	5.9635	0.8894
Fig. 3.20 (a),(c), and (d)	22.2547	0.8607	2.9695	5.5574	0.8400
Fig. 3.20 (b),(c), and (d)	21.9836	0.8119	2.8986	5.4101	0.8460
Fig. 3.21 (a),(b), and (d)	19.6532	0.8874	1.8538	5.6973	0.8953
Fig. 3.21 (a),(c), and (d)	19.3368	0.8448	1.4480	5.5478	0.8966
Fig. 3.21 (b),(c), and (d)	19.9019	0.8542	1.3277	5.2921	0.8892

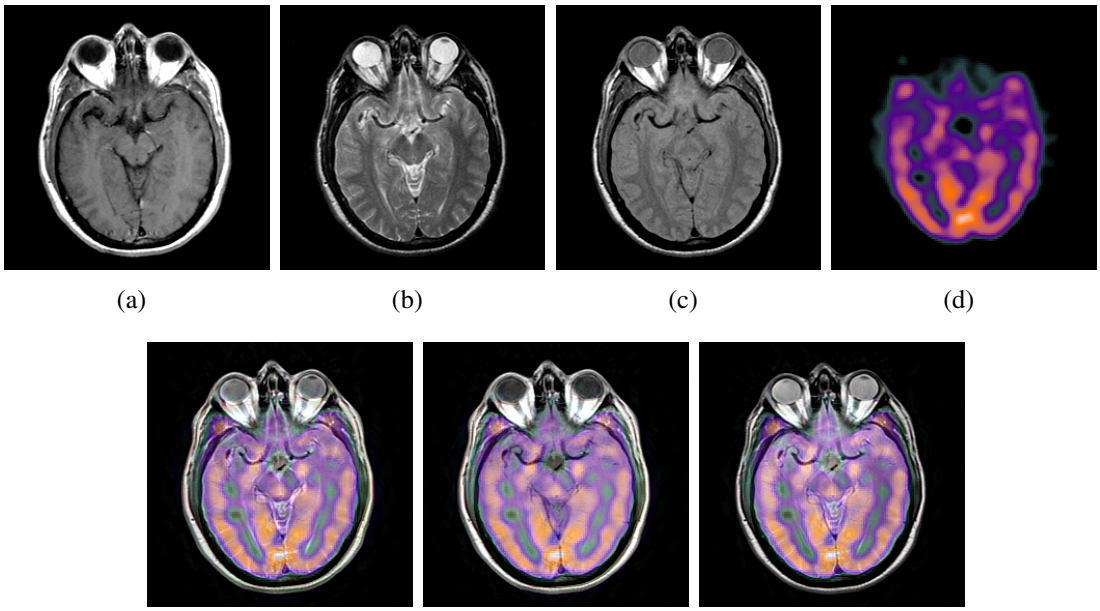
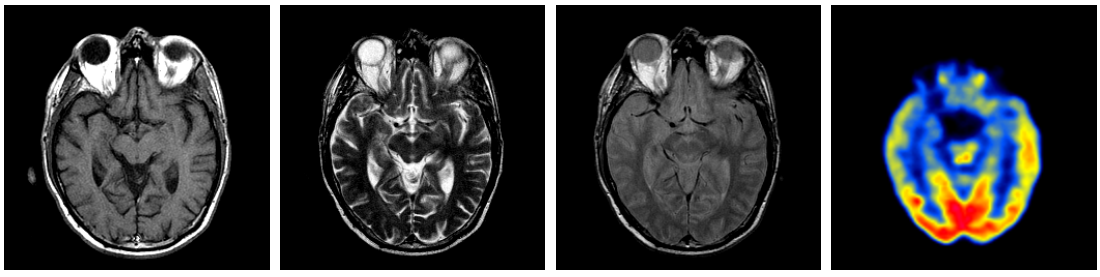
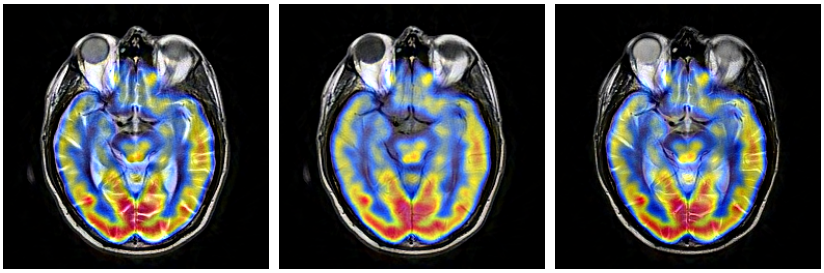


Figure 3.20: Multimodal fusion of three 2-D real neuroimages (MR-SPECT combinations) of a lyme encephalopathy patient from the Whole Brain Atlas database. (a) MR-T1 image. (b) MR-T2 image. (c) MR-PD image. (d) SPECT image. (e) Fused image of (a), (b), and (d). (f) Fused image of (a), (c), and (d). Fused image of (b), (c) and (d). NOTE: Figs. 3.20 (a), (b), and (c) are, respectively, the same as Figs. 3.16 (a), (b), and (c), and are repeated here for convenience.

are well enhanced when any two of the MR images and the SPECT image are fused unlike fusing any one of the MR images and SPECT image. Similar experiments are carried out for the source images, MR-T1, MR-T2, MR-PD, and PET images shown in Fig. 3.17 (a), (b), (c), and (d) respectively, of the mild Alzheimer's disease patient. The corresponding objective fusion results are given in Table 3.12. By comparing these results with that shown in Tables 3.10 and 3.12, it is observed that the proposed method yields the highest values in all the measures considered except the mutual information and entropy measures, when any two of the MR images and the PET image are fused. The subjective fusion results are shown in Fig. 3.21. It is clear from the fused images that the features are well enhanced when any two of the MR images and the PET image are fused unlike fusing any one of the MR images and PET image.



(a) (b) (c) (d)



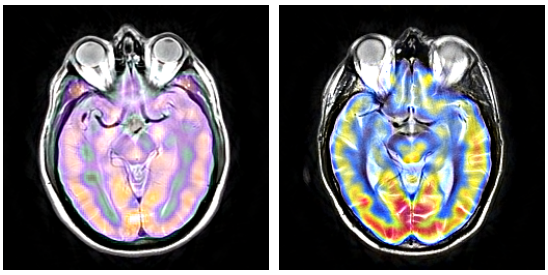
(e) (f) (g)

Figure 3.21: Multimodal fusion of three 2-D real neuroimages (MR-PET combinations) of a mild Alzheimer's disease patient from the Whole Brain Atlas database. (a) MR-T1 image. (b) MR-T2 image. (c) MR-PD image. (d) FDG-PET image. (e) Fused image of (a), (b), and (d). (f) Fused image of (a), (c), and (d). Fused image of (b), (c) and (d). NOTE: Figs. 3.21 (a), (b), and (c) are, respectively, the same as Figs. 3.17 (a), (b), and (c), and are repeated here for convenience.

Finally, we consider the fusion of four input images, MR-T1, MR-T2, MR-PD, and SPECT, shown in Figs. 3.20 (a), (b), (c), and (d), respectively, of the lyme encephalopathy patient. The objective quality of the fused image obtained using as measured by the various metrics is summarized in Table 3.13. By comparing these results with that shown in Tables 3.9 and 3.12, it is observed that the proposed method yields the highest values in all the measures considered except the mutual information and entropy and measures, when any

Table 3.13: Objective Fusion Results of four images (three MR images and one PET/SPECT image)

Source Images	**PSNR**	**SSIM**	**MI**	**Entropy**	**Q**$^{\frac{f_1 f_2 f_3 f_4}{F}}$
Fig. 3.20 (a),(b),(c), and (d)	23.5487	0.9783	2.8745	5.1237	0.8981
Fig. 3.21 (a),(b),(c), and (d)	20.1342	0.8983	1.8755	5.6332	0.8997



(a) (b)

Figure 3.22: Multimodal fusion of four 2-D real neuroimages (MR-PET combinations) from the Whole Brain Atlas database. (a) Fused image of Figs. 3.20 (a), (b), (c), and (d). (b) Fused image of Figs. 3.21 (a), (b), (c), and (d).

one or two of the MR images and the SPECT image are fused. We believe that these scores could be further increased by considering more number of functional images. We also notice that the proposed technique boosts the normalized weighted performance metric, $Q^{\frac{f_1 f_2}{F}}$ as the number of images to be fused increases. Fig. 3.22 (a) is the subjective fusion result. It is clear from the fused image that when we integrate the structural features from different MR images along with the functional data, the features are well enhanced unlike having only two or three images. Similar experiments are carried out for the input images, MR-T1, MR-T2, MR-PD, and PET, shown in Figs. 3.21 (a), (b), (c), and (d), respectively, of the mild Alzheimer's disease patient. The corresponding objective fusion results are given in Table 3.13 and the subjective fusion result is shown in Fig. 3.22 (b). As mentioned earlier, in our experiments we have chosen two scales, one with six and the other with ten directions. However, the performance could be further improved by increasing the number of scales and directions at the expense of increased complexity.

3.6 Summary

In this chapter, a novel neuroimaging fusion algorithm has been proposed based on the statistical properties of the nonsubsampled shearlet transform (NSST) and an energy maximization fusion rule. The multiscale and multidirectional capabilities of the NSST for optimally representing neuroimaging data have been explored. Then, a linear forward model has been presented to describe the relationship between observed brain images and the underlying heterogeneous features of the brain acquired by multiple sensors, along with the application of the inverse of the linear forward model to neuroimaging fusion. Moreover, a parametric brain image modeling based on the statistical properties of NSST coefficients has been proposed. The marginal statistics of the detail subband coefficients are also modeled by the student's t location-scale PDF, which has heavier tails, making it more prone to outliers. Considering the discriminative features contained in high-frequency subbands, a new fusion rule for multimodal fusion based on maximizing the energy in such subbands has been developed as well. Experiments were carried out on fusing two or more multimodal neuroimages taken from the BrainWeb, Alzheimer's disease neuroimaging initiative (ADNI), and Whole Brain Atlas databases. The subjective and objective results have shown that the proposed neuroimaging fusion method is robust and significantly outperforms the state-of-the-art methods even for noisy images.

Chapter 4

LHAttNet: A Lightweight Deep Convolutional Neural Network Extracting Local and Global Contextual Features for the Classification of Alzheimer's Disease

4.1 Introduction

Recent advancements in the classification of Alzheimer's disease have leveraged the automatic feature generation capability of convolutional neural networks (CNNs) using neuroimaging biomarkers. However, as pointed out in Section 1.5, most of the existing CNN-based methods often disregard the local features of the brain data, which leads to a loss of

subtle fine-grained features in the brain imaging data. Moreover, the existing CNN architectures, which rely mainly on global features, do not pay much attention to the discriminability of the extracted features for the task of classification of Alzheimer's disease. In [82], a novel semi-supervised classification framework using an unsupervised autoencoder to select a subset from given structural and clinical features has been proposed. Then, the linear and nonlinear relationships of the features have been explored, followed by a supervised multinomial logistic layer to discriminate the patients having Alzheimer's disease (AD), mild cognitive impairment (MCI), and healthy control (HC). In [83], a stacked sparse autoencoder-based method has been proposed for data imputation and to extract the discriminative structural and clinical features. In [84], a deep learning model in which a recurrent neural network (RNN) and a pre-trained deep CNN have been combined to extract inter-slice and intra-slice features from the 3D MR data for the classification of Alzheimer's disease. In [85], a deep learning model using an ensemble of hybrid deep learning architectures to leverage more complete spatial information from MRI data has been proposed. To leverage more complete spatial information, including the local and global features of the brain, most of the aforementioned deep networks have concatenated the features extracted from multiple CNN architectures for the classification of Alzheimer's disease. Although these networks provide acceptable performance, their practical utility in real-world scenarios is constrained by the use of a large number of parameters aimed at enriching the extracted features. These networks have treated all the extracted features uniformly without taking into consideration the varying importance of the different features in brain imaging data for the classification of Alzheimer's disease.

In view of the above limitations, this chapter proposes a lightweight deep CNN capable of extracting both local and global contextual features by using an attention mechanism for the task of classification of Alzheimer's disease [86], [87]. The main idea used in designing the proposed network is to process the local and global features separately by using

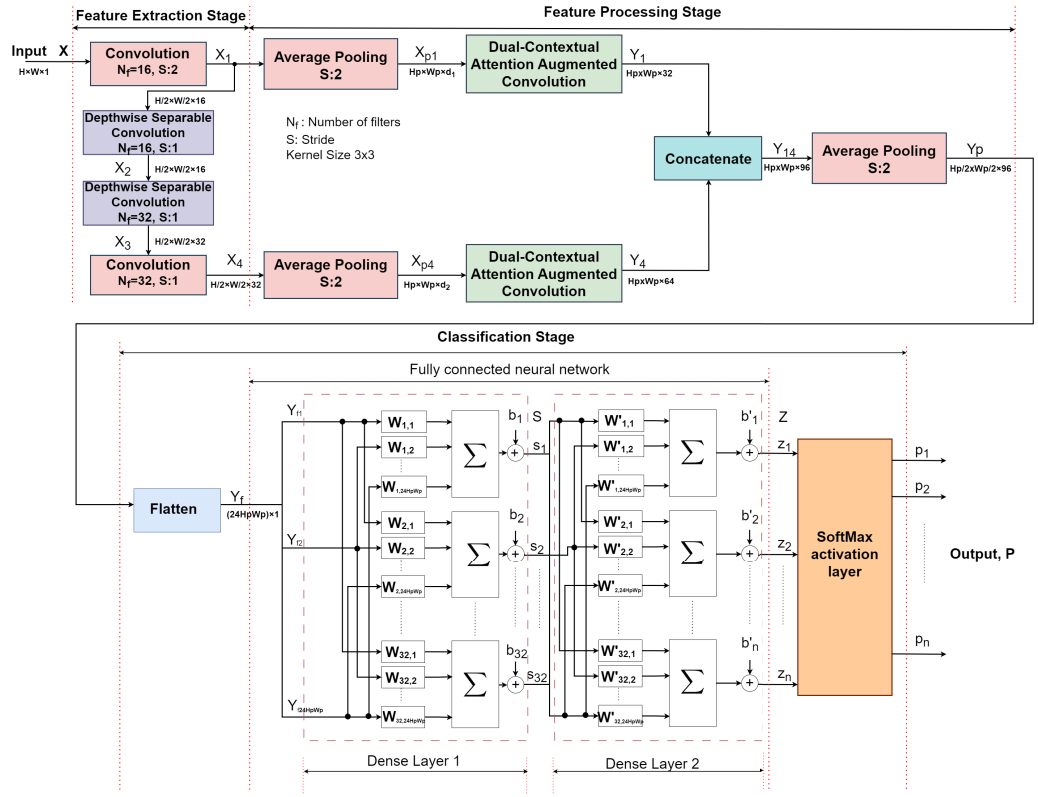
modules designed to emphasize contextual features relevant to Alzheimer's disease classification. The fused local and global contextual features are then used to classify Alzheimer's disease. The performance of the CNN model developed is studied using images of a given modality for the tasks of binary and multiclass classifications of Alzheimer's disease.

4.2 Proposed Scheme for Classification of Alzheimer's Disease

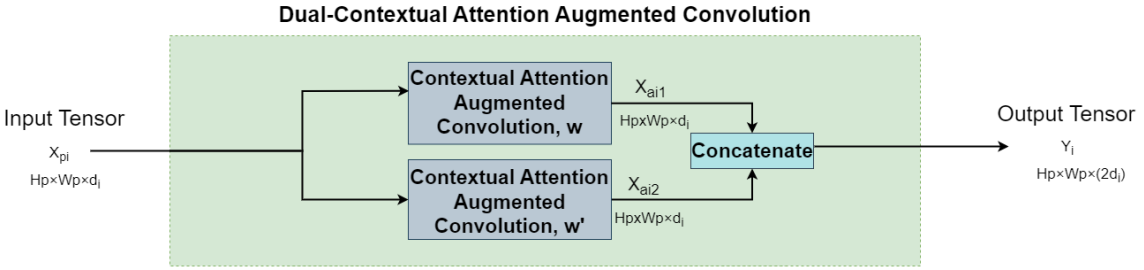
In this section, we develop the architecture of our proposed convolutional neural network (CNN) referred to as LHAttNet for the classification of Alzheimer's disease. The overall architecture of our proposed CNN for the classification of Alzheimer's disease is shown in Fig. 4.1(a). As seen from this figure, the architecture consists of three stages, namely, feature extraction, feature processing, and classification. In the feature extraction module, features are first extracted at four different hierarchical levels from the input MRI data. In the feature processing stage, the features extracted from two of the levels that best represent the local and global features are chosen and processed separately by paying attention to the salient features capable of discriminating the features representing the various classes of Alzheimer's disease. Finally, in the third stage, the processed local and global features are fused and used for the classification of Alzheimer's disease. In the following subsections, the three stages of the overall architecture are described in detail.

4.2.1 Feature Extraction Stage

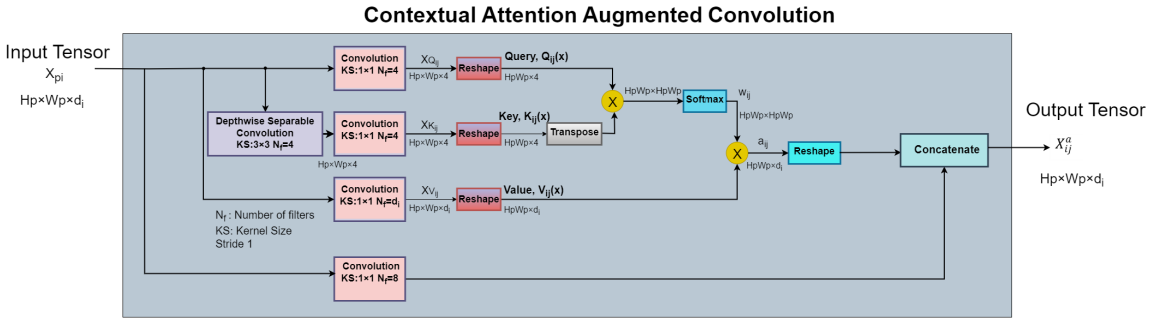
The input image X of size $H \times W$ representing a single slice of MRI data is passed through a sequence of four convolutional layers in order to extract features at four hierarchical levels. At the first hierarchical level, a standard convolution operation with 16 filters and a stride of 2 is applied to X, resulting in a tensor X_1 consisting of 16 feature maps. In



(a)



(b)



(c)

Figure 4.1: Architecture of the proposed Alzheimer's disease classification scheme.

view of using a stride of 2, the size of each map in X_1 is $\frac{H}{2} \times \frac{W}{2}$. At the second and third hierarchical levels, separable convolutions employing 16 and 32 filters, each with a stride of 1, are performed to generate tensors X_2 and X_3 of sizes $\frac{H}{2} \times \frac{W}{2} \times 16$ and $\frac{H}{2} \times \frac{W}{2} \times 32$ respectively. At the fourth level, again a standard convolution operation with 32 filters and a stride of 1 is applied to X_3, producing a tensor X_4 consisting of 32 feature maps each of size $\frac{H}{2} \times \frac{W}{2}$. The kernel size used in all the convolution operations in this stage is 3×3.

The choice of using standard convolution at the first and fourth hierarchical levels, and separable convolution at the intermediate levels in the feature extraction stage of the proposed architecture, instead of using either standard or separable convolution at all the four hierarchical levels, is aimed at balancing the trade-offs between computational efficiency and the classification accuracy.

4.2.2 Feature Processing Stage

In the second stage, first, the lowest and the highest level features extracted by the first stage, which respectively correspond to the local and global features of the input image, are individually processed through two separate streams. The process initiates with an average pooling operation using a kernel of size 3×3 with a stride of 2, resulting in a pooled feature tensor $X_{pi}(i = 1 \text{ or } 4)$ of size $H_p \times W_p \times d_i$, where $H_p = \left\lfloor \frac{(H/2-1)}{2} \right\rfloor + 1$, $W_p = \left\lfloor \frac{(W/2-1)}{2} \right\rfloor + 1$, and $d_i = 16$ for $i = 1$ and $d_i = 32$ for $i = 4$. The idea behind these streams is to obtain the set of features Y_i ($i = 1$ or 4) from the local or global pooled features X_{pi} through the mechanism of self-attention weighting provided by the dual-contextual attention augmented convolution block, so as to enhance the classification. The resulting weighted local and global feature tensors, Y_1 and Y_4, are concatenated to obtain a set of fused feature tensor, Y_{14}. Finally, in this stage, the fused feature tensor Y_{14} is average-pooled to obtain reduced spatial size feature maps Y_p.

Fig. 4.1(b) shows the top-level architecture of the dual-contextual attention augmented

convolution block that consists of a pair of contextual attention augmented convolution blocks whose detailed architecture is shown in Fig. 4.1(c). The outputs of the two contextual attention augmented convolution blocks, which are connected in parallel (i.e., they have the same input X_{pi}), are concatenated to produce the output Y_i of size $H_p \times W_p \times 2d_i$ of the dual- contextual attention augmented convolution block. The parallel implementation of a pair of contextual attention augmented convolution blocks enables the model to independently focus on salient features by assigning weights to the local or global pooled features, thus producing attention-weighted features X_{ij}^a ($j = 1$ or 2) that are crucial for the classification task.

Using the input tensor X_{pi}, three operations are performed in parallel to produce three different maps called the query, key, and value maps, denoted by Q_{ij}, K_{ij}, and V_{ij}, respectively. To produce the query Q_{ij}, a standard convolution using a kernel of size 1×1 with 4 filters is applied to the input X_{pi}. The resulting feature maps $X_{Q_{ij}}$ of size $H_p \times W_p \times 4$ are then arranged to have the query map of size $H_pW_p \times 4$. The key map is also produced in the same way as the query, except that in the computation of the key, neighbourhood information is exploited by passing the input through a depthwise separable convolution using a kernel of size 3×3 with 4 filters. The resulting feature maps $X_{K_{ij}}$ of size $H_p \times W_p \times 4$ are then reshaped to have the key map of size $H_pW_p \times 4$. The value map V_{ij} is produced by applying a convolution using a kernel of size 1×1 with the d_i filters (i.e., the number of filters used is the same as the number of channels in the input) to the input X_{pi}. The resulting feature maps Xv_{ij} of size $H_p \times W_p \times d_i$ are then reshaped to have a map of size $H_pW_p \times d_i$. The key and query maps are then used to compute the weight matrix as given by

$$W_{ij} = \frac{\exp(K_{ij}^T Q_{ij})}{\sum \exp(K_{ij}^T Q_{ij})}, \quad i = 1 \text{ or } 4; \quad j = 1 \text{ or } 2 \tag{4.1}$$

Using the weight matrix W_{ij} given by (4.1) and the value map, V_{ij}, attention weighted

feature tensor X_{ij}^{a} is obtained as

$$X_{ij}^{a} = \textit{Resh}[W_{ij}V_{ij}], \quad i = 1 \text{ or } 4; \quad j = 1 \text{ or } 2 \tag{4.2}$$

where *Resh*(.) denotes the reshape operation used to convert a 2D feature map to a 3D feature tensor. The resulting attention-weighted feature tensor X_{ij}^{a} is of size $H_p \times W_p \times d_i$. The weighted local and global feature tensors, Y_1 and Y_4, are of sizes $H_p \times W_p \times 32$ and $H_p \times W_p \times 64$, respectively. Thus, the size of the fused feature tensor Y_{14} is $H_p \times W_p \times 96$ and that of average-pooled feature maps Y_p is $\frac{H_p}{2} \times \frac{W_p}{2} \times 96$.

At this point, it is important to point out the main difference in the features of X_i and those of Y_i ($i = 1$ or 4). The features of Y_i have been obtained from X_i through a process of attention mechanism used in the second stage in which the values of its features have been discriminatively weighted in terms of their importance to the task of classification. This process has been carried out individually for the local and global features and then the two types of features have been concatenated leading to the generation of the features in Y_p that can be expected to be highly discriminative and rich.

4.2.3 Classification Stage

In this stage, the feature tensor Y_p obtained from the feature processing stage is used to perform the classification of the MRI data X input to the network. This stage consists of four layers: a flatten layer, two dense layers, and an activation layer. The flatten layer is used to convert the feature tensor Y_p of size $\frac{H_p}{2} \times \frac{W_p}{2} \times 96$ to a one-dimensional feature vector Y_f of size $(24H_pW_p) \times 1$. The resulting flattened vector is input to the first fully connected dense layer consisting of 32 neurons. The weighted sum produced by each of the 32 neurons is biased to produce one of the 32 components of the output S of this dense layer. This output is then fed to the second dense layer consisting of n neurons, where n

is the number of classes in the input data (for example, for binary classification $n = 2$). The resulting output Z has n components. Finally, Z is passed through a layer of softmax activation functions to obtain the probabilities $P = [p_1, p_2, \ldots, p_n]^T$ of the n classes.

4.3 Experimental Results

In this section, we study the performance of the proposed deep convolutional neural network (CNN) designed for the classification of Alzheimer's disease using single modality images, namely, T1-weighted MR, T2-weighted MR, and 18F-FDG PET images. We organize this section into a number of subsections. We first present the description of the input dataset used in our study for training and evaluating the proposed CNN. Since the MR and PET imaging provides three-dimensional data and the input to our CNN is a 2D image, we next explain how 2D slices are obtained from the 3D data. This is followed by providing details of the hardware and software platforms used to train and test the proposed network along with the implementation details. In the succeeding subsection, we compare the performance of the proposed method with that of the state-of-the-art methods. In the final subsection, we provide results of some ablation studies demonstrating the effectiveness of the various ideas used in the design of the proposed architecture.

4.3.1 Dataset Description

The input images used in this study are downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (`https://adni.loni.usc.edu`) [80]. This site has been collecting data on various biomarkers of AD, including neuroimaging data such as MRI and PET since 2004. We have selected 756 patients comprising 179, 192, 173, and 212 patients, respectively, and have categorized into AD, pMCI, sMCI, and HC categories of the disease based on the following criteria:

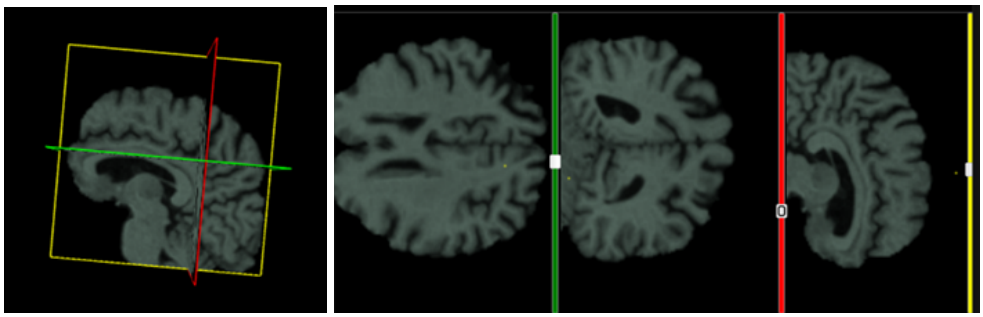
- HC: Patients who are diagnosed as healthy controls at their baseline visit.
- sMCI: Patients who are diagnosed as MCI at baseline and did not progress to AD within 36 months of the baseline visit.
- pMCI: Patients who are diagnosed as MCI at baseline and progress to AD within 36 months of the baseline visit.
- AD: Patients who are diagnosed as full-fledged AD at their baseline visit.

The ADNI database contains both raw and processed data for each scan of the selected patients. We have downloaded processed T1-weighted structural MRI, T2-weighted structural MRI, and 18F-FDG PET images of each of the selected patients at their baseline visit. We register the processed 3D images on the standard MNI152 template [88] to remove global linear differences, including global translation, scale, and rotation differences. Then, we remove the non-brain tissues in all the registered images using a skull-stripping method provided in FSL toolbox [89]. The preprocessed images are standardized to have a size of $160 \times 128 \times 128$ voxels.

In this work, 80% of the 756 selected and preprocessed 3D images are used for training, and the remaining 20% are used for testing. Within the training set, 20% of the 3D images are used for validation, while the remaining 80% are employed for actual training of the proposed LHAttNet.

4.3.2 Choosing 2D Slices

The data obtained from the ADNI database is three-dimensional. While training a 3D deep convolutional neural network using 3D data can leverage the complete spatial information, in practice, it comes with increased computational complexity. As an alternative approach, as shown in Fig. 4.2, we chose to transect the brain along the x, y, and z planes commonly referred to as the coronal, axial, and sagittal planes, respectively. Among these, sagittal



(a) (b)

Figure 4.2: Transecting a 3D brain of an AD patient into axial (yellow box), coronal (red box), and sagittal (green box) planes.

plane slices are known to offer clearer information about crucial brain regions such as the hippocampus, amygdala, and corpus callosum. These slices provide insights into the extent of degeneration in AD [90]. Hence, we selected the most informative 100 sagittal slices according to the decreasing order of their entropy values.

4.3.3 Hardware and Software Platforms and Implementation Details

The proposed network is trained and tested on a hardware platform equipped with an NVIDIA GeForce RTX 2070 GPU with dedicated 6 GB memory, and an Intel(R) Core(TM) i7-8750H CPU with 32 GB RAM and 6 cores. The proposed CNN architecture is implemented in Python using an open-source machine learning framework, TensorFlow [91].

The network is trained with a maximum of 100 epochs, each using a batch size of 32. The cross-entropy loss function is used as the objective function, which is minimized using the Adam optimizer with an initial learning rate of 10^{-5}. The learning rate is decreased by a factor of 0.3 when there is no improvement in the validation loss metric. Moreover, if there is no improvement in the validation loss metric for a specified number of consecutive epochs (with the patience parameter set to 10), the process of training is stopped, even if the maximum number of 100 epochs has not yet been reached, and the model at the onset of degradation is chosen as the final trained model. Otherwise, the training is continued up

to the 100 epochs, and the model at that epoch is selected as the final trained model. Each epoch takes approximately 90 seconds to complete the training.

4.3.4 Classification Performance and Comparison

In this section, we present experimental results on the proposed classification scheme using sagittal slices of single modality images, namely, T1-weighted MR, T2-weighted MR, and 18F-FDG PET images. For each experiment, we first provide the training and validation curves of our method for two binary classification tasks: AD vs. HC and pMCI vs. sMCI. We then present the classification results obtained from the trained model and compare these results with state-of-the-art methods for AD vs. HC and pMCI vs. sMCI binary classification tasks. It is worth acknowledging that an exact comparison among these methods is not feasible due to the inability to know the test sets used by the different methods, even though such sets are selected from the same dataset. The classification results for the state-of-the-art methods are taken as reported in the respective works. We then provide the training and validation curves of our proposed method for multiclass classification of four classes: AD, pMCI, sMCI, and HC. Then, we present the multiclass classification results obtained from the trained model. Since the dual-contextual attention augmented convolution modules are the cores of the feature processing stage of the proposed network, we also present and discuss the visual illustration of a typical feature map resulting from these modules.

Finally, we also provide the results on the computational complexity of the proposed network in terms of the number of parameters used.

Experiment 1: Classification Results using T1-weighted MR Images

In this experiment, we study the performance of our proposed LHAttNet using sagittal T1-weighted MR images as input for binary classifications, AD vs. HC and pMCI vs. sMCI,

as well as for multiclass classification AD vs. pMCI vs. sMCI vs. HC.

Results on AD vs. HC and pMCI vs. sMCI Binary Classifications: In order to evaluate the performance of the proposed method, the network is trained on sagittal T1-weighted MR images for the AD vs. HC and pMCI vs. sMCI binary classification tasks individually. To assess the process of training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs.4.3(a) and 4.3(c) show the training and validation losses, for the AD vs. HC and pMCI vs. sMCI classifications, respectively. Figs. 4.3(b) and 4.3(d) show the corresponding training and validation accuracy curves, respectively. It is seen from Figs. 4.3(a) and 4.3(c) that the

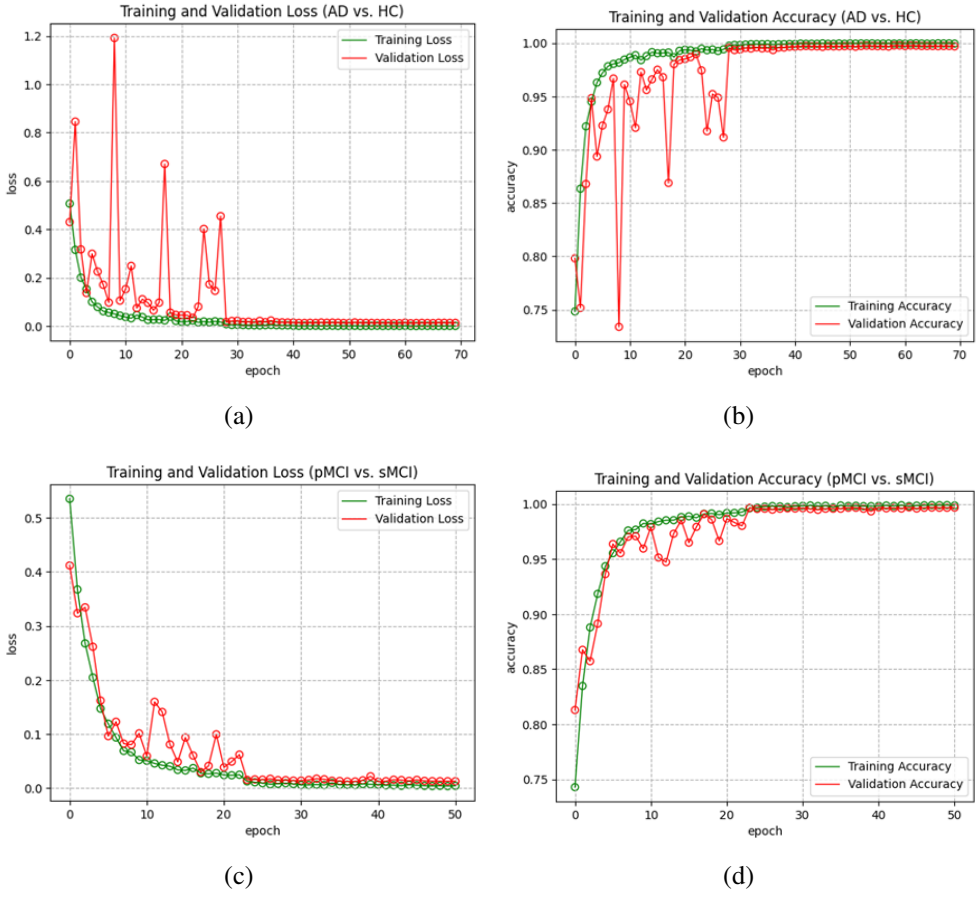


Figure 4.3: Training and validation performance curves on T1-weighted MR images for two binary classifications: (a) and (b) for AD vs. HC, and (c) and (d) for pMCI vs. sMCI.

validation losses decrease and converge starting at the 28^{th} and 23^{th} epochs, respectively. The gap between the validation and training losses is very small. Figs.4.3(b) and 4.3(d) show the training and validation accuracy curves for the AD vs. HC and pMCI vs. sMCI classifications, respectively. It is seen from these figures that both the training and validation accuracies are high and reasonably close to each other. From the above observations, it is concluded that the two models are appropriately trained.

To evaluate the performance of the trained models, we classify the sagittal T1-weighted sMR images annotated as AD and HC in the testing dataset (not used in the training or validation) as either AD or HC using the first trained model. Similarly, the images annotated as pMCI and sMCI in the testing dataset are classified using the second trained model. Table 4.1 gives the performance results in terms of the ACC, SEN, SPE, and AUC metrics provided by the proposed and the four state-of-the-art methods, namely, 3D DenseNet and spherical harmonics (DN-SH) [92], hierarchical fully convolutional network (H-FCN) [93], dual attention multi-instance deep learning network (DA-MIDL) [94], and 3D Residual attention deep neural network (3D ResAttNet) [95], for both the binary classification

Table 4.1: Performance results, in percentage, of LHAttNet and four state-of-the-art methods for AD vs. HC and pMCI vs. sMCI classifications on the T1-weighted MR images from the ADNI dataset

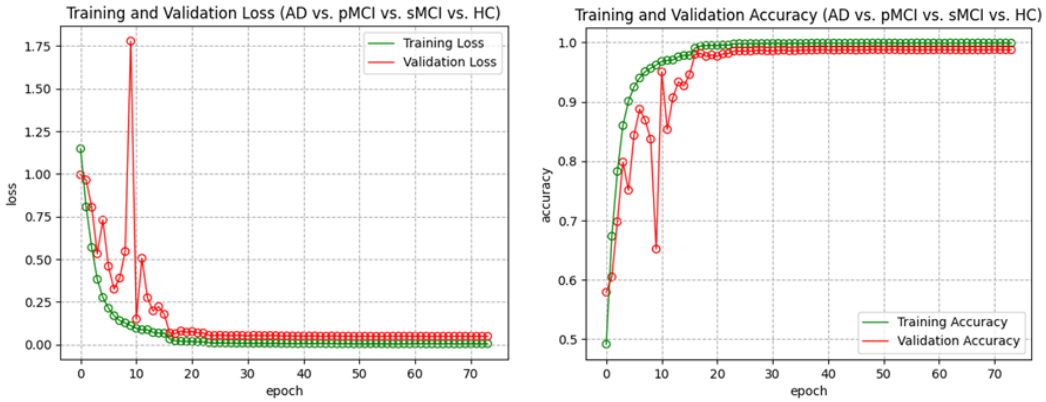
Methods	AD vs. HC				pMCI vs. sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
DN-SH [92]	**92.29**	**90.63**	93.72	96.95	75	73.33	76.19	79.70
H-FCN [[93]]	90.3	82.4	96.5	95.1	**80.9**	52.6	85.4	78.1
DA-MIDL [94]	92.4	91	**93.8**	96.5	80.2	**77.1**	**82.6**	**85.1**
3D ResAttNet [95]	91.3	91	91.9	98.4	82.1	81.2	80.9	92
LHAttNet (Proposed)	97.43	93.21	96.34	**96.62**	96.55	94.43	92.32	95.68

The quantities underscored in red fonts, italic in blue fonts, and bold in green fonts indicate, respectively, the best, second-best, and third-best results.

tasks. It is seen from this table that for the AD vs. HC binary classification, our proposed LHAttNet provides the best results in terms of the ACC and SEN metrics with substantial margins from the methods providing the second-best results for these two metrics. In terms of the SPE metric, even though the proposed scheme ranks second, its performance is very close to that of the second-best method. In terms of the AUC metric, the proposed network ranks third. However, the values of the AUC metric of our network and that of the second-best performing scheme are, respectively, 1.78 and 1.45 from that of the best performing scheme. For the pMCI vs. sMCI binary classification, the proposed LHAttNet scheme provides the best results in terms of all the four metrics among all the classification schemes. The values of the ACC, SEN, SPE, AUC metrics of the proposed scheme are higher than that of the second-best performing scheme by 14.55, 13.43, 6.92, and 3.68, respectively. It is clear from the results of Table 4.1 that the proposed scheme is capable of providing a performance that is substantially higher than that of other state-of-the-art schemes for the task of binary classification of Alzheimer's disease.

Results on AD vs. pMCI vs. sMCI vs. HC Multiclass Classification: We now consider the task of multiclass classification to classify a sagittal T1-weighted MR image into one of the four classes, namely, AD, pMCI, sMCI, and HC. To assess the process of multiclass training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 4.4(a) and 4.4(b) show the training and validation losses and accuracies, respectively. It is seen from these figures that the validation loss decrease and converge starting at the 23^{rd} epoch. The gap between the validation and training losses is very small. It is also seen from Fig. 4.4(b) that both the training and validation accuracies are high and reasonably close to each other. From these observations, it is concluded that the model is appropriately trained.

To evaluate the performance of the model trained for multiclass classification, we classify the images in the test set of sagittal T1-weighted MR images annotated as AD, pMCI,



(a) (b)

Figure 4.4: Training and validation performance curves on T1-weighted MR images for multiclass classification task (AD vs. sMCI vs. pMCI vs. HC): (a) Loss and (b) Accuracy.

sMCI, or HC. Table 4.2 presents the performance results provided by the proposed scheme in terms of the ACC, SEN, SPE, PREC, F1, and AUC metrics for the task of multiclass classification. It is to be noted from this table that the proposed scheme provides a performance that is larger than 95 % for all the metrics, except for the sensitivity metric for the AD class, which has a value of 94 %. It should be pointed out that the proposed scheme provides the lowest accuracy of 97.76 % for the AD class and the highest accuracy of 99.70

Table 4.2: Performance results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC classifications on T1-weighted MR images from the ADNI dataset

Classes	AD vs. pMCI vs. sMCI vs. HC					
	ACC	SEN	SPE	PREC	F1	AUC
AD	97.76	94	98.75	96.34	95.16	96.46
pMCI	98.02	96.24	98.62	95.96	96.10	97.43
sMCI	98.45	96.66	98.99	96.68	96.67	97.82
HC	99.70	98.57	98.91	96.84	97.70	98.66

% for the HC class. Table 4.3 gives the average performance results, in terms of the various metrics, provided by the proposed LHAttNet method and the state-of-the-art method of [96] for the multiclass classification. It is clear from this table that the proposed scheme provides the results in terms of all the metrics, which are significantly superior to that provided by the other method.

Table 4.3: Average performance results, in percentage, of LHAttNet and the state-of-the-art method of [96] for AD vs. pMCI vs. sMCI vs. HC multiclass classification on the T1-weighted MR images from the ADNI dataset

Methods	AD vs. pMCI vs. sMCI vs. HC					
	ACC	SEN	SPE	PREC	F1	AUC
Deep-stacked CNN-BiLSTM [96]	92.62	94.02	N/A	94.02	92.56	N/A
LHAttNet (Proposed)	**98.48**	**96.37**	**98.82**	**96.45**	**96.40**	**97.59**

Since the dual-contextual attention augmented convolution modules are the cores of the feature processing stage of the proposed network, we choose two typical feature maps from Y_1 and Y_4 resulting from the two dual-contextual attention augmented convolution modules, respectively, corresponding to an image belonging to each of the four classes, and display them in Fig 4.5.

Fig. 4.5 provides a visual illustration of the effectiveness of using the dual-contextual attention augmented convolution modules in the proposed network. Each row in this figure shows an image from one of the four classes and a typical feature map taken from each of the sets of feature maps X_1, Y_1, X_4, and Y_4 corresponding to that image. A comparison of the typical feature map taken from X_1 with that from Y_1 shows that they both contain the local features of the input image. However, the local features in the map of Y_1 are more enhanced than that of the map in X_1. Similarly, a comparison between the typical feature

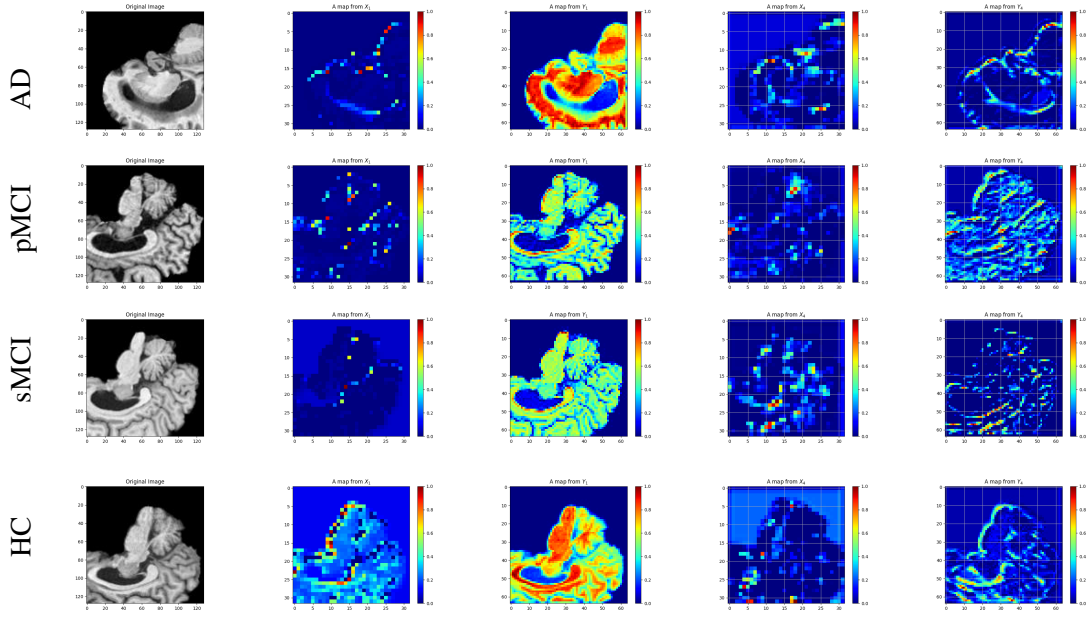


Figure 4.5: Visual illustration of the effectiveness of the dual contextual attention-augmented convolution modules using T1-weighted MR images. (a) Images selected from the AD, pMCI, sMCI, and HC classes. (b) Typical feature map from X_1 resulting from the first convolutional layer corresponding to an image in (a). (c) Typical feature map resulting from the dual contextual attention-augmented convolution module processing the local features obtained from an image in (a). (d) Typical feature map from X_4 resulting from the fourth convolutional layer corresponding to an image in (a). (e) Typical feature map resulting from the dual contextual attention-augmented convolution module processing the global features obtained from an image in (a).

map taken from X_4 and that from Y_4 shows that they both include global features of the input image. However, the global features in the map of Y_4 are relatively more pronounced.

Experiment 2: Classification Results using T2-weighted MR Images

In this experiment, we study the performance of our proposed LHAttNet using sagittal T2-weighted MR images as input for binary classifications, AD vs. HC and pMCI vs. sMCI, as well as for multiclass classification AD vs. pMCI vs. sMCI vs. HC.

Results on AD vs. HC and pMCI vs. sMCI Binary Classifications: In order to evaluate

the performance of the proposed method, the network is trained on sagittal T2-weighted MR images for the AD vs. HC and pMCI vs. sMCI binary classification tasks individually. To assess the process of training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 4.6(a) and 4.6(c) show the training and validation losses, for the AD vs. HC and pMCI vs. sMCI classifications, respectively. Figs. 4.6(b) and 4.6(d) show the corresponding training and validation accuracy curves, respectively. It is seen from Figs. 4.6(a) and 4.6(c) that the validation losses decrease and converge starting around 33^{th} and 55^{th} epochs, respectively. The gap between the validation and training losses is very small. Figs.4.6(b) and 4.6(d)

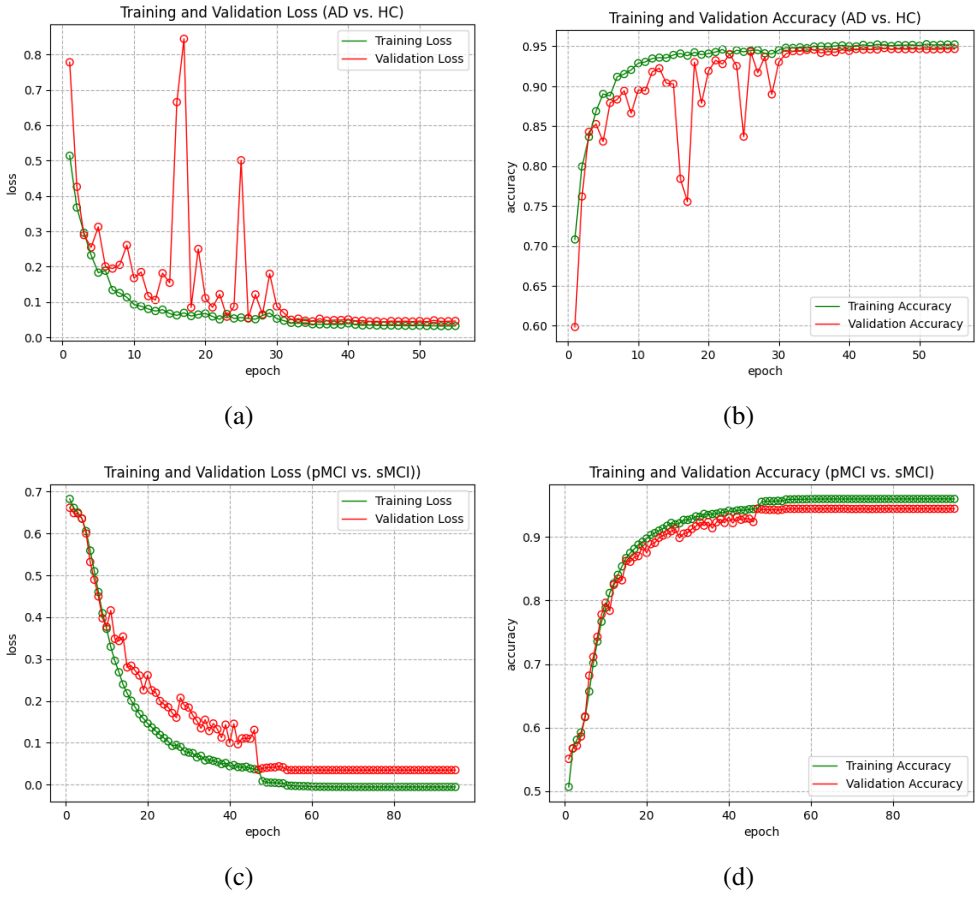


Figure 4.6: Training and validation performance curves on T2-weighted MR images for two binary classifications: (a) and (b) for AD vs. HC, and (c) and (d) for pMCI vs. sMCI.

show the training and validation accuracy curves for the AD vs. HC and pMCI vs. sMCI classifications, respectively. It is seen from these figures that both the training and validation accuracies are around 95% and reasonably close to each other. From the above observations, it is concluded that the two models are appropriately trained.

To evaluate the performance of the trained models, we classify the sagittal T2-weighted MR images annotated as AD and HC in the testing dataset (not used in the training or validation) as either AD or HC using the first trained model. Similarly, the images annotated as pMCI and sMCI in the testing dataset are classified using the second trained model. Table 4.4 gives the performance results in terms of the ACC, SEN, SPE, and AUC metrics provided by the proposed and two state-of-the-art methods. One of the state-of-the-art methods uses Leung-Malik filtered bank features [97], while the other employs hybrid enhanced independent component analysis (ICA) for AD vs. HC binary classification task. The results for the pMCI vs. sMCI binary classification task are not available in the respective works. It is seen from this table that for the AD vs. HC binary classification task, the proposed LHAttNet provides the best results in terms of the ACC, SEN, and SPE metrics with substantial margins from the methods providing the second-best results for these three metrics. For the pMCI vs. sMCI binary classification task, the proposed LHAttNet scheme provides the acceptable results in terms of all the four metrics. The values of the ACC,

Table 4.4: Performance Results, in percentage, of LHAttNet and two state-of-the-art methods for AD vs. HC and pMCI vs. sMCI classifications on the T2-weighted MR images from the ADNI dataset

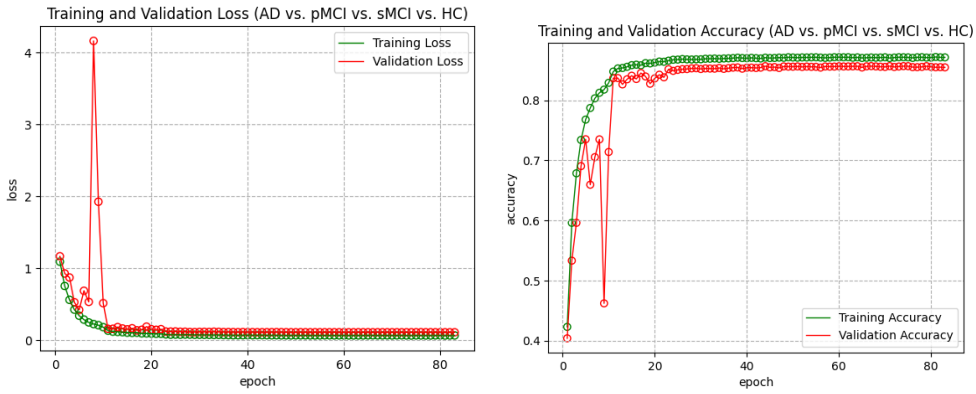
Methods	AD vs. HC				pMCI vs. sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Leung-Malik [97]	84.24	89.85	79.22	N/A	N/A	N/A	N/A	N/A
ICA [98]	90.47	86.66	N/A	N/A	N/A	N/A	N/A	N/A
LHAttNet (Proposed)	**93.81**	**90.25**	**93.63**	**93.70**	**91.78**	**90.54**	**93.44**	**90.11**

SEN, SPE, and AUC metrics of the proposed scheme are 91.78, 90.54, 93.44, and 90.11, respectively. It is clear from the results of Table 4.4 that the proposed scheme significantly outperforms state-of-the-art methods in the task of AD vs. HC binary classification for Alzheimer's disease. However, due to the unavailability of results pertaining to the pMCI vs. sMCI binary classification task, a comparative analysis remains unfeasible.

For the AD vs. HC binary classification task, the values of the ACC, SEN, SPE, and AUC metrics of the proposed scheme when using T2-weighted MR images, as reported in Table 4.1, are lower than those obtained using T1-weighted MR images by 3.62, 2.96, 2.71, and 2.92, respectively. In the pMCI vs. sMCI binary classification task, the values of the ACC, SEN, and AUC metrics of the proposed scheme when using T2-weighted MR images, as reported in Table 4.1, are lower than those obtained using T1-weighted MR images by 4.77, 3.89, and 5.57, respectively, except for the SPE metric, which is higher by 1.12.

Results on AD vs. pMCI vs. sMCI vs. HC Multiclass Classification: We now consider the task of multiclass classification to classify a sagittal T2-weighted MR image into one of the four classes, namely, AD, pMCI, sMCI, and HC. To assess the process of multiclass training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 4.7(a) and 4.7(b) show the training and validation losses and accuracies, respectively. It is seen from these figures that the validation loss decrease and converge starting around the 12^{th} epoch. The gap between the validation and training losses is very small. It is also seen from Fig. 4.7(b) that both the training and validation accuracies are high and reasonably close to each other. From the above observation, it is concluded that the model is appropriately trained.

To evaluate the performance of the model trained for multiclass classification, we classify the images in the test set of sagittal T2-weighted MR images annotated as AD, pMCI, sMCI, or HC. Table 4.5 presents the performance results provided by the proposed scheme



(a) (b)

Figure 4.7: Training and validation performance curves on T2-weighted MR images for multiclass classification task (AD vs. sMCI vs. pMCI vs. HC): (a) Loss and (b) Accuracy.

in terms of the ACC, SEN, SPE, and AUC metrics for the task of multiclass classification. It is to be noted from this table that the proposed scheme provides a performance that is larger than 84 % for all the metrics. It should be pointed out that the proposed scheme provides the lowest accuracy of 85 % for the sMCI class and the highest accuracy of 86.30 % for the HC class.

Table 4.6 gives the performance results, in terms of the various metrics, provided by the proposed LHAttNet method for the multiclass classification. For multiclass classification task, the values of ACC, SEN, SPE, AUC metrics of the proposed scheme on using T2-weighted images is lower than that on using T1-weighted MR images, as reported in Table 4.3. Since the dual-contextual attention augmented convolution modules are the cores of the feature processing stage of the proposed network, we choose two typical feature maps

Table 4.5: Performance Results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC classifications on the T2-weighted MR images from the ADNI dataset

Class	ACC	SEN	SPE	AUC
AD	85.50	86.78	89.78	84.72
pMCI	86	87.32	90.92	85.91
sMCI	85	91.84	93.75	89.65
HC	86.30	92.38	92.87	90.64

Table 4.6: Average performance Results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC multiclass classification on the T2-weighted sagittal MR images from the ADNI dataset

Methods	AD vs. pMCI vs. sMCI vs. HC			
	ACC	SEN	SPE	AUC
LHAttNet (Proposed)	**85.70**	**89.58**	**91.83**	**87.73**

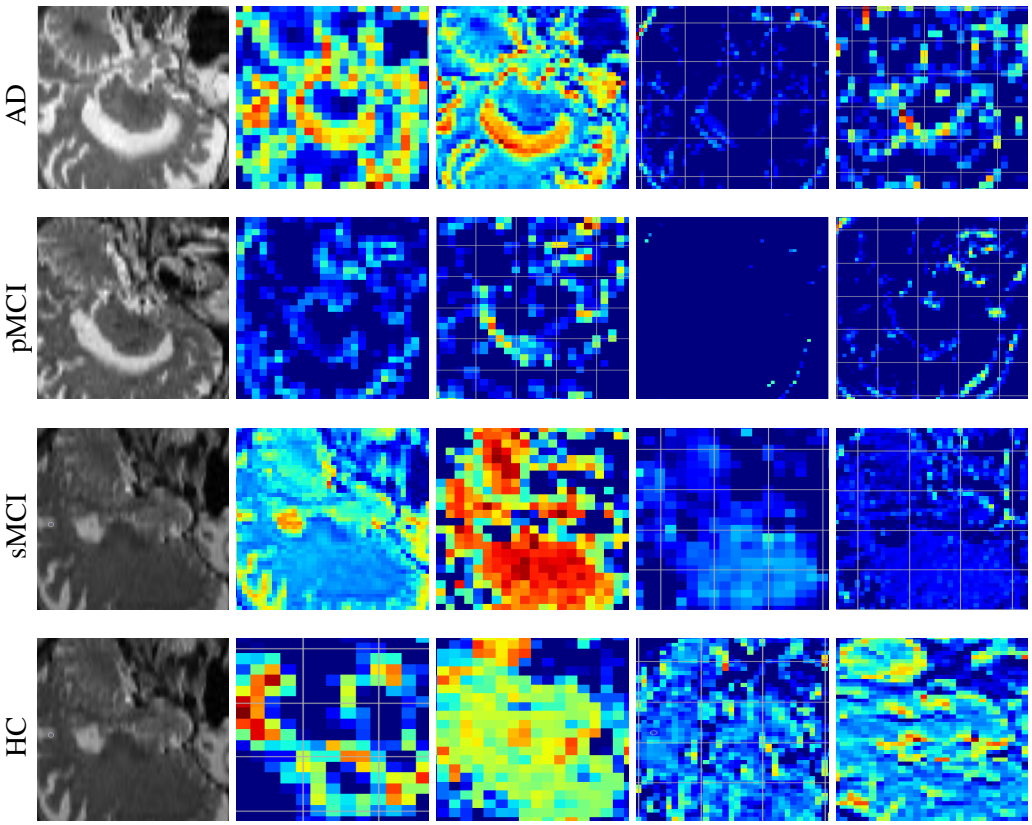


Figure 4.8: Visual illustration of the effectiveness of the dual contextual attention-augmented convolution modules using T2-weighted MR images.. (a) Images selected from the AD, pMCI, sMCI, and HC classes. (b) Typical feature map from X_1 resulting from the first convolutional layer corresponding to an image in (a). (c) Typical feature map resulting from the dual contextual attention-augmented convolution module processing the local features obtained from an image in (a). (d) Typical feature map from X_4 resulting from the fourth convolutional layer corresponding to an image in (a). (e) Typical feature map resulting from the dual contextual attention-augmented convolution module processing the global features obtained from an image in (a).

from Y_1 and Y_4 resulting from the two dual-contextual attention augmented convolution modules, respectively, corresponding to an image belonging to each of the four classes, and display them in Fig 4.8.

Fig. 4.8 provides a visual illustration of the effectiveness of using the dual-contextual attention augmented convolution modules in the proposed network. Each row in this figure shows an image from one of the four classes and a typical feature map taken from each of the sets of feature maps X_1, Y_1, X_4, and Y_4 corresponding to that image. A comparison of the typical feature map taken from X_1 with that from Y_1 shows that they both contain the local features of the input image. However, the local features in the map of Y_1 are more enhanced than that of the map in X_1. Similarly, a comparison between the typical feature map taken from X_4 and that from Y_4 shows that they both include global features of the input image. However, the global features in the map of Y_4 are relatively more pronounced.

Experiment 3: Classification Results using 18F-FDG PET Images

In this experiment, we study the performance of our proposed LHAttNet using 18F-FDG PET images as input for binary classifications, AD vs. HC and pMCI vs. sMCI, as well as for multiclass classification AD vs. pMCI vs. sMCI vs. HC.

Results on AD vs. HC and pMCI vs. sMCI Binary Classifications: In order to evaluate the performance of the proposed method, the network is trained on sagittal 18F-FDG PET images for the AD vs. HC and pMCI vs. sMCI binary classification tasks individually. To assess the process of training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 4.9(a) and 4.9(c) show the training and validation losses, for the AD vs. HC and pMCI vs. sMCI classifications, respectively. Figs. 4.9(b) and 4.9(d) show the corresponding training and validation accuracy curves, respectively. It is seen from Figs. 4.9(a) and 4.9(c) that the validation losses decrease and converge starting at the 23^{th} and 40^{th} epochs,

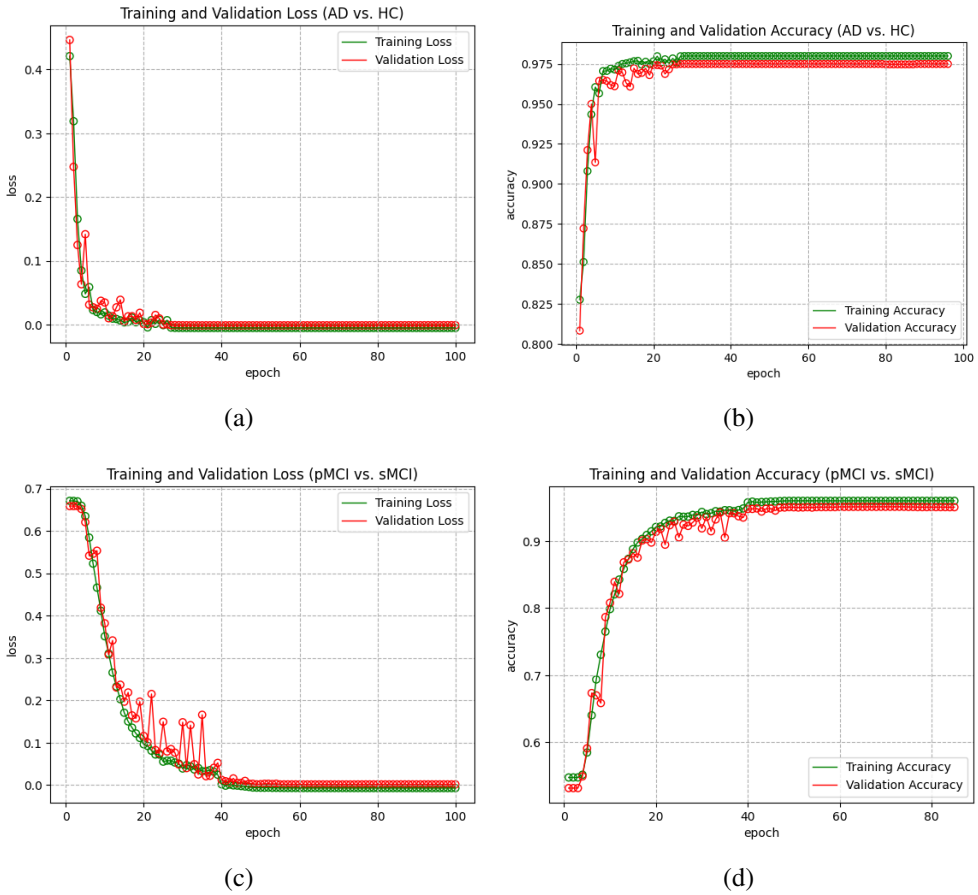


Figure 4.9: Training and validation performance curves on 18F-FDG PET images for two binary classifications: (a) and (b) for AD vs. HC, and (c) and (d) for pMCI vs. sMCI.

respectively. The gap between the validation and training losses is very small. Figs.4.9(b) and 4.9(d) show the training and validation accuracy curves for the AD vs. HC and pMCI vs. sMCI classifications, respectively. It is seen from these figures that both the training and validation accuracies are high and reasonably close to each other. From the above observations, it is concluded that the two models are appropriately trained.

To evaluate the performance of the trained models, we classify the sagittal 18F-FDG PET images annotated as AD and HC in the testing dataset (not used in the training or validation) as either AD or HC using the first trained model. Similarly, the images annotated as pMCI and sMCI in the testing dataset are classified using the second trained model. Table 4.7 gives the performance results in terms of the ACC, SEN, SPE, and AUC

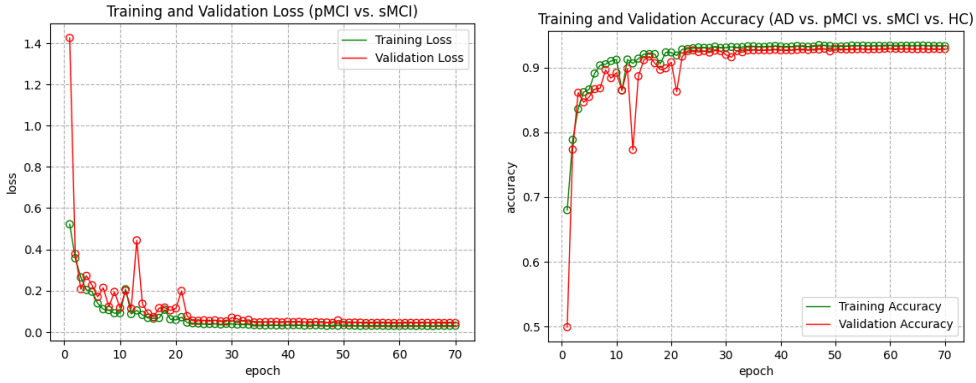
Table 4.7: Performance results, in percentage, of LHAttNet and three state-of-the-art methods for AD vs. HC and pMCI vs. sMCI classifications on the 18F-FDG PET images from the ADNI dataset

Methods	AD vs. HC				pMCI vs. sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
CNN_RNN [99]	91.2	91.4	91	91.8	N/A	N/A	N/A	N/A
Radiomics [100]	N/A	N/A	N/A	N/A	90.62	87.50	93.39	90
Static_Dynamic [101]	N/A	N/A	N/A	N/A	72.42	36.36	97.83	64.43
LHAttNet (Proposed)	**96.51**	**94.38**	**95.42**	**96.70**	**94.77**	**91.38**	**94.15**	**92.80**

metrics provided by the proposed and the four state-of-the-art methods for both the binary classification tasks. It is seen from this table that for the AD vs. HC binary classification, our proposed LHAttNet provides the best results in terms of all the metrics. For the pMCI vs. sMCI binary classification, the proposed LHAttNet scheme provides the best results in terms of all the four metrics among all the classification schemes. The values of the ACC, SEN, SPE, AUC metrics of the proposed scheme are significantly higher than that of the second-best performing scheme. It is clear from the results of Table 4.7 that the proposed scheme is capable of providing a performance that is substantially higher than that of other state-of-the-art schemes for the task of binary classification of Alzheimer's disease.

Results on AD vs. pMCI vs. sMCI vs. HC Multiclass Classification: We now consider the task of multiclass classification to classify a sagittal 18F-FDG PET image into one of the four classes, namely, AD, pMCI, sMCI, and HC. To assess the process of multiclass training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 4.10(a) and 4.10(b) show the training and validation losses and accuracies, respectively. It is seen from these figures that the validation loss decrease and converge starting at the 23^{rd} epoch. The gap between the validation and training losses is very small. It is also seen from Fig. 4.10(b) that both the

training and validation accuracies are high and reasonably close to each other. From the above observation, it is concluded that the model is appropriately trained. To evaluate the



(a) (b)

Figure 4.10: Training and validation performance curves using 18F-FDG PET images for multiclass classification task (AD vs. sMCI vs. pMCI vs. HC): (a) Loss and (b) Accuracy.

performance of the model trained for multiclass classification, we classify the images in the test set of sagittal MRI images annotated as AD, pMCI, sMCI, or HC. Table 4.8 presents the performance results provided by the proposed scheme in terms of the ACC, SEN, SPE, PREC, F1, and AUC metrics for the task of multiclass classification. It is to be noted from this table that the proposed scheme provides a performance that is larger than 90 % for all the metrics. It should be pointed out that the proposed scheme provides the lowest accuracy of 92.84 % % for the pMCI class and the highest accuracy of 98.01 % for the HC class. Table 4.9 gives the performance results, in terms of the various metrics, provided by the proposed LHAttNet method for the multiclass classification. It is clear from this table that

Table 4.8: Performance results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC classifications on the 18F-FDG PET images from the ADNI dataset

Class	ACC	SEN	SPE	AUC
AD	93.43	92.48	92.89	91.54
pMCI	92.84	92.50	94.90	93.55
sMCI	97.69	92.52	96.91	95.56
HC	98.01	92.54	98.96	97.51

Table 4.9: Average performance results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC multiclass classification on the 18F-FDG PET images from the ADNI dataset

Methods	AD vs. pMCI vs. sMCI vs. HC			
	ACC	SEN	SPE	AUC
LHAttNet (Proposed)	**95.49**	**92.51**	**95.89**	**94.54**

the proposed scheme provides the acceptable results in terms of all the metrics.

Since the dual-contextual attention augmented convolution modules are the cores of the feature processing stage of the proposed network, we choose two typical feature maps from Y_1 and Y_4 resulting from the two dual-contextual attention augmented convolution modules, respectively, corresponding to an image belonging to each of the four classes, and display them in Fig. 4.11.

Fig. 4.11 provides a visual illustration of the effectiveness of using the dual-contextual attention augmented convolution modules in the proposed network. Each row in this figure shows an image from one of the four classes and a typical feature map taken from each of the sets of feature maps X_1, Y_1, X_4, and Y_4 corresponding to that image. A comparison of the typical feature map taken from X_1 with that from Y_1 shows that they both contain the local features of the input image. However, the local features in the map of Y_1 are more enhanced than that of the map in X_1. Similarly, a comparison between the typical feature map taken from X_4 and that from Y_4 shows that they both include global features of the input image. However, the global features in the map of Y_4 are relatively more pronounced.

Table 4.10 gives the number of parameters employed by the various schemes listed in Tables 4.1 and 4.3. It is seen from this table that the number of parameters employed by the proposed LHAttNet is only a small fraction of that employed by the other schemes. Specifically, the number of parameters in the proposed network is about 18 % of that of the scheme of [94], which uses the lowest number of parameters of approximately 2M among

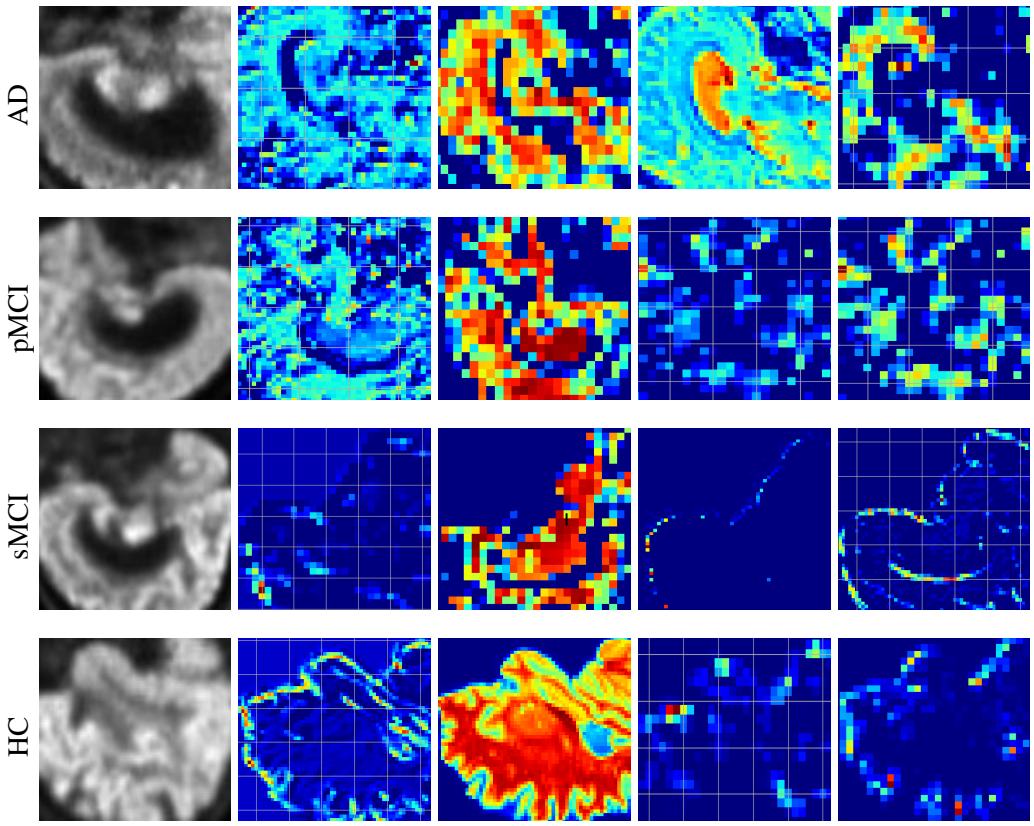


Figure 4.11: Visual illustration of the effectiveness of the dual contextual attention-augmented convolution modules using 18F-FDG PET Images. (a) Images selected from the AD, pMCI, sMCI, and HC classes. (b) Typical feature map from X_1 resulting from the first convolutional layer corresponding to an image in (a). (c) Typical feature map resulting from the dual contextual attention-augmented convolution module processing the local features obtained from an image in (a). (d) Typical feature map from X_4 resulting from the fourth convolutional layer corresponding to an image in (a). (e) Typical feature map resulting from the dual contextual attention-augmented convolution module processing the global features obtained from an image in (a).

all the other schemes used for comparison.

Table 4.10: Comparison of Total Number of Parameters (in Millions)

Type of Classification	**Methods**	**No. of parameters* (in Millions)**
Binary	DN-SH [92]	~ 12
	H-FCN [93]	~ 100
	DA-MIDL [94]	~ 2
	3D ResAttNet [95]	~ 64
	LHAttNet (Proposed)	0.5383
Multiclass	Deep-stacked CNN-BiLSTM [96]	~ 8
	LHAttNet (Proposed)	0.5384

*The number of parameters given in the table are the estimated numbers for all the networks except that for the proposed network.

4.4 Summary

In this chapter, a novel lightweight deep CNN that extracts the local and global contextual features for the classification of Alzheimer's disease has been proposed. A new module has been proposed and used in this network to process separately the local and global feature maps of sagittal slices of input data to pay special attention to discriminative contextual features. The features resulting from the dual-contextual attention augmented convolution modules operating on the local and the global features are fused and used to classify Alzheimer's disease. Experiments have been performed to demonstrate the effectiveness of the proposed network for the task of binary and multiclass classifications of Alzheimer's disease using sagittal plane slices of three single modality images, namely, T1-weighted MR, T2-weighted MR, and 18F-FDG PET images. The experimental results on T1-weighted MR images have shown that the performance of the proposed network is

generally much superior to that provided by the state-of-the art methods and the results obtained on T2-weighted MR or 18F-FDG PET images. It is to be noted that the proposed network has employed only a small fraction of the number of parameters used by the state-of-the art networks.

Chapter 5

Classification of Alzheimer's Disease using LHAttNet on Fused Images

5.1 Introduction

In this chapter, we develop a novel framework for Alzheimer's disease classification, focusing on evaluating the effectiveness of our proposed neuroimaging fusion method, which was introduced in Chapter 3. This framework is designed to enhance the accuracy of Alzheimer's disease classification through the use of our lightweight deep convolutional network, LHAttNet, detailed in Chapter 4. The neuroimaging fusion method produces a single composite image from multiple imaging modalities, containing complementary details present in each of the images used for fusion. The LHAttNet is capable of extracting both local and global contextual features using an attention mechanism. In the proposed framework, LHAttNet extracts both the local and global contextual features from the fused neuroimaging data. Our aim is to demonstrate how this integrated approach can improve the classification accuracy in diagnosing Alzheimer's disease, leveraging the strengths of both the neuroimaging fusion method and LHAttNet. We also compare the effectiveness of our proposed method to the results obtained using single modality images in Chapter 4.

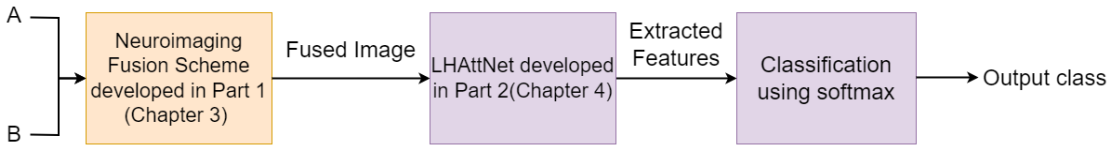


Figure 5.1: Proposed framework for the classification of Alzheimer's disease with multi-modal neuroimaging fusion.

5.2 Proposed Framework for Classification of Alzheimer's Disease

In this section, we introduce a framework designed to classify Alzheimer's disease using the local and global contextual features extracted from fused multi-modal neuroimaging data. Fig. 5.1 outlines the step-by-step process of our proposed classification framework. As seen from this figure, the proposed framework consists of three stages, namely neuroimaging fusion, feature extraction, and classification. First, using the proposed neuroimaging fusion method detailed in Chapter 3, a fused image is obtained from multiple modality images, A and B. The local and global contextual features are then extracted from the fused image using the proposed LHAttNet detailed in Chapter 4 . Finally, the classifier consisting of a fully connected (FC) layer and a softmax layer is used to classify Alzheimer's disease.

5.3 Experimental Results

As described in Section 4.3.1, we consider T1-weighted, T2-weighted structural MRI scans and 18F-FDG PET of each of these patients at their baseline visit. These modalities were selected for their complementary nature; T1-weighted and T2-weighted images provide structural details with different contrast, as well as 18F-FDG PET scans highlight metabolic activity, which are crucial for identifying areas affected by Alzheimer's disease. We register each of the downloaded 3D images on the standard MNI152 template [88] to remove global linear differences, including global translation, scale, and rotation differences. Then, we

remove the non-brain tissues in all the registered images using a skull-stripping method provided in FSL toolbox [89]. The preprocessed images are standardized to have a size of $160 \times 128 \times 128$ voxels.

In this work, 80% of the 756 preprocessed 3D images are used for training, and the remaining 20% are used for testing. Within the training set, 20% of the 3D images are used for validation, while the remaining 80% are employed for actual training of the proposed CNN. The 2D slices are obtained using the approach described in Section 4.3.2. Then the corresponding sagittal slices are fused using the implementation details described in Section 3.4.3. The hardware and software platforms and parameters set for training the network is same as the details given in Section 4.3.3.

Experiment 1: Classification Results on Fused T1-weighted and T2-weighted MR Images

In this experiment, we study the performance of our proposed LHAttNet using fused sagittal T1-weighted and T2-weighted MR images as input for binary classifications, AD vs. HC and pMCI vs. sMCI, as well as for multiclass classification AD vs. pMCI vs. sMCI vs. HC.

Results on AD vs. HC and pMCI vs. sMCI Binary Classifications: In order to evaluate the performance of the proposed method, the network is trained on fused sagittal T1-weighted and T2-weighted MR images for the AD vs. HC and pMCI vs. sMCI binary classification tasks individually. To assess the process of training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 5.2(a) and 5.2(c) show the training and validation losses, for the AD vs. HC and pMCI vs. sMCI classifications, respectively. Figs. 5.2(b) and 5.2(d) show the corresponding training and validation accuracy curves, respectively. It is seen from Figs. 5.2(a) and 5.2(c) that the validation losses decrease and converge starting at the

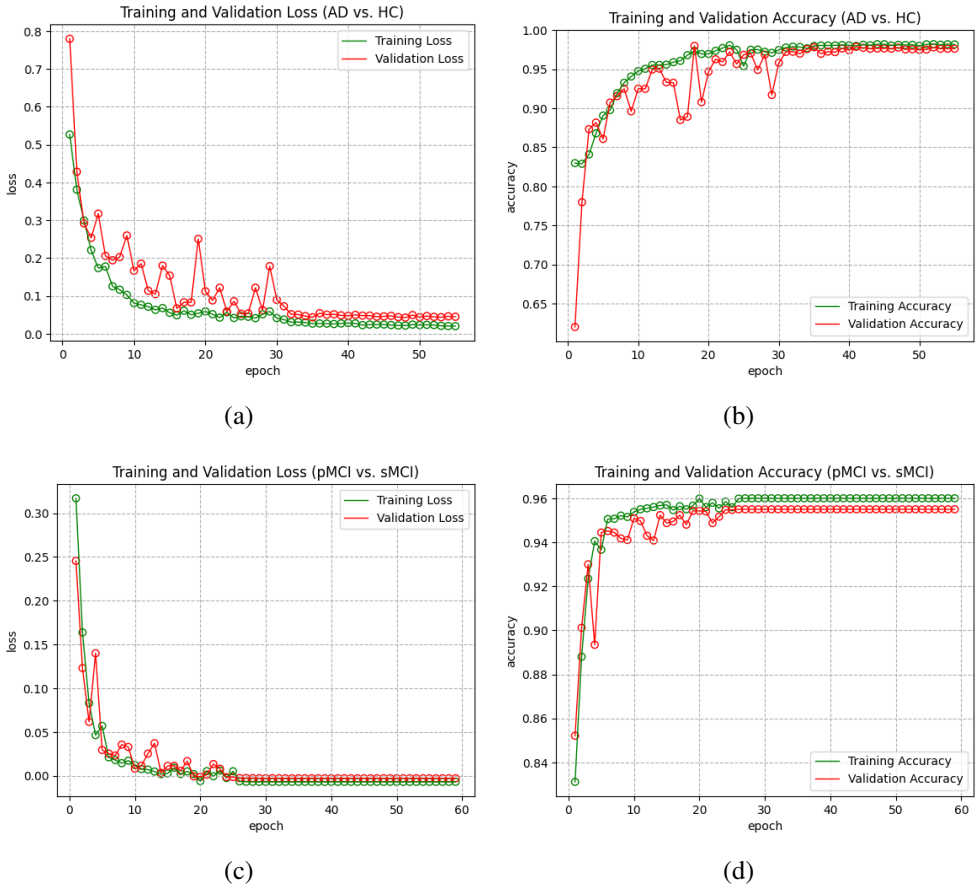


Figure 5.2: Training and validation performance curves on fused T1-weighted MR and T2-weighted MR images for two binary classifications: (a) and (b) for AD vs. HC, and (c) and (d) for pMCI vs. sMCI.

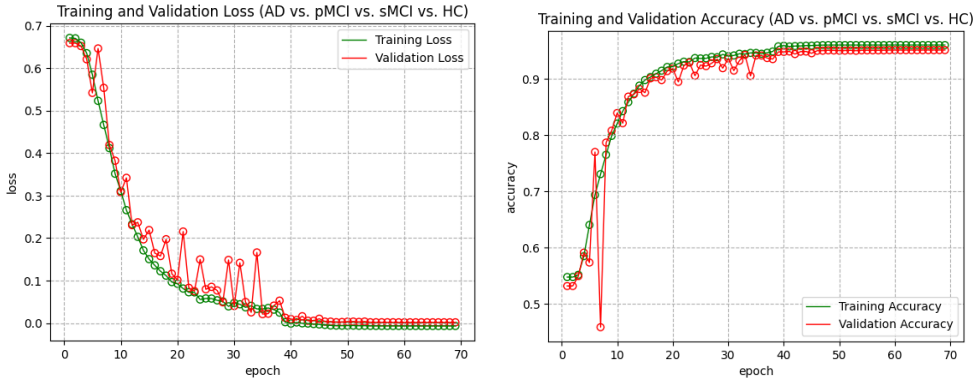
34^{th} and 27^{th} epochs, respectively. The gap between the validation and training losses is very small. Figs.5.2(b) and 5.2(d) show the training and validation accuracy curves for the AD vs. HC and pMCI vs. sMCI classifications, respectively. It is seen from these figures that both the training and validation accuracies are high and reasonably close to each other. From the above observations, it is concluded that the two models are appropriately trained.

To evaluate the performance of the trained models, we classify the sagittal fused T1-weighted and T2-weighted MR images annotated as AD and HC in the testing dataset (not used in the training or validation) as either AD or HC using the first trained model. Similarly, the images annotated as pMCI and sMCI in the testing dataset are classified

using the second trained model. Table 5.1 gives the performance results in terms of the ACC, SEN, SPE, and AUC metrics provided by the proposed method for both the binary classification tasks.

It is seen from this table that for the AD vs. HC binary classification, the proposed LHAttNet provides the classification accuracy of 97%. For the pMCI vs. sMCI binary classification, the proposed LHAttNet scheme provides the classification accuracy of 96.80 %.

Results on AD vs. pMCI vs. sMCI vs. HC Multiclass Classification: We now consider the task of multiclass classification to classify a sagittal fused T1-weighted and T2-weighted MR images into one of the four classes, namely, AD, pMCI, sMCI, and HC. To



(a) (b)

Figure 5.3: Training and validation performance curves on fused T1-weighted and T2-weighted MR images for multiclass classification task (AD vs. sMCI vs. pMCI vs. HC): (a) Loss and (b) Accuracy.

Table 5.1: Performance Results, in percentage, of LHAttNet for AD vs. HC and pMCI vs. sMCI classifications on fused T1-weighted and T2-weighted MR images from the ADNI dataset

Methods	AD vs. HC				pMCI vs. sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
LHAttNet (Proposed)	97.65	94.55	96.55	96.85	96.80	94.75	96.55	96.05

Table 5.2: Performance Results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC multiclass classification on the fused T1-weighted and T2-weighted MR images from the ADNI dataset

Methods	**AD vs. pMCI vs. sMCI vs. HC**			
	ACC	**SEN**	**SPE**	**AUC**
LHAttNet (Proposed)	97.96	94.65	96.15	93.86

assess the process of multiclass training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 5.3(a) and 5.3(b) show the training and validation losses and accuracies, respectively. It is seen from these figures that the validation loss decrease and converge starting at the 40^{th} epoch. The gap between the validation and training losses is very small. It is also seen from Fig. 5.3(b) that both the training and validation accuracies are high and reasonably close to each other. From the above observation, it is concluded that the model is appropriately trained.

To evaluate the performance of the model trained for multiclass classification, we classify the images in the test set of sagittal fused T1-weighted and T2-weighted MR images annotated as AD, pMCI, sMCI, or HC. Table 5.2 presents the performance results provided by the proposed scheme in terms of the ACC, SEN, SPE,and AUC metrics for the task of multiclass classification. The proposed method provides classification accuracy of 97.96% for multiclass classification task.

Experiment 2: Classification Results on Fused T1-weighted MR and FDG-PET Images

In this experiment, we study the performance of our proposed LHAttNet using fused sagittal T1-weighted and FDG-PET images as input for binary classifications, AD vs. HC and pMCI vs. sMCI, as well as for multiclass classification AD vs. pMCI vs. sMCI vs. HC.

Results on AD vs. HC and pMCI vs. sMCI Binary Classifications: In order to evaluate the performance of the proposed method, the network is trained on fused sagittal T1-weighted and FDG PET images for the AD vs. HC and pMCI vs. sMCI binary classification tasks individually. To assess the process of training of the proposed network, we plot

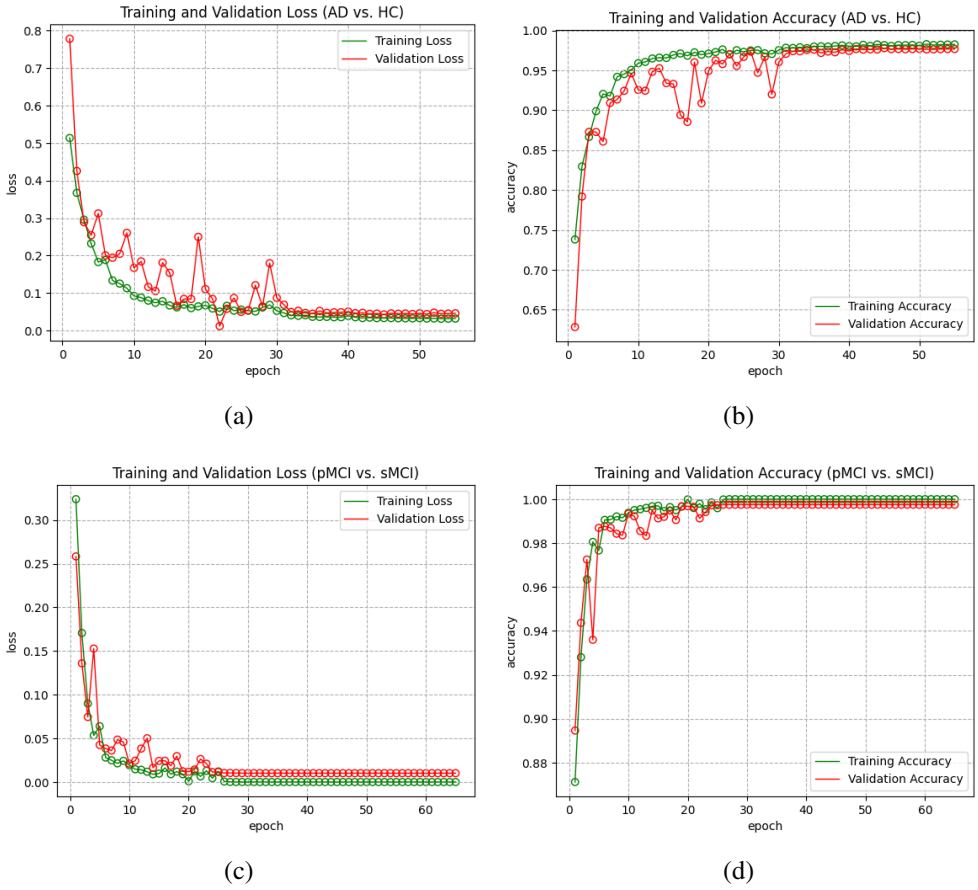
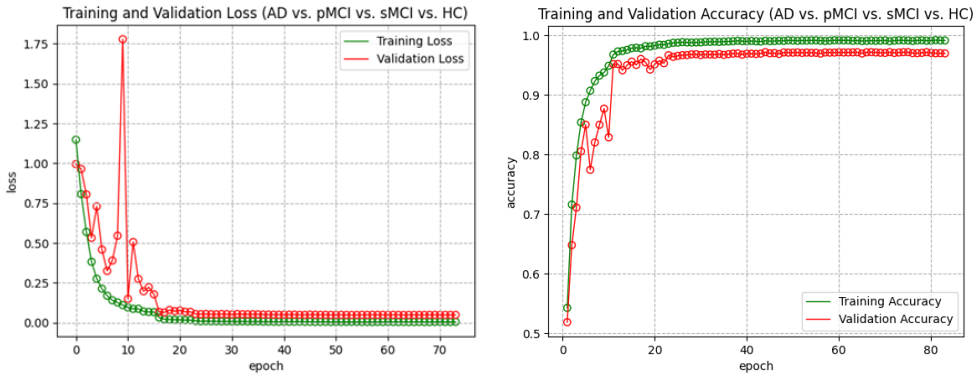


Figure 5.4: Training and validation performance curves on fused T1-weighted and FDG-PET images images for two binary classifications: (a) and (b) for AD vs. HC, and (c) and (d) for pMCI vs. sMCI.

Table 5.3: Performance Results, in percentage, of LHAttNet for AD vs. HC and pMCI vs. sMCI classifications on the fused T1-weighted MR and FDG-PET images

Methods	AD vs. HC				pMCI vs. sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
LHAttNet (Proposed)	98.05	96.38	98.59	98.62	98.83	97.49	97.71	98.65

curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 5.4(a) and 5.4(c) show the training and validation losses, for the AD vs. HC and pMCI vs. sMCI classifications, respectively. Figs. 5.4(b) and 5.4(d) show the corresponding training and validation accuracy curves, respectively. It is seen from Figs. 5.4(a) and 5.4(c) that the validation losses decrease and converge starting at the 33^{th} and 25^{th} epochs, respectively. The gap between the validation and training losses is very small. Figs.5.4(b) and 5.4(d) show the training and validation accuracy curves for the AD vs. HC and pMCI vs. sMCI classifications, respectively. It is seen from these figures that both the training and validation accuracies are high and reasonably close to each other. From the above observations, it is concluded that the two models are appropriately trained.



(a) (b)

Figure 5.5: Training and validation performance curves on fused T1-weighted MR and FDG-PET images for multiclass classification task (AD vs. sMCI vs. pMCI vs. HC): (a) Loss and (b) Accuracy.

To evaluate the performance of the trained models, we classify the sagittal fused T1-weighted MR and FDG-PET images annotated as AD and HC in the testing dataset (not used in the training or validation) as either AD or HC using the first trained model. Similarly, the images annotated as pMCI and sMCI in the testing dataset are classified using the second trained model. Table 5.3 gives the performance results in terms of the ACC, SEN, SPE, and AUC metrics provided by the proposed method for both the binary classification

tasks. It is seen from this table that for the AD vs. HC binary classification, the proposed LHAttNet provides classification accuracy of 98.05 %. For the pMCI vs. sMCI binary classification, the proposed LHAttNet scheme provides the accuracy of 98.62 % for the task of binary classification of Alzheimer's disease.

Table 5.4: Performance Results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC multiclass classification on fused T1-weighted MR and FDG-PET images from the ADNI dataset

Methods	**AD vs. pMCI vs. sMCI vs. HC**			
	ACC	**SEN**	**SPE**	**AUC**
LHAttNet (Proposed)	98.99	98.48	97.38	98.41

Results on AD vs. pMCI vs. sMCI vs. HC Multiclass Classification: We now consider the task of multiclass classification to classify a sagittal fused T1-weighted MR and FDG-PET images into one of the four classes, namely, AD, pMCI, sMCI, and HC. To assess the process of multiclass training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 5.5(a) and 5.5(b) show the training and validation losses and accuracies, respectively. It is seen from these figures that the validation loss decrease and converge starting at the 18^{th} epoch. The gap between the validation and training losses is very small. It is also seen from Fig. 5.5(b) that both the training and validation accuracies are high and reasonably close to each other. From the above observation, it is concluded that the model is appropriately trained.

To evaluate the performance of the model trained for multiclass classification, we classify the images in the test set of sagittal fused T1-weighted MR and FDG-PET images annotated as AD, pMCI, sMCI, or HC. Table 5.4 presents the performance results provided by the proposed scheme in terms of the ACC, SEN, SPE, and AUC metrics for the task of multiclass classification. It is to be noted from this table that the proposed scheme provides

a classification accuracy of 98.99 %.

Experiment 3: Classification Results on Fused T2-weighted MR and FDG-PET Images

In this experiment, we study the performance of our proposed LHAttNet using fused sagittal T2-weighted and FDG-PET images as input for binary classifications, AD vs. HC and pMCI vs. sMCI, as well as for multiclass classification AD vs. pMCI vs. sMCI vs. HC.

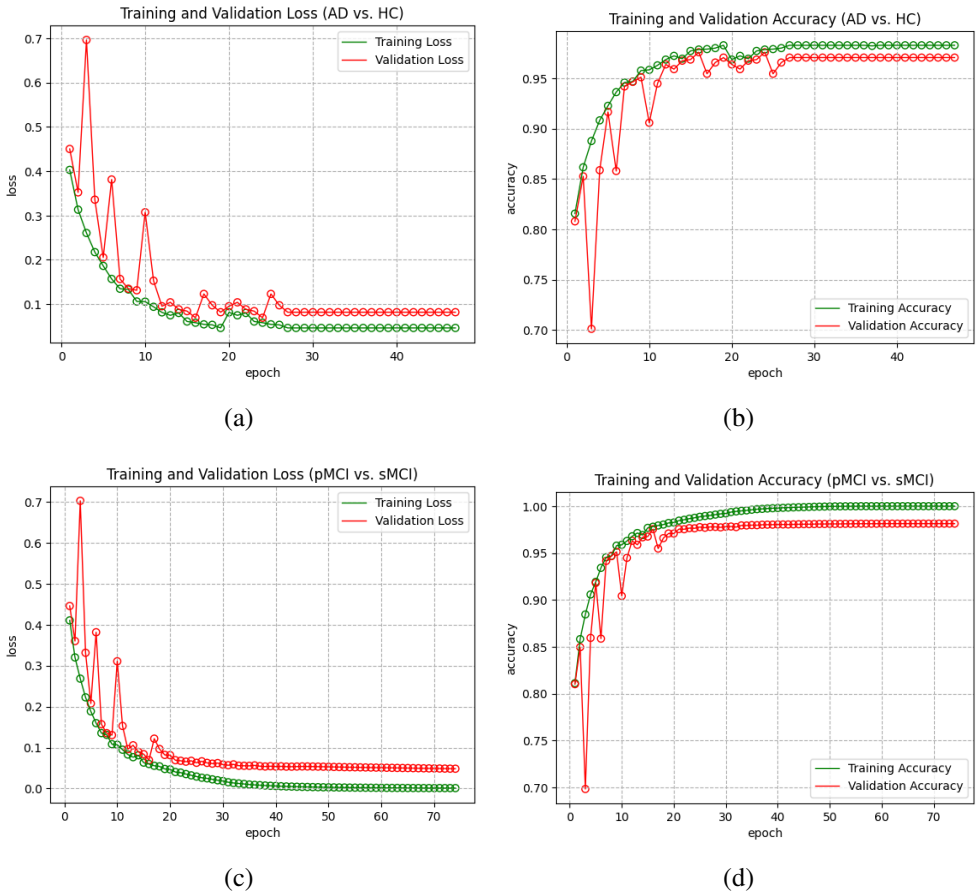


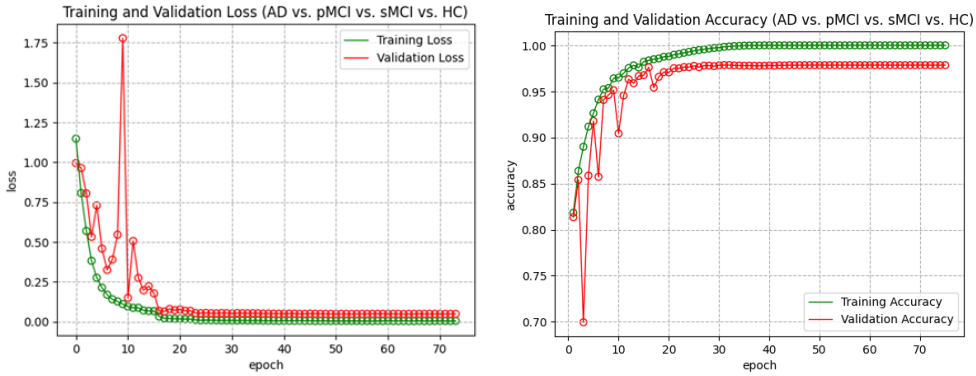
Figure 5.6: Training and validation performance curves on fused T2-weighted and FDG-PET images images for two binary classifications: (a) and (b) for AD vs. HC, and (c) and (d) for pMCI vs. sMCI.

Results on AD vs. HC and pMCI vs. sMCI Binary Classifications: In order to evaluate the performance of the proposed method, the network is trained on fused sagittal T2-weighted and FDG PET images for the AD vs. HC and pMCI vs. sMCI binary classification tasks individually. To assess the process of training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 5.6(a) and 5.6(c) show the training and validation losses, for the AD vs. HC and pMCI vs. sMCI classifications, respectively. Figs. 5.6(b) and 5.6(d) show the corresponding training and validation accuracy curves, respectively. It is seen from Figs. 5.6(a) and 5.6(c) that the validation losses decrease and converge starting at the 33^{th} and 25^{th} epochs, respectively. The gap between the validation and training losses is very small. Figs.5.6(b) and 5.6(d) show the training and validation accuracy curves for the AD vs. HC and pMCI vs. sMCI classifications, respectively. It is seen from these figures that both the training and validation accuracies are high and reasonably close to each other. From the above observations, it is concluded that the two models are appropriately trained.

To evaluate the performance of the trained models, we classify the sagittal fused T2-weighted MR and FDG-PET images annotated as AD and HC in the testing dataset (not used in the training or validation) as either AD or HC using the first trained model. Similarly, the images annotated as pMCI and sMCI in the testing dataset are classified using the second trained model. Table 5.5 gives the performance results in terms of the ACC, SEN, SPE, and AUC metrics provided by the proposed method for both binary classification tasks. It is seen from this table that for the AD vs. HC binary classification, the

Table 5.5: Performance Results, in percentage, of LHAttNet for AD vs. HC and pMCI vs. sMCI classifications on the fused T2-weighted MR and FDG-PET images

Methods	AD vs. HC				pMCI vs. sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
LHAttNet (Proposed)	97.76	95.83	95.54	96.82	95.32	93.48	95.62	93.12



(a) (b)

Figure 5.7: Training and validation performance curves on fused T2-weighted MR and FDG-PET images for multiclass classification task (AD vs. sMCI vs. pMCI vs. HC): (a) Loss and (b) Accuracy.

proposed LHAttNet provides a classification accuracy of 97.76 %. For the pMCI vs. sMCI binary classification, the proposed LHAttNet scheme provides an accuracy of 95.32 % for the task of binary classification of Alzheimer's disease.

Table 5.6: Performance Results, in percentage, of LHAttNet for AD vs. pMCI vs. sMCI vs. HC multiclass classification on fused T2-weighted MR and FDG-PET images from the ADNI dataset

Methods	**AD vs. pMCI vs. sMCI vs. HC**			
	ACC	**SEN**	**SPE**	**AUC**
LHAttNet (Proposed)	96.65	94.53	95.13	94.36

Results on AD vs. pMCI vs. sMCI vs. HC Multiclass Classification: We now consider the task of multiclass classification to classify a sagittal fused T2-weighted MR and FDG-PET images into one of the four classes, namely, AD, pMCI, sMCI, and HC. To assess the process of multiclass training of the proposed network, we plot curves showing the training and validation losses and accuracies as functions of the number of epochs. Figs. 5.7(a) and 5.7(b) show the training and validation losses and accuracies, respectively. It is seen from

these figures that the validation loss decreases and converges starting at the 18^{th} epoch. The gap between the validation and training losses is very small. It is also seen from Fig. 5.7(b) that both the training and validation accuracies are high and reasonably close to each other. From the above observation, it is concluded that the model is appropriately trained.

To evaluate the performance of the model trained for multiclass classification, we classify the images in the test set of sagittal fused T2-weighted MR and FDG-PET images annotated as AD, pMCI, sMCI, or HC. Table 5.6 presents the performance results provided by the proposed scheme in terms of the ACC, SEN, SPE, and AUC metrics for the task of multiclass classification. It is to be noted from this table that the proposed scheme provides a classification accuracy of 96.65 %.

5.4 Comparison of Different Input Modalities

In the previous section, we have studied the performance of the proposed LHAttNet on fused images generated through the multimodal neuroimaging fusion algorithm developed in Chapter 3 for binary and multiclass classifications of Alzheimer's disease. This section compares the performance of LHAttNet on single modality and fused images.

Table 5.7 presents the performance results for six different input modalities in binary classification tasks, distinguishing between Alzheimer's disease (AD) versus healthy controls (HC), and progressive mild cognitive impairment (pMCI) versus stable mild cognitive impairment (sMCI). It is seen from the table that the proposed scheme on fused T1-weighted and FDG-PET images provides superior results in terms of all the metrics compared to that on fused T1-weighted and T2-weighted MR images or fused T2-weighted and FDG PET images or any other individual modalities for both binary classification tasks.

Table 5.7: Performance results in percentage of different input modalities for binary classification tasks

Input Modality	**AD vs. HC**				**pMCI vs. sMCI**			
	ACC	**SEN**	**SPE**	**AUC**	**ACC**	**SEN**	**SPE**	**AUC**
T1-weighted MR Images	97.43	93.21	96.34	96.62	96.55	94.43	92.32	95.68
T2-weighted MR Images	93.81	90.25	93.63	93.70	91.78	90.54	93.44	90.11
FDG-PET Images	96.51	94.38	95.42	96.70	94.77	91.38	94.15	92.80
Fused T1-weighted and T2-weighted MR Images	97.65	94.55	96.55	96.85	96.80	94.75	96.55	96.05
Fused T1-weighted MR and FDG-PET Images	**98.05**	**96.38**	**98.59**	**98.62**	**98.83**	**97.49**	**97.71**	**98.65**
Fused T2-weighted MR and FDG-PET Images	97.76	95.83	95.54	96.82	95.32	93.48	95.62	93.12

Table 5.8 presents the performance results for six different input modalities in multiclass classification tasks. It is seen from the table that the proposed scheme using fused T1-weighted and FDG-PET images outperforms all other modalities across all metrics, except for sensitivity when using T1-weighted MR images as input for the multiclass classification task, where the increase is 1.5 %. It is observed that fusing two different modalities of the same types of details such as structural from T1-weighted and T2-weighted MR images does not offer improvement over individual modalities, on the other hand, details of different types such as structural and functional provides improvement on the multiclass classification of Alzheimer's disease.

Table 5.8: Performance results in percentage of different input modalities for multiclass classification tasks

Input Modality	**ACC**	**SEN**	**SPE**	**AUC**
T1-weighted MR Images	98.48	96.37	**98.82**	97.59
T2-weighted MR Images	85.70	89.58	91.83	87.73
FDG-PET Images	95.49	92.51	95.89	94.54
Fusion of T1-weighted and T2-weighted MR Images	97.96	94.65	96.15	93.86
Fusion of T1-weighted MR and FDG-PET Images	**98.99**	**98.48**	97.38	**98.41**
Fusion of T2-weighted MR and FDG-PET Images	96.65	94.53	95.13	94.36

5.5 Summary

In this chapter, the effectiveness of our proposed neuroimaging fusion method has been investigated for the classification of Alzheimer's disease using our proposed deep lightweight deep convolutional network, LHAttNet. It is observed that fusing two different modalities of the same types of details such as structural from T1-weighted and T2-weighted MR images does not offer improvement over individual modalities, on the other hand, details of different types such as structural and functional provides a significant improvement on the multiclass classification of Alzheimer's disease. The shows the importance of incorporating different types of information, such as structural and functional information, for the classification of Alzheimer's disease.

Chapter 6

Conclusion

6.1 Concluding Remarks

Fusion of medical imaging data obtained from different modalities is crucial for a better understanding of diseases. A fused image can reveal enhanced details that might not be visible with a single imaging modality. Traditional model-based fusion approaches have leveraged multiscale and multidirectional transforms because of their capabilities for superior data representation, energy compaction, and reduced complexity. However, the generalized Gaussian distribution traditionally used as a prior to model the transform domain coefficients does not provide an accurate fit to the empirical distribution, which exhibits heavier tails. Moreover, in order to enhance the energy level of the fused image, each of its transform-domain coefficients is replaced by the corresponding coefficient with the largest magnitude among the constituent images. Since increasing the energy level is key to enhancing the quality of a fused image, it is imperative to investigate other fusion rules that could further elevate the energy level of a fused image.

One of the most common brain diseases is Alzheimer's disease affecting millions of people around the world. The detection and classification of the disease are crucial for the care of the patient. The recent advancements in convolutional neural networks (CNNs)

have leveraged the classification of Alzheimer's disease because of their automatic feature generation capability. However, most of the existing CNN-based methods often disregard the local features of the brain data, which leads to a loss of subtle fine-grained features in the brain imaging data. The existing CNN architectures, which mainly rely on global features, do not pay much attention to the discriminability of the extracted features for the classification of Alzheimer's disease. Moreover, the existing architectures often end up using a large number of parameters to enhance the richness of the extracted features.

The objective of this thesis has been to develop a multimodal neuroimaging fusion algorithm using a novel energy maximization fusion rule and to design a lightweight deep CNN capable of extracting both local and global contextual features for improved Alzheimer's disease classification performance. This objective has been achieved by carrying out the work of this thesis in two parts.

In the first part of the thesis, a novel method for multimodal neuroimaging fusion has been proposed. The proposed method has employed NSST transform of the images whose coefficients are first modeled using the student's t location-scale distribution. It has been shown that this distribution results in a better approximation of the empirical NSST coefficients than the traditional generalized Gaussian distribution does. A Bayesian MAP estimator has been devised by using this distribution as a prior distribution to obtain the noise-free coefficients of the images. A new fusion rule has been proposed and used to fuse the obtained high-frequency coefficients of the constituent images. The main idea in proposing the new fusion rule is that the energy level of the fused image can be enhanced more effectively if the real and imaginary parts of the coefficients are treated individually instead of together, as is done in the traditional fusion rules. Extensive experiments have been carried out using both synthetic and real brain data to assess the performance of the proposed method for image fusion. It has been shown that the proposed method is able to preserve the structural and functional details and that it significantly outperforms the other

traditional methods in terms of the various performance metrics, peak signal-to-noise ratio, structural similarity, mutual information, and entropy. The proposed method is completely general in that it can be applied to any number of neuroimages.

In the second part of the thesis, a novel lightweight deep CNN that extracts local and global contextual features for the task of classification of Alzheimer's disease has been developed. A new module has been proposed and used in this network to process separately the local and global feature maps of input data to produce contextual and more relevant corresponding features. The features resulting from the dual contextual attention augmented convolution modules operating on the local and the global features are fused and used for the classification of Alzheimer's disease. It is to be noted that the proposed network employs only a small fraction of the number of parameters used by the other networks. Experiments have been performed to demonstrate the effectiveness of the proposed network for the task of binary and multiclass classifications using single modality T1-weighted structural MR, T2-weighted structural MR, and 18F-FDG-PET images taken from the ADNI database. The experimental results have shown that the performance of the proposed network is generally superior to that provided by the state-of-the art methods.

Finally, the effectiveness of the neuroimaging fusion on the classification of Alzheimer's disease has also been investigated. This investigation has been carried out using the proposed fusion scheme and the proposed CNN for the task of binary and multiclass classifications of Alzheimer's disease using fused T1-weighted and T2-weighted MR images, fused T1-weighted MR and FDG-PET images, and fused T2-weighted MR and FDG-PET images. From the experimental results, it has been observed that fusing two different modalities of different types of features such as structural and functional generally provides the best results for the classification of Alzheimer's disease.

6.2 Scope for Further Investigations

In the first part of the thesis, a neuroimaging fusion method has been developed developed for fusing brain imaging data obtained from different modalities such as MRI and PET. Future research could explore the application of the proposed method to other neuroimaging modalities, such as functional MRI and diffusion tensor imaging. Investigating how the proposed fusion method performs across different modalities could enhance its utility in neuroimaging analysis. The proposed fusion rule can also be further investigated by applying it to transform domain approaches in other fusion applications.

In the second part of the thesis, a deep convolutional neural network (CNN) extracting local and global contextual features has been developed for the classification of Alzheimer's disease into four different categories, namely, Alzheimer's disease (AD), progressive mild cognitive impairment (pMCI), stable mild cognitive impairment (sMCI), and healthy control (HC). The deep CNN developed for Alzheimer's disease classification can be used for the classification of other diseases such as Parkinson's disease, Huntington's disease, and frontotemporal dementia. Further evaluation of the proposed deep CNN architecture on larger and different datasets could also provide increased classification performance and generalizability.

Finally, in this thesis, a fusion-combined CNN framework has been developed to investigate the impact of fused images obtained using our proposed fusion method on the classification of Alzheimer's disease using our proposed CNN architecture. The fusion-combined CNN framework can be applied to different classification tasks across different domains such as other medical imaging diagnosis tasks beyond Alzheimer's disease classification using multimodal neuroimaging data.

References

[1] "2023 Alzheimer's disease facts and figures," vol. 19, no. 4, pp. 1598–1695, Mar. 2023.

[2] Y.L. Rao, B. Ganaraja, B.V. Murlimanju, T. Joy, A. Krishnamurthy, and A. Agrawal, "Hippocampus and its involvement in Alzheimer's disease: A review," *3 Biotech*, vol. 12, no. 2, p. 55, Feb. 2022.

[3] S.M. Nestor, R. Rupsingh, M. Borrie, *et al.*, "Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," *Brain*, vol. 131, no. 9, pp. 2443–2454, Sep. 2008.

[4] M. Park and W.J. Moon, "Structural MR imaging in the diagnosis of Alzheimer's disease and other neurodegenerative dementia: Current imaging approach and future perspectives," *Korean Journal of Radiology*, vol. 17, no. 6, pp. 827–845, Nov. 2016.

[5] A. Calvi, L. Haider, F. Prados, C. Tur, D. Chard, and F. Barkhof, "In vivo imaging of chronic active lesions in multiple sclerosis," *Multiple Sclerosis Journal*, vol. 28, no. 5, pp. 683–690, Apr. 2022.

[6] K. Misquitta, M. Dadar, D.L. Collins, M.C. Tartaglia, Alzheimer's Disease Neuroimaging Initiative, *et al.*, "White matter hyperintensities and neuropsychiatric

symptoms in mild cognitive impairment and Alzheimer's disease," *NeuroImage: Clinical*, vol. 28, p. 102 367, Jan. 2020.

[7] M.A. Oghabian, S.A.H. Batouli, M. Norouzian, M. Ziaei, and H. Sikaroodi, "Using functional magnetic resonance imaging to differentiate between healthy aging subjects, mild cognitive impairment, and Alzheimer's patients," *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, vol. 15, no. 2, p. 84, Mar. 2010.

[8] D.M. Wilson, M.R. Cookson, L. Van Den Bosch, H. Zetterberg, D.M. Holtzman, and I. Dewachter, "Hallmarks of neurodegenerative diseases," *Cell*, vol. 186, no. 4, pp. 693–714, Feb. 2023.

[9] G. Musa, A. Slachevsky, C. Muñoz-Neira, *et al.*, "Alzheimer's disease or behavioral variant frontotemporal dementia? Review of key points toward an accurate clinical and neuropsychological diagnosis," *Journal of Alzheimer's Disease*, vol. 73, no. 3, pp. 833–848, Jan. 2020.

[10] R.T. Vieira and L. Caixeta, "Subcortical atrophy in frontotemporal dementia and Alzheimer's disease: Significance for differential diagnosis and correlation with clinical manifestations," *Dementia & Neuropsychologia*, vol. 2, pp. 284–288, Oct. 2008.

[11] R.A. Armstrong, P.L. Lantos, and N.J. Cairns, "Overlap between neurodegenerative disorders," *Neuropathology*, vol. 25, no. 2, pp. 111–124, Jun. 2005.

[12] Chelsea Ekstrand, "Neuroimaging," in *Oxford Research Encyclopedia of Psychology*, Dec. 2022.

[13] S. Das and M.K. Kundu, "A neuro-fuzzy approach for medical image fusion," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3347–3353, Dec. 2013.

[14] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Trans. Instrum. Meas.*, no. 99, pp. 1–16, Jun. 2018.

[15] J. Du, W. Li, and B. Xiao, "Anatomical-functional image fusion by information of interest in local Laplacian filtering domain," *IEEE Trans. Image Processing*, vol. 26, no. 12, pp. 5855–5866, Dec. 2017.

[16] D.P. Bavirisetti, V. Kollu, X. Gang, and R. Dhuli, "Fusion of MRI and CT images using guided image filter and image statistics," *International Journal of Imaging Systems and Technology*, vol. 27, no. 3, pp. 227–237, Sep. 2017.

[17] Shuqin Liu, "Study on medical image enhancement based on wavelet transform fusion algorithm," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 2, pp. 388–392, Apr. 2017.

[18] W. Zhao and H. Lu, "Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2283–2294, Sep. 2017.

[19] J.J. Zong and T.S. Qiu, "Medical image fusion based on sparse representation of classified image patches," *Biomedical Signal Processing and Control*, vol. 34, pp. 195–205, Apr. 2017.

[20] V.D. Calhoun and J. Sui, "Multimodal fusion of brain imaging data: A key to finding the missing links in complex mental illness," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 1, no. 3, pp. 230–244, May 2016.

[21] Y. Liu, X. Chen, R.K. Ward, and Z.J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Oct. 2016.

[22] M. Manchanda and R. Sharma, "A novel method of multimodal medical image fusion using fuzzy transform," *Journal of Visual Communication and Image Representation*, vol. 40, pp. 197–217, Oct. 2016.

[23] J. Du, W. Li, B. Xiao, and Q. Nawaz, "Union Laplacian pyramid with multiple features for medical image fusion," *Neurocomputing*, vol. 194, pp. 326–339, Jun. 2016.

[24] J. Du, W. Li, K. Lu, and B. Xiao, "An overview of multi-modal medical image fusion," *Neurocomputing*, vol. 215, pp. 3–20, Nov. 2016.

[25] V. Bhateja, H. Patel, A. Krishn, A. Sahu, and A. Lay-Ekuakille, "Multimodal medical image sensor fusion framework using cascade of wavelet and contourlet transform domains," *IEEE Sensors J.*, vol. 15, no. 12, pp. 6783–6790, Aug. 2015.

[26] S. Singh, D. Gupta, R.S. Anand, and V. Kumar, "Nonsubsampled shearlet based CT and MR medical image fusion using biologically inspired spiking neural network," *Biomedical Signal Processing and Control*, vol. 18, pp. 91–101, Apr. 2015.

[27] L. Wang, B. Li, and L. Tian, "Multimodal medical volumetric data fusion using 3-D discrete shearlet transform and global-to-local rule," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 1, pp. 197–206, Jan. 2014.

[28] R. Shen, I. Cheng, and A. Basu, "Cross-scale coefficient selection for volumetric medical image fusion," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1069–1079, Apr. 2013.

[29] G. Bhatnagar, Q.M.J. Wu, and Z. Liu, "Directive contrast based multimodal medical image fusion in NSCT domain," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1014–1024, Aug. 2013.

[30] F.E. Ali, I.M. El-Dokany, A.A. Saad, and F.E. Abd El-Samie, "A curvelet transform approach for the fusion of MR and CT images," *Journal of Modern Optics*, vol. 57, no. 4, pp. 273–286, Feb. 2010.

[31] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Information Fusion*, vol. 33, pp. 100–112, Jan. 2017.

[32] M. Kumar and S. Dass, "A total variation-based algorithm for pixel-level image fusion," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2137–2143, Sep. 2009.

[33] H. Li, Z. Yu, and C. Mao, "Fractional differential and variational method for image fusion and super-resolution," *Neurocomputing*, vol. 171, pp. 138–148, Jan. 2016.

[34] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, pp. 147–164, Jul. 2015.

[35] A. Loza, D. Bull, N. Canagarajah, and A. Achim, "Non-Gaussian model-based fusion of noisy images in the wavelet domain," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 54–65, Jan. 2010.

[36] P.A. Hagargi and D.C. Shubhangi, "Brain tumor MR image fusion using most dominant features extraction from wavelet and curvelet transforms," *Brain*, vol. 5, no. 05, May 2018.

[37] S. Paris, S. W. Hasinoff, and J. Kautz, "Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid," *Communications of the ACM*, vol. 58, no. 3, pp. 81–91, Jul. 2015.

[38] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, Sep. 2006.

[39] G. Kutyniok and D. Labate, "Introduction to shearlets," *Shearlets*, pp. 1–38, Jan. 2012.

[40] D. D-Y Po and M.N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. Image Processing*, vol. 15, no. 6, pp. 1610–1620, May 2006.

[41] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, Jul. 2008.

[42] R.R. Coifman and D.L. Donoho, "Translation-invariant de-noising," *Wavelets and statistics*, vol. 103, pp. 125–150, May 1995.

[43] W.-Q. Lim, "The discrete shearlet transform: A new directional transform and compactly supported shearlet frames," *IEEE Trans. Image Processing*, vol. 19, no. 5, pp. 1166–1180, Jan. 2010.

[44] J.M. Fadili and L. Boubchir, "Analytical form for a bayesian wavelet estimator of images using the Bessel K form densities," *IEEE Trans. Image Processing*, vol. 14, no. 2, pp. 231–240, Jan. 2005.

[45] J. Zhao, R. Laganiere, and Z. Liu, "Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement," *International Journal of Innovative Computing, Information and Control*, vol. 3, no. 6, pp. 1433–1447, Dec. 2007.

[46] L. Alzubaidi, J. Zhang, A.J. Humaidi, *et al.*, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 53, pp. 1–74, Mar. 2021.

[47] A. Gupta, M. Ayhan, and A. Maida, "Natural image bases to represent neuroimaging data," in *Proc. Int. Conf. Mach. Learn.*, Atlanta, GA, USA, May 2013, pp. 987–994.

[48] B. Cheng, M. Liu, D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Robust multi-label transfer feature learning for early diagnosis of Alzheimer's disease," *Brain Imaging Behav.*, vol. 13, no. 1, pp. 138–153, Feb. 2019.

[49] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," in *Proc. Int. Conf. Pattern Recognit. App. Methods (2)*, Lisbon, Portugal, Feb. 2015, pp. 355–362.

[50] E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network," in *Proc. IEEE Int. Conf. Image Process.*, Phoenix, AZ, USA, Sep. 2016, pp. 126–130.

[51] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Melbourne, VIC, Australia, Apr. 2017, pp. 835–838.

[52] R. Cui and M. Liu, "Hippocampus analysis by combination of 3-D densenet and shapes for Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2099–2107, Sep. 2019.

[53] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[54] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Med. Imaging*, vol. 40, no. 9, pp. 2354–2366, Sep. 2021.

[55] X. Zhang, L. Han, W. Zhu, L. Sun, and D. Zhang, "An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's

disease diagnosis using structural MRI," *IEEE J Biomed Health Inform.*, vol. 26, no. 11, pp. 5289–5297, Nov. 2022.

[56] A. Ng, "Sparse autoencoder," vol. 72, pp. 1–19, Jan. 2011.

[57] C. Li, N.M.A. Elsayed Bakheet, W. Huang, and S. Wang, "Machine learning for automatic Alzheimer's disease detection: Addressing domain shift issues for building robust models," *Radiology Science*, vol. 2, no. 1, pp. 10–21, Mar. 2023.

[58] G.B. Karas, P. Scheltens, S.A. Rombouts, *et al.*, "Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease," *Neuroimage*, vol. 23, no. 2, pp. 708–716, Aug. 2004.

[59] J. Ashburner and K.J. Friston, "Voxel-based morphometry—the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, Jun. 2000.

[60] A.L. Da Cunha, J. Zhou, and M.N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.

[61] K. Guo, D. Labate, W.-Q. Lim, G. Weiss, and E. Wilson, "Wavelets with composite dilations and their MRA properties," *Applied and Computational Harmonic Analysis*, vol. 20, no. 2, pp. 202–236, Mar. 2006.

[62] K. Guo, D. Labate, and W.-Q. Lim, "Edge analysis and identification using the continuous shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 27, no. 1, pp. 24–46, Jul. 2009.

[63] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," *Optics & Photonics*, vol. 2005, 59140U–59140U, Aug. 2005.

[64] L. Wang, B. Li, and L. Tian, "Multimodal medical volumetric data fusion using 3-d discrete shearlet transform and global-to-local rule," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 1, pp. 197–206, Jan. 2014.

[65] Haitao Yin, "Tensor sparse representation for 3-D medical image fusion using weighted average rule," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2622–2633, Nov. 2018.

[66] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques—an introduction, review and comparison," *ISPRS Journal of photogrammetry and Remote Sensing*, vol. 62, no. 4, pp. 249–263, Sep. 2007.

[67] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[68] R.C. Gonzalez and R.E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006, ISBN: 013168728X.

[69] C.S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

[70] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "Statistical modeling of multimodal neuroimaging data in non-subsampled shearlet domain using the student's t location-scale distribution," in *Proc. 2017 IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Baltimore, MD, USA, May 2017, pp. 1–4.

[71] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "Multimodal neuroimaging fusion in nonsubsampled shearlet domain using location-scale distribution by maximizing the high frequency subband energy," *IEEE Access*, vol. 7, pp. 97 865–97 886, Jul. 2019.

[72] Min Xu, *Image registration and image fusion: Algorithms and performance bounds*. Syracuse University, 2009.

[73] L. Sendur and I.W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 438–441, Dec. 2002.

[74] Simon Jackman, *Bayesian analysis for the social sciences*. John Wiley & Sons, 2009.

[75] K.L. Lange, R.J.A. Little, and J.M.G. Taylor, "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, Dec. 1989.

[76] D.M. Endres and J.E. Schindelin, "A new metric for probability distributions," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1858–1860, Jul. 2003.

[77] Tania Stathaki, *Image fusion: algorithms and applications*. Academic Press, 2011.

[78] J. Du, W. Li, and B. Xiao, "Anatomical-functional image fusion by information of interest in local Laplacian filtering domain," *IEEE Trans. Image Processing*, vol. 26, no. 12, pp. 5855–5866, Dec. 2017.

[79] R.K.S. Kwan, A.C. Evans, and G.B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, Nov. 1999.

[80] C.R. Jack, M.A. Bernstein, N.C. Fox, *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods [online]," *Journal of magnetic resonance imaging*, vol. 27, no. 4, pp. 685–691, Apr. 2008.

[81] K.A. Johnson and J.A. Becker, "The whole brain atlas [online]," Available: http://www.med.harvard.edu/aanlib/ (accessed May. 11, 2018).

[82] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "Deep structural and clinical feature learning for semi-supervised multiclass prediction of Alzheimer's disease," in *Proc. 2018 IEEE 61st Int. Midwest Symp. on Circuits and Systems (MWSCAS)*, Windsor, ON, Canada, Aug. 2018, pp. 791–794.

[83] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "Missing structural and clinical features imputation for semi-supervised Alzheimer's disease classification using stacked sparse autoencoder," in *Proc. 2018 IEEE Biomedical Circuits and Systems Conf. (BioCAS)*, Cleveland, OH, USA, Oct. 2018, pp. 1–4.

[84] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "Hybrid feature fusion using RNN and pre-trained CNN for classification of Alzheimer's disease (poster)," in *Proc. 2019 22nd Int. Conf. on Information Fusion (FUSION)*, Ottawa, ON, CA, Jul. 2019, pp. 1–4.

[85] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "Classification of Alzheimer's disease from MRI data using an ensemble of hybrid deep convolutional neural networks," in *Proc. 2019 IEEE 62nd Int. Midwest Symp. on Circuits and Systems (MWSCAS)*, Dallas, TX, USA, Aug. 2019, pp. 481–484.

[86] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "Classification of Alzheimer's disease from MRI data using a lightweight deep convolutional model," in *Proc. 2022 IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Austin, TX, USA, May 2022, pp. 1279–1283.

[87] E. Jabason, M.O. Ahmad, and M.N.S. Swamy, "A lightweight deep convolutional neural network extracting local and global contextual features for the classification of Alzheimer's disease using structural MRI*," submitted for publication.

[88] J. Mazziotta, A. Toga, A. Evans, *et al.*, "A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM),"

Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, vol. 356, no. 1412, pp. 1293–1322, Aug. 2001.

[89] S.M. Smith, M. Jenkinson, M.W. Woolrich, *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23, S208–S219, Jan. 2004.

[90] R. Sharma, T. Goel, M. Tanveer, and R. Murugan, "FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans," *Appl. Soft Comput.*, vol. 115, no. 108099, pp. 1–11, Jan. 2022.

[91] M. Abadi, A. Agarwal, P. Barham, *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: `https://www.tensorflow.org/`.

[92] R. Cui and M. Liu, "Hippocampus analysis by combination of 3-D densenet and shapes for Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2099–2107, Sep. 2019.

[93] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[94] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Med. Imaging*, vol. 40, no. 9, pp. 2354–2366, Sep. 2021.

[95] X. Zhang, L. Han, W. Zhu, L. Sun, and D. Zhang, "An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's

disease diagnosis using structural MRI," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 11, pp. 5289–5297, Nov. 2022.

[96] S. El-Sappagh, T. Abuhmed, S.M.R. Islam, and K.S. Kwak, "Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, pp. 197–215, Oct. 2020.

[97] S. Basheera and M.S. Ram, "Alzheimer's disease classification using Leung-Malik filtered bank features and weak classifier," *Int. J. Recent Technol. Eng*, vol. 8, pp. 1956–61, Sep. 2019.

[98] S. Basheera and M.S. Ram, "Convolution neural network–based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 5, pp. 974–986, Jan. 2019.

[99] M. Liu, D. Cheng, W. Yan, and Alzheimer's Disease Neuroimaging Initiative, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Frontiers in neuroinformatics*, vol. 12, p. 35, Jun. 2018.

[100] P. Zhou, R. Zeng, L. Yu, *et al.*, "Deep-learning radiomics for discrimination conversion of Alzheimer's disease in patients with mild cognitive impairment: A study based on 18F-FDG PET imaging," *Frontiers in Aging Neuroscience*, vol. 13, Oct. 2021.

[101] L. Teng, Y. Li, Y. Zhao, *et al.*, "Predicting MCI progression with FDG-PET and cognitive scores: A longitudinal study," *BMC neurology*, vol. 20, pp. 1–10, Apr. 2020.