IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 9, 2021, accepted March 24, 2021, date of publication April 12, 2021, date of current version April 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3072559

A Survey on Classification Algorithms of Brain Images in Alzheimer's Disease Based on Feature Extraction Techniques

RUHUL AMIN HAZARIKA[1], ARNAB KUMAR MAJI[1], SAMARENDRA NATH SUR[2], (Senior Member, IEEE), BABU SENA PAUL[3], (Member, IEEE), AND DEBDATTA KANDAR[1]

[1]Department of Information Technology, North Eastern Hill University, Shillong 793022, India
[2]Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Gangtok 737136, India
[3]Institute for Intelligent Systems, University of Johannesburg, Johannesburg 2006, South Africa

Corresponding author: Debdatta Kandar (kdebdatta@gmail.com)

ABSTRACT Alzheimer's disease (AD) is one of the most serious neurological disorders for elderly people. AD affected patient experiences severe memory loss. One of the main reasons for memory loss in AD patients is atrophy in the hippocampus, amygdala, etc. Due to the enormous growth of AD patients and the paucity of proper diagnostic tools, detection and classification of AD are considered as a challenging research area. Before a Cognitively normal (CN) person develops symptoms of AD, he may pass through an intermediate stage, commonly known as Mild Cognitive Impairment (MCI). MCI is having two stages, namely StableMCI (SMCI) and Progressive MCI (PMCI). In SMCI, a patient remains stable, whereas, in the case of PMCI, a person gradually develops few symptoms of AD. Several research works are in progress on the detection and classification of AD based on changes in the brain. In this paper, we have analyzed few existing state-of-art works for AD detection and classification, based on different feature extraction approaches. We have summarized the existing research articles with detailed observations. We have also compared the performance and research issues in each of the feature extraction mechanisms and observed that the AD classification using the wavelet transform-based feature extraction approaches might achieve convincing results.

INDEX TERMS Alzheimer's disease (AD), hippocampus, magnetic resonance imaging (MRI), mild cognitive impairment (MCI), progressive MCI (PMCI), stable MCI (SMCI)

I. INTRODUCTION

Alzheimer's disease (AD) is a neurological disorder that mainly destroys the memory cells in the human brain. In AD, patients experience symptoms like memory loss, visual changes, confusions, etc [1]. AD destroys the memory and thinking skills slowly and in the end, it kills the capability to carry out even the simplest tasks [1]. The effects of AD can be observed from the age of the early 60s. In 2019, ''National Institute on Aging, U.S.A'', has done a survey and found that around 6 million people from the U.S.A are affected by AD [2]. In a similar kind of report, the ''Alzheimer's and Dementia Resources'' has concluded that in India, more than 4 million people are suffering from AD [3]. Worldwide, the growth of AD patients is enormous and alarming.

MCI is a dementia stage, where a patient experiences more cognitive declination than a CN individual of the same ages [4]. Although people in MCI experience problems with languages, memories, thinking skills, etc., their symptoms are not as severe as those of AD. According to a research report, 8 out of 10 people with MCI develop AD within 7 years, whereas the probability of converting to AD from CN is 3 out of 10 [5]. Therefore, MCI is considered an early stage of AD. MCI is classified into two stages namely SMCI and PMCI. After following up the dementia stage for an MCI patient, over the years, if the dementia stage remains stable, or if it shows a very slow rate of progression to the AD, then the patients are said to be in SMCI dementia stage [6]. Some MCI subjects experience a high rate of cognitive declination over

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng.

TABLE 1. Summary of some of the commonly used AD diagnosis approaches introduced by several associations.

AD Diagnosis Approaches	Introduced by	Advantages	Limitations
Based on the score obtained from the Mini-Mental State Examination (MMSE), Short Blessed Test, Clinical Dementia Rating (CDR), Clinical Impression of Global Change (CIBIC), Alzheimer Disease Assessment Scale (ADAS), etc.	American Psychiatric Association	The easiest way for the diagnosis of AD. Less technical steps involved [40].	Time-consuming. May not provide accurate results all the time [40].
AD diagnosis based on the blood- based Biomarkers.	Alzheimer's Precision Medicine Initiative	A blood test is one of the most feasible approaches in world-wide settings. The process is less costly and less time consuming [41].	A variety of molecules such as, proteins, peptides, nucleic acids, lipids, metabolites etc. presents in the blood that can be seen in plasma, exosomes, cellular compartments etc. To study such a large number of components is a challenge for the researchers [42].
Aβ/amyloid- Tau/Neurofibrillary-Neurodegenerative (ATN) based AD diagnosis (the decision is made based on the amount of Aβ/amyloid and Tau deposits).	National Institute on Aging and Alzheimer's Association Research Framework for AD	This approach imposes a little financial burden to the patients which is easily accessible too. Moreover, this approach does not comprise acquaintance to the radioactivity [43].	The differentiation based on this biomarker is sometimes arbitrary as in many diseases other than AD, almost the same type of differentiation can be seen [44].
Transcranial Magnetic Stimulation (TMS)-based AD diagnosis approach (a non-invasive therapeutic method where the concept of a changing magnetic field is used. A paired-pulse TMS differentiates AD patients).	Alzheimer's Disease and Related Disorders Association	The process is non-invasive. In this approach the patients can continues with their regular routines [45].	The process is time-consuming (at an average 30 actions within 6 weeks). Patients may experience nervousness before and throughout the treatment.
Electroencephalography (EEG) based AD diagnosis approach (it archives the electrical action of nerve cells and hence ramblingly signifies fundamental intelligence function in brain).	American EEG Society	One major benefit is the capability to perceive brain activity as it discloses in real-time (in milliseconds) [46].	One major drawback is that it's difficult to determine the exact brain position from the electrical activity comes [47].
ELECTROVESTIBULOGRAPHY (EVestG) based AD diagnosis approach (This approach is based on a test that determines the field potential activity in the exterior ear channel in response to vestibular stimuli).	NeuralDx, Monash University, Clayton	In this approach, the signals are determined painlessly as well as non-invasively. The approach is cost-effective too [48].	The main disadvantage is that the signal is depending on the patient's physical condition also [49].

the ages, and after a few years they may progress to the stage of AD, which are known as the PMCI subjects [7].

A. CLASSIFICATION OF AD USING BRAIN IMAGES

Dr. Alois Alzheimer discovered AD in the year 1906 [8]. Since then, researchers have been trying to develop a mechanism that can detect AD accurately. Some of the popular approaches for AD diagnosis are described in table 1.

The manual classification of AD by the neurologist is time consuming and may not provide accurate results all the time. Many factors may affect the manual diagnosis process, such as the patient's age, nervousness, poor eyesight, etc. Researchers have been trying to determine the complicated changes in brain tissues that occur in early stages as well as during the progression of AD [9]. According to the research reports, changes in the brain tissues may begin much before a person develops AD, i.e., toxic changes occur in the brain before the symptoms of AD occurs [10], [11]. When a person develops AD, the brain experiences some unusual transformations of proteins from amyloid plaques and tau tangles [12], [13]. The process gradually makes the healthy neurons stop working and connection among all other neurons starts splitting [14].

The initial damage occurs in the areas of the brain which are responsible for forming memories, such as the Hippocampus, Entorhinal Cortex, Amygdala, etc. [15], [16] [17]. Gradually, more neurons die, which causes additional damage in other parts of the brain and leads to volumetric shrinkage. Structural imaging tools, such as MRI, Computed Tomography (CT), etc. can provide information about the shapes, positions, and volumes of brain tissues [18]. Hence, these brain imaging tools can be used for diagnosing the AD [19], [20]. By segmenting the affected tissues in the brain images, it is possible to observe the neurological changes that occur in AD, and the information can be used to train a classifier to classify AD more accurately [21], [22].

Challenges in the detection of AD: Diagnosing AD is a complex task. Based on the memory test, it is a challenge for the neurologist to decide whether a person is developing AD or not, because, significant memory loss (more or less) is common in normal aging too [23]. Hence, classification of AD, based on the bio-markers in the brain tissues is preferable. During the progression stage of AD, the most major affected brain regions are responsible for the cognitive operations [24]. The major challenges in classifying AD using brain images are proper segmentation of brain MRI's,

detection of Region of Interest (RoI), extraction of appropriate feature set, and comparison amongst the tissues of all the subject groups [25], [26]. Hence, a proper feature extraction technique is absolutely necessary for AD classification using brain images.

B. COMMONLY USED METHODOLOGY FOR CLASSIFICATION OF AD USING BRAIN IMAGES

For effective classification of any images, numerous steps are involved. The initial task for classification is pre-processing which includes removal of noise etc. Then appropriate methodology is invoked for effective feature extraction. The distinct features are selected for the appropriate classification and a common methodology for the same is shown in figure 1.

1) DATA COLLECTION

The first step towards the classification approach is to obtain a sufficient amount of brain images as well as the patient details for different subject groups, such as CN, MCI, and AD patients. The most common types of brain images used by the researchers, namely Structural MRI, Fluid-attenuated inversion recovery (FLAIR), Magnetization Prepared Rapid Acquisition Gradient Echo (MP-RAGE), T2 Weighted, Functional MRI (FMRI), Positron Emission Tomography (PET), etc. Apart from the brain images, some other relevant information, such as the patient's medical history, Mini-Mental State Exam (MMSE) score, genetic information, etc. are also may be required in the study of AD. The most commonly used publicly available online data sources for AD classification, namely Alzheimer's Disease Neuroimaging Initiative (ADNI) [27], Open Access Series of Imaging Studies (OASIS) [28], etc.

2) PRE-PROCESSING

In the perspective of image classification, pre-processing is the collection of operations to extract relevant information before processing to further computational process [29]. To get an accurate classification result, pre-processing is one of the most important steps [30]. The most commonly used pre-processing steps for AD classification include; image resizing, noise filtering, skull stripping, morphological operations, etc. The brain imaging mechanisms capture some unwanted pixels in the form of the skull, hence among all the pre-processing steps, skull stripping is considered one of the most essential steps [31].

Some of the most commonly used software's/toolboxes for medical image processing are: Diffeomorphic Anatomic Registration Through Exponentiated Lie (DARTEL) for image registration [32], Montreal Neurological Institute template (MNI) for affine registration [33], Statistical Parametric Mapping (SPM) for realignment, smoothing and spatial normalization [34], FreSurfer for volumetric feature extraction [35], Voxel-based morphometry (VBM) for bias-correction segmentation, morphological changes estimation, etc. [36], FMRIB Software Library (FSL) for segmentation [37], International Consortium for Brain

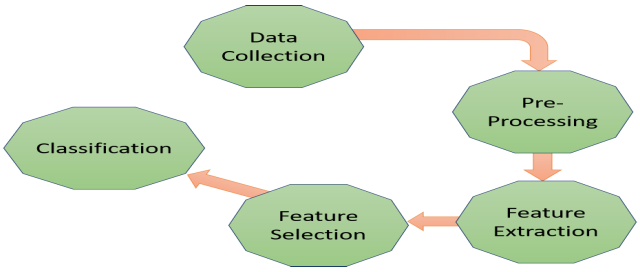

FIGURE 1. Block diagram of the methodology for classification of AD using brain images.

Mapping (ICBM) for extraction of the region of interests, etc [38]. Although VBM-8 can produce convincing results in various pre-processing steps, but sometimes it fails to determine the brain morphological changes accurately. Computational Anatomy Toolbox (CAT) is another toolbox for estimating the morphometric changes in the human brain. The combination of CAT-12 and the VBM8 toolboxes can provide more accurate results in estimating the morphological changes in the brain [39].

3) FEATURE EXTRACTION

Feature extraction (FE) is a dimensionality reduction technique, that proficiently characterizes the fascinating portions of input images as the feature vectors [50]. FE is the process of producing novel features from the existing ones and extracts the essential features that are useful for classifying an object. The newly compact features can recapitulate utmost information confined in the original set of features [51]. feature extraction helps the classification model for better training, reducing time complexity, and producing a better accuracy. The most useful features in medical images, namely colors, shapes, textures, etc. In some medical images, where color information is less explainable, the texture and shape-based feature extraction techniques [52] is used. The most commonly used feature extraction approaches in medical image processing include; Principle Components Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Locally Linear Embedding (LLE), t-distributed Stochastic Neighbor Embedding (t-SNE), Autoencoders, etc.

4) FEATURE SELECTION

Sometimes, feature extraction methods extract some less relevant features that force a model to learn falsely, and the accuracy of the final classification gets affected. Feature selection is the procedure to select only the most relevant features by eliminating the unwanted features by following a particular classification paradigm [53]. Some of the major advantages of using a proper feature selection approach, namely faster training of the algorithm, reducing the complexity & over-fitting issues of the model, improving the accuracy of the model, etc [54].

Feature selection can be done in two types of approaches, by using the supervised methods, or by using the

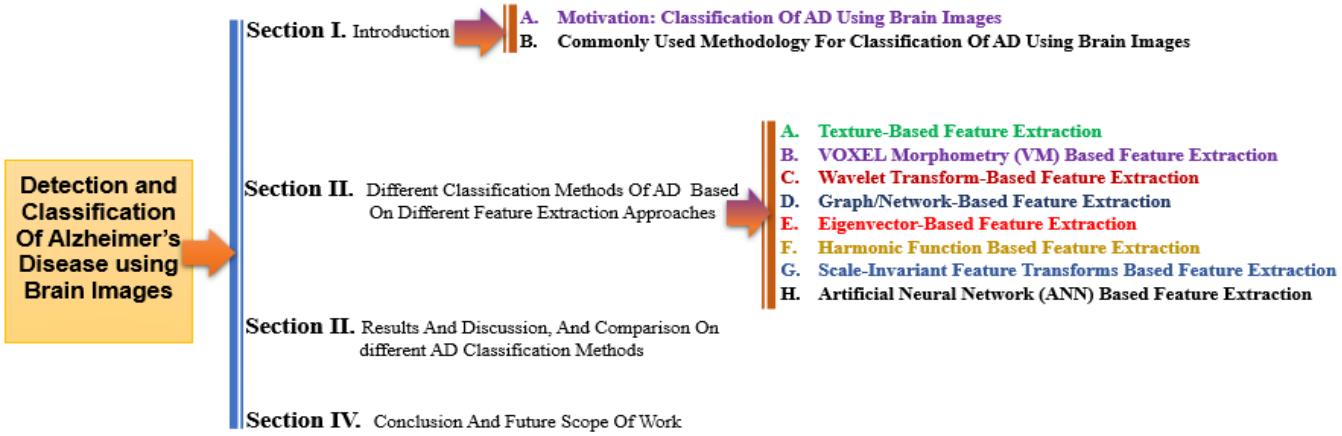

FIGURE 2. Organization of the paper.

unsupervised methods. An unsupervised method uses the concept of correlation and eliminates non-used features while ignoring the target variables. But in supervised methods, target variables are compared with the input variables to eliminate the irrelevant features [55]. Among all supervised feature selection methods, wrapper, filter, and intrinsic/embedded are the most commonly used methods in medical image processing. In wrapper methods, firstly a performance evaluation metric is created, and then the suitable algorithm is used to generate several models having different sets of input features. Finally, an effective feature is selected which contributes the most for exploring the best performing model [56]. One of the commonly used examples of wrapper feature selection methods is the Recursive Feature Elimination (RFE) method [57]. In filter methods, some statistical algorithms are used to estimate the correlation among the input and target variables, and then based on the correlation scores, the best variables are considered in the model [58]. Some of the commonly used filter feature selection methods are Pearson's Correlation, Linear Discriminant Analysis (LDA), etc. Embedded is a feature selection approach, where the model is trained by acquiring a large dataset, and selection of the most relevant features is carried out automatically [53]. Among all the most widely used embedded feature selection approaches, Least Absolute Shrinkage and Selection Operator (LASSO), and Decision Trees, are the most common methods.

The most commonly used feature selection methods include; SVM-REF (Support Vector Machine- Recursive Feature Elimination), Genetic algorithm based technique, Pearson's Correlation Coefficient (PCC), T-test Score (TS), Fisher Criterion (FC), Gini index (GI), Statistical Dependency (SD), Mutual Information (MI), Information Gain (IG), etc.

5) CLASSIFICATION

Image classification is a procedure to categorize a group of pixels, based on some protocols by using spectral or textural features. Classification methods can be categorized into two parts, namely supervised classification and unsupervised classification [59].

In supervised classification, the training data are selected visually, and then data are assigned to some pre-determined categories such as roads, buildings, etc. and after that, some statistical measures are created to apply in the entire image [60]. Using the sample data, for categorizing entire images, the two most commonly used methods, namely maximum likelihood, and minimum distance. On the other hand, unsupervised classification is an automatic procedure where training data is not used. In unsupervised classification, some particular properties of an image are thoroughly determined using an appropriate algorithm [61].

Some of the commonly used image classification techniques, namely Logistic Regression [62], Naïve Bayes [63], Stochastic Gradient Descent [64], K-Nearest Neighbours [65], Decision Tree [66], Random Forest [67], Support Vector Machine (SVM) [68] etc.

a: ORGANIZATION OF THE PAPER

The organization of this paper includes 4 sections and a total of 10 subsections. The organization is represented pictorially in figure 2.

II. SURVEY ON DIFFERENT CLASSIFICATION METHODS OF AD BASED ON DIFFERENT FEATURE EXTRACTION APPROACHES

Research is going on to match the finest feature extraction approach for classifying AD accurately. Several approaches have been proposed by the researchers. Some of the commonly used feature extraction approaches for AD classification are discussed below.

A. TEXTURE-BASED FEATURE EXTRACTION

The texture is a set of repetitive information in an image with a uniform interval [69]. Texture generally refers to a particular region (region of interest) of an image, that provides the same information such as shape, density, pixel value, etc [69]. The procedure to extract texture features from an image is called

texture feature extraction. Texture features play a major role in medical image processing [70]. Texture features help in finding discriminative features from a brain image for the classification of neurological disorders such as AD, MCI, etc. [71].

Gray-Level Co-occurrence Matrix (GLCM) is one of the most widely used texture feature extraction approaches proposed by Haralick et.al. [72]. GLCM determines the statistical features, according to the gray intensity values of a pixel. In the literature, [73], and [74], it is described that the texture features extraction by co-occurrence matrices provides a better result than other texture discrimination methods. One of the first approaches in texture analysis in the whole brain volume using GLCM is discussed in the literature [75]. In the literature [75], the authors used the GLCM based approach in Dopamine Active Transfer scan (DaTSCAN) brain images to determine the patterns that change due to the Parkinson's disease (PD), then used the achieved texture based information for final classification, and achieved a convincing result. Similarly, in the literature [76], the authors proposed a PD classification method based on the texture based feature extraction by using the GLCM approach. The authors have used the 123I-ioflupane imaging to achieve the best texture based features. Based on the features extracted by the GLCM approach, a SVM classifier is used which finally achieved an accuracy of 97.4% while classifying the PD. One of the major advantages of the co-occurrence matrix is that the co-occurring sets of pixels can be spatially related in different directions by taking reference of the distance and angular spatial relationships [77]. One of the major drawbacks of GLCM is that, it is a sparse matrix with many elements valued as zero, which is unnecessary for texture features calculation, hence it is computationally expensive [78], [79]. Some literature on AD classification using texture based feature extraction approaches and their research issues are discussed below.

In the articles [80], [81], the authors have proposed an approach of AD classification, where GLCM is used for feature extraction. In the literature, [80], Gabor filter is used along with the GLCM. After extracting a sufficient number of features, the authors have applied the SVM- RFE (Recursive Feature Elimination) method for selecting the most appropriate features. In the literature, [81], texture features, such as edge information, color, and boundary information, etc. are mined from the whole MR images, and clinical features such as Functional Activities Questionnaire (FAQ), Neuropsychiatric Inventory (NPI), Geriatric Depression Scale (GDS), etc. are mined from the Grey Matter(GM), White Matter(WM), and Cerebrospinal Fluid (CSF) segmented regions using the GLCM approach.

Krishnakumar Vaithinathana, *et al.* proposed a classification framework for the classification of AD, based on texture information extraction [77]. The voxels, which are selected as Region of Interests (RoIs), are mined and combined, and different textures are determined collectively. To select the best features, the authors have used 3 approaches, namely Fisher score, Elastic net regularization, and the SVM Recursive Feature Elimination (SVM-RFE) technique. For classification, the authors have used random forest, linear SVM, and the k-Nearest Neighbor (kNN).

Based on brain structural changes, and hippocampal shape, G. Wiselin Jiji, *et al.* proposed a novel method for detecting AD [82]. For extracting the features, the authors have identified the busy texture information. Busy textures are those for which there are rapid changes of intensities from one pixel to its neighbor. The spatial frequency of intensity alteration is very high. Therefore, suppression of contract aspect, from the information about the spatial rate of change in intensity, indicates the degree of business of a texture. The authors have used SVM for classification.

The performance comparison for different AD classification techniques using Texture-based feature extraction approaches is presented in table 2.

From table 2, it can be observed that, amongst several AD classification techniques using texture-based feature extraction approaches, the maximum classification performance is claimed by Altaf *et al.* [81] with an accuracy rate of 97.8%. But the overall highest average performance is acquired by Xiao *et al.* [80] with a rate of 93.96%.

B. VOXEL MORPHOMETRY (VM) BASED FEATURE EXTRACTION

For brain structure study, Morphometry analysis is one of the most common approaches that determine a comprehensive quantity of structural variances within or across the clusters in the whole brain [83]. Voxel-based morphometry (VBM) is a commonly used method for measuring the variances in local concentrations of brain cells, by performing a voxel-wise evaluation, taking reference from several brain images of the same group [84]. VBM can be applied to determine the volumetric changes, especially in Grey Matter (GM) regions among different subject groups such as AD, MCI, CN, etc [85], [86].

One of the first uses of the voxel-wise statistical test in medical image processing is proposed in the article [87]. In all the pre-processed SPECT images, a t-test based approach is used to determine the voxel-wise features, and finally, the model returns a total of 3816 no's of features for each subject. Based on the features extracted, the authors have used the SVM classifier and achieved a convincing result while classifying AD subjects.

In a similar work, I. Alvarez, *et al.* discussed a novel AD classification method in the literature [88]. The authors have used the concept of ensemble SVM classifiers trained on different parts of the brain images (SPECT and PET), such as the majority-voting, least squares estimation, as well as the double layer hierarchical combing, and then a pasting-votes technique-based approach is used to ensemble the classifiers. The authors claimed that the accuracy of the proposed method is 97.5% for SPECT and 100% for PET images.

Based on an SVM-based classifier, a novel AD classification approach is proposed in the literature [89]. The SPECT images are pre-processed using a filtered back

TABLE 2. Performance comparison of the AD classification techniques using texture-based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
Zhe Xiao, et al. [80]	AD vs. CN: 92.86%, MCI vs. CN: 97.22%, AD vs. MCI: 91.18%	AD vs. CN: 87.04%, MCI vs. CN: 95.23%, AD vs. MCI: 100%	AD vs. CN: 98.28%, MCI vs. CN: 100%, AD vs. MCI: 83.33%	AD vs. CN: 97.28%, MCI vs. CN: 100%, AD vs. MCI: 84.21%	AD vs. CN: 89.06%, MCI vs. CN: 93.75% AD vs. MCI: 100%	NA
Tooba Altaf, et al. [81]	AD vs. CN: 97.8%, AD vs. MCI: 85.3%, MCI vs. CN: 91.8%	AD vs. CN: 100%, AD vs. MCI: 75%, MCI vs. CN: 90%	AD vs. CN: 95.65%, AD vs. MCI: 94.29%, MCI vs. CN: 93.33%	NA	NA	NA
Krishnakumar Vaithinathana, et al. [77]	AD vs. CN: 87.39%, MCI vs. CN: 63.48%, AD vs. MCI: 63.16%, PMCI vs SMCI: 66.38%	NA	NA	NA	NA	NA
G. Wiselin Jiji, et al. [82]	NA	AD vs. CN: 80%	AD vs. CN: 60%	NA	NA	NA

projection (FBP) algorithm and a Butterworth noise removal filter. By using the SPM toolbox, all the images are spatially normalized to 95 × 69 × 79 voxel representation, where the voxels contain the grey level intensities. The most relevant features are selected by using a Fisher linear discriminant ratio based approach. For classification, the authors have used the SVM classifier with Radial Basis Function (RBF) kernel and achieved a convincing accuracy.

A novel CAD developed for detecting AD from the SPECT images is described in the literature [90]. The brain images are divided into different components as the chains of successive voxels in 3 orthogonal ways; coronal, axial, and sagittal, and then used as feature vectors in the SVM classifier. In order the select the most relevant set of voxels for the final classification, the authors have used the concept of Classification and Regression Trees. The performance evaluation shows that the combination of the SVM and the classification trees can produce a better classification result.

Some of the recently published literature, where AD classification is performed using VBM based feature extraction approaches, are discussed below.

In the literature [91], authors proposed an AD classification method using Voxel-based morphometry (VBM) based feature extraction to obtain the brain regions, where the grey matter volumes decreased significantly, and hence marked those regions as a 3D mask. The 3D masks are then applied in all the pre-processed images to extract the voxel values as raw-feature vectors.

For AD classification, in an article [92], authors have proposed a classification mechanism to identify AD vs CN subjects from the structural MR images, based on the Voxel-based morphometry (VBM) analysis. The authors have used Statistical Parametric Mapping (SPM)8, Voxel-based morphometry (VBM)8 toolbox, and Diffeomorphic Anatomical Registration using The Exponentiated Lie algebra (DARTEL) toolboxes with Voxel-based morphometry (VBM) for the enhancement correction of images. The authors have used only grey matter information in their study, which is further spatially smoothed with a Gaussian smoothing-based kernel. Further, the changes in grey matter volumes are determined by a voxel-based analysis. For identifying the Volume of Interests (VoIs), a 3D mask is generated from the atrophy regions, depending on the results of the Voxel-based morphometry (VBM) and DARTEL approaches. For extracting the regions, where grey matter volumes decreased, Voxel-based morphometry (VBM) analysis of each training data set is used through a 3D mask.

To classify AD vs CN, MCI vs CN, and MCI vs AD, from brain MR images, in literature [93], the hippocampal morphometry of CN, MCI, and the AD subjects are mainly focused. For extracting the hippocampus, the authors have used a 3D Automated Anatomical Labeling (AAL)-based approach, where 3D images are superimposed on the AAL, and the voxels which are categorized as hippocampal are selected. The authors have used 2D Circular Harmonic Functions (CHF) to select the contradicted patterns, slice by slice basis. For analyzing the reduction in the hippocampus volume, authors have counted the cerebrospinal fluid voxels in the hippocampal region and concluded that CN people have less cerebrospinal fluid than MCI and AD affected people.

A multi-atlas based classification approach for AD diagnosis, based on the morphometry features, is described in literature [94]. For obtaining the features on multiple atlases, the authors have performed a registration approach for spatial normalization followed by a quantification approach for morphometric measurements. Then a grey matter density map is extracted for feature representation from the brain images. The watershed segmentation technique on the correlation map is applied for selecting a set of Region of Interests (RoIs). For selecting the most relevant voxels, Pearson Correlation (PC) method is applied. Then all the neighboring voxels, for which there is no increment of the PC, are included iteratively. For each atlas space, 1,500 most discriminative Region of Interests (RoIs) features are designated as the representation of a subject. The authors proposed a View-Centralized Multi-Atlas (VCMA) approach with the help of the Accelerated Proximal Gradient (APG) method, for selecting the appropriate features from each atlas.

In the literature, [95], a novel classification technique for classifying AD, MCI, and CN subjects is proposed based on the morphometry feature analysis. The authors have used FMRIB Software Library (FSL) package to segment brain parts into 3 different tissues, namely grey matter, white matter, and cerebrospinal fluid. Image registration is performed to obtain subject-labeled images, based on a template consist of 93 manual labels. The grey matter tissue volume of each

region is computed and used as a feature, which is then aligned to its particular image using a rigid transformation. The average intensity of each Voxel of Interest is computed as another feature. For each subject, a total of 93 features are obtained from MRI and additional 93 features from a PET image. The authors have used the multi-task feature selection technique, which can preserve both the multi-modality correlation within the same subject and the relationship across modalities between different subjects. The Accelerated Proximal Gradient (APG) method is applied for obtaining the optimal solution of the proposed approach.

Using the structural MR images, a classification framework to classify AD vs CN subject is proposed in the literature [96]. For the Diffusion Tensor Imaging (DTI) images, the authors have used the Brain-Visa toolbox while performing the eddy current correction. Then diffusion tensors are determined and then Apparent Diffusion Coefficient (ADC) and Fractional Anisotropy (FA) maps are extracted. Using the Statistical Parametric Mapping (SPM)2 toolbox, the segmentation operation is performed to divide the structural images into 3 parts, namely grey matter, white matter, and cerebrospinal fluid. The ADC maps are also segmented into two parts namely cerebrospinal fluid, and the non-cerebrospinal fluid maps. The Fractional Anisotropy images are also segmented into two parts. They are white matter and non-white matter maps. Then the maps of DTI-grey matter are determined by the intersection of non-cerebrospinal fluid and non-white matter maps. The intersection of DTI-grey matter and structural grey matter map resulted in the final grey matter map. The authors proposed to calculate the common minimal brain volume. A binary mask is applied to all the normalized images for calculating their intersection. Then the common binary mask is mapped to the Automated Anatomical Labeling (AAL) to retain the Region of Interests (RoIs). Finally, 73 out of 90 Region of Interests (RoIs) are considered from the AAL. The mean diffusivity (mean ADC) is calculated from the Region of Interests (RoIs), and then the voxel-wise multimodal properties are obtained from the ADC and grey matter concentration ratio.

A multi-modality classification framework to classify AD vs CN, MCI vs CN, and Progressive MCI (PMCI) vs Stable MCI (SMCI) subjects is proposed in the literature [97], where the morphometry properties have been considered for feature extraction. The authors have proposed the framework to partition the subject images into 93 regions of interests (RoIs) with the help of atlas wrapping. From all the 93 Region of Interests (RoIs), the grey matter tissue volume is calculated. For the PET images, each image is rigidly aligned with its respective MR image, and for each Region of Interests (RoI), the average PET signal value is computed. Finally, for each subject, a total of 93 features from MRI and 93 features from PET image are obtained. Discriminative Multi-Task Feature Selection (DMTFS) model deliberates the integral relations among multimodality information and the distribution data of both the intra-class and the inter-class subjects from all the modalities. The proposed technique formulates the feature selection on multi-modality information as a multi-task learning problem, then, two regularized terms are included, namely; i) group-sparsity regularization, for ensuring only the common brain region-specific features, jointly selected from multimodality data, ii) Laplacian regularization for preserving the compactness of intra-class subjects and the separability of inter-class subjects to induce more dissimilar features.

A feature-ranking-based classification framework to classify AD vs CN subjects is proposed in the literature [98], where all input images are analyzed using a voxel-wise parametric mapping. The grey matter volume changes are detected by using the voxel-based analysis over the whole brain. For isolating the Region of Interests (RoIs), Voxel-based morphometry (VBM) based mining procedure, and Diffeomorphic Anatomical Registration using The Exponentiated Lie algebra (DARTEL) analysis is applied. The regions, where a significant decrement of grey matter is taken place are segmented using a 3D mask and the ''MarsBaR region of interest'' toolbox.

A multi-modality, multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification is proposed in [99]. Initially, the brain MRI is partitioned into 93 Region of Interests (RoIs) by using the Jacob template. For each subject, the pre-processed PET images are aligned to their respective MR images using affine registration, and finally, 93 features from the MR image and 93 features from the PET images are acquired for each subject. The authors have proposed a multitask feature selection technique for preserving the complementary inter-modality information by taking the feature selection from each modality as a separate task. A constraint for preserving the inter-modality relationship is imposed, and enforce sparseness of the selected features from each modality separately.

A hierarchical fusion of features and classifier decisions for AD classification is proposed in the literature [100]. The authors have performed a t-test for selecting the voxels with significant group differences, considering the threshold value (P) smaller than 0.05. The mean of the P-value is calculated for all the selected voxels and sorted in ascending order. For capturing both imaging, and structural information, 2 kinds of features are mined for each of the patches, namely i) local imaging features, based on the grey matter densities of the patch, ii) correlations among local patches, known as the spatial-correlation features.

Using the probability distribution function, a classification framework to classify AD vs CN subjects is proposed in the literature [101]. To extract and isolate the Volume of Interests (VoIs), Diffeomorphic Anatomical Registration using The Exponentiated Lie algebra (DARTEL), and Voxel-based morphometry (VBM) analysis is applied. The regions which experience a significant grey matter volumetric loss, are identified by the DARTEL-Voxel-based morphometry (VBM) approach and segmented using a 3D mask. The data are divided randomly into 10 folds with

an equal number of AD and NC subjects in each fold. For every iteration, 1 fold is used for testing and 9 folds are used for training. Based on each training dataset, the authors have performed Voxel-based morphometry (VBM)-DARTEL analysis to reveal regions of decreased grey matter volume in patients as a 3D mask. From a total of 59,395 to 69,170 voxels, 10 different masks with different lengths are defined.

Based on cortical and sub-cortical features, an AD classification framework is proposed in the literature [102]. The cortical thickness and sub-cortical volume are extracted from brain images using the Freesurfer toolbox. A total of 110 features are extracted from the 3D Structural Magnetic Resonance Imaging (SMRI) T1-weighted image. Freesurfer provides the ability to construct surface-based morphometry (SBM) for representations of the cortex, from which neuroanatomic volume, cortical thickness, and surface area can be derived. The cortical surface lies either at the white matter/grey matter tissue interface or in the grey matter/cerebrospinal fluid tissue interface.

Using the structural MR images, an AD classification framework, based on the Voxel Morphometry (VM) based feature extraction approach is discussed in the literature [103]. From the normalized tissues of grey matter, white matter, and the cerebrospinal fluid, the down-sampling operation is performed on the densities of the voxels of 1 mm size. By performing the simple averaging, the voxels are sampled into a size of 8 mm, which contains not more than 10% density values and ignores more than 50% parts of the total image for further analysis. The dimensions of the maps are $22 \times 27 \times 22$ voxels. From the maps, feature vectors are introduced. To eliminate the cerebellum from entire data sets, a Region of Interests (RoI) is drawn on the custom template. To select the suitable features from the grey matter, white matter, and cerebrospinal fluid densities, the linear SVM based principle is applied. After selecting the most appropriate features for classification, 26 neighborhood voxels (in a $3 \times 3 \times 3$ cube) carrying non-zero weight are also considered in the classification. Newly formed weight vectors are then represented as the highest absolute weight in the surroundings of $3 \times 3 \times 3$ cube of voxels. The weight vectors are then considered as a threshold to get top-ranked voxels.

For early diagnosis of AD, a novel classification mechanism using combined features from voxel-based morphometry, cortical, sub-cortical, and hippocampus regions are described in the literature [104]. For the extraction of features, such as Voxel-Based Morphometry (VBM), Cortical and subcortical volumetric features, and Hippocampus volume (HV), the authors have used FreSurfer and Statistical Parametric Mapping (SPM)12 toolbox. The Voxel-based morphometry (VBM) performs voxel-wise statistical assessments for determining the volume transformations among different parts of the brain. The affine transformation by the Statistical Parametric Mapping (SPM) templates is used for data standardization to compensate for the size differences. Adopting the unified tissue segmentation technique, all the input images are segmented into three parts, namely grey matter, white matter, and cerebrospinal fluid. All the linearly distorted and segmented images are then non-linearly distorted by applying the Diffeomorphic Anatomical Registration (DARTEL) method. Based on the Montreal Neurological Institute (MNI)152 template, all images are modulated and smoothed by applying an 8 mm full breadth at half maximum kernel for creating the modified template of DARTEL. By using the default constraints of the cross-sectional automated Freesurfer routine, the important features from cortical and subcortical sections are extracted. Volumetric quantities for all lobes are mined by the Freesurfer. The Desikan-Killiany atlas is used, which labeled the whole cortex in 68 sections for each hemisphere. As hippocampus volume is one of the most commonly used biomarkers in the diagnosis of AD, the authors segmented both left and right hippocampus volume using the FreeSurfer toolbox.

Based on the Mann–Whitney–Wilcoxon U-Test, an AD detection approach is described in the literature [105]. All images are reconstructed with a total of 67200 voxels. The intensities of each voxel lie between 0 to 255. To reduce the dimensionality, at first, the voxels with an intensity value less than 70 are excluded. After that, using the Mann–Whitney–Wilcoxon (MWW) U-Test, the most fitted voxels for the classification are selected. By performing a factor analysis, the selected voxels are then modeled. The voxels are labeled according to linear combinations of the factors. The factor loadings are determined for describing the discriminability between the selected voxels, which help in reducing the dimension of the data.

A classification framework to classify AD vs CN subjects by sparse representation is proposed in the literature [106]. The average voxel intensity is calculated, and voxels having an intensity that is less than half of the average intensity value, are discarded, and determined as the background voxels. For MR images, after performing the spatial normalization using Voxel-based morphometry (VBM)-T1 template, input images are segmented into three parts; grey matter, white matter, and cerebrospinal fluid. Then in all images, the Sparse Representation Classifiers (SRC) dictionary-based approach is applied to choose the most effective voxels, and to remove the voxels having less information. For the voxels which are selected by the SRC method, Welch's t-test is performed separately. However, the authors have used separate activation levels for different image types.

A local MRI analysis approach for diagnosing AD is described in the literature [107]. With the help of the atlas-based method, some Volume of Interest (VoI) is specified. Once the templates are ready, a rigid registration technique is used to map them onto the target MRI, and the correlation coefficients are determined from the extracted VoIs. The process returns VoIs having the best correlation value. For this study, a total of 9 VoIs are extracted. The process of VoI selection and extraction is done with the help of Insight Segmentation and Registration Toolkit (ITK), FMRIB's Linear Image Registration Tool (FLIRT), and Matrix Laboratory (MATLAB) toolbox. The extracted

9 VoIs are then filtered with 18 different kinds of filters. The neighborhood voxels of each selected voxels are compared, and then, the Gaussian mean, standard deviation, range, entropy, and Mexican-hat filters are determined.

An optimal decisional space based method for the classification of AD and MCI is proposed in the literature [108]. To generate the variable vector discriminator for each subject, the Intracranial Volume (ICV) of all the variables are determined and combined with the MMSE score. For determining the consequence of each variable, Student's t-test is adopted between AD and CN, or between MCI and CN. The variables, whose t-test scores with a p-value are less than the significance level, are nominated and ranked. For reducing the dimensionality, an incremental error analysis method is applied only to the top-ranked variables, determined by the t-test.

A novel grading biomarker for the prediction of conversion of MCI to AD is described in the literature [109]. For selecting the relevant features, the authors have adopted the Elastic Net (EN) technique. One advantage of using EN is that, even if the number of features is very large, it can easily select the most appropriate features. The popular Least Absolute Shrinkage and Selection Operator (LASSO) regression is used, which helped in selecting the highly correlated features from different groups. The implementation of EN is done in the Sparse Modeling Software (SPAMS) toolbox. Finally, the authors identified a biomarker by circulating the disease labels of NC and AD to the MCI subjects. A global grading value is determined for MCI subjects which are used as a biomarker in the classification step. The CN and the AD subjects are considered as the train population. The members of the train population are compared along with their relationship with the MCI subjects and assigned a new grading value for them.

An inherent structure-based multi-view learning with multi-template feature representation is proposed for AD classification in the literature [110]. For extracting the most relevant features, an affinity propagation (AP) clustering procedure is applied. The input images are partitioned into several clusters. The centroid for each cluster is determined and used as templates. The bisection method is applied for determining a suitable predilection value to find out the most similar data points from the centroids. A total of 10 templates are selected in this study. The authors have applied a mass-preserving shape transformation mechanism for capturing the morphometric shapes of every considered subject with the help of multiple templates. The segmentation and registration steps are applied next to extract the volumetric structures. The clustering tissues are adapted into the Region of Interests (RoIs) of each template space to extract the relevant features. Moreover, by adopting a subclass-based approach, the inherent patterns of each template space are also extracted.

A classification approach to classify AD vs CN, MCI vs CN, and Progressive MCI (PMCI) vs Stable MCI (SMCI) subjects, using the volume-based morphometry is proposed in the literature [111]. The Statistical Parametric Mapping (SPM) toolbox is used to convert input images into numerous tissue probability maps, where a grey matter probability map is also included. The grey matter segmentation is performed by using a Bayesian based image segmentation procedure, known as the New Segment. With Jacobian determinants of deformations, the map is spatially smoothed, and warped for a referencing space to allow a voxel-wise assessment of different subjects. Moreover, the reference space is iteratively enhanced from grey matter and white matter probability maps using the Diffeomorphic Anatomical Registration using The Exponentiated Lie algebra (DARTEL) algorithm. FreeSurfer toolbox is used to segment the input images in a large number of anatomical constructions, and then, calculated volumes of all the corresponding segments. The toolbox is primarily absorbed in temporal grey matter, total grey matter, hippocampus, and ventricular volumes output by FreeSurfer to determine the latent biomarkers from AD related brain atrophy. The authors have implemented a brain volumetric procedure, known as MorphoBox, that combines simple and fast image analysis approaches for performing the Volume-Based Morphometry (VolBM).

Some of the research articles on AD classification, using Voxel Morphometry (VM) based feature extraction techniques, are discussed, and the performance comparison is presented in table 3.

From table 3, it can be noticed that, among all the discussed methods, the highest performance is achieved by Mesrob *et al.* [96] with a rate of 99.60% accuracy, 99.25% sensitivity, and 99.95% specificity.

C. WAVELET TRANSFORM-BASED FEATURE EXTRACTION

Wavelet transform (WT), is a well-known approach for analyzing signals, from where the detailed information of an object can be evaluated. Since WT is defined in both spatial frequency 'v' as well as the spatial position 't', it can be written as a function in the form of WT(v, t) [112]. WT breaks down the signal with limited energy, from the spatial area to a set of functions. WT is also one of the most widely used feature extraction tools in image processing [113]. In WT, the fusion of input images is transformed from the spatial domain to a wavelet domain [114]. Wavelet domain characterizes the wavelet coefficient, and then, wavelet decomposition is done by moving the image through a sequence of low-pass and high-pass filters. Several filter bands forms, where every band produces a separate resolution and orientations.

One of the first AD classification methods using wavelet analysis is proposed by Padilla *et al.* [115]. The authors have introduced the concept of the Gabor wavelet (GW) based brain analysis approach on the input SPECT images followed by a Fisher Discriminant Ratio (FDR) based feature extraction technique. Then Principle Components Analysis (PCA) feature selection technique is used to select the most relevant features. The final classification is done by using the SVM

TABLE 3. Performance comparison of the AD classification techniques using Voxel Morphometry (VM) based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
Iman Beheshti, et al. [91]	AD vs. CN: 93.01%, PMCI vs. SMCI: 75%	AD vs. CN: 89.13%, PMCI vs. SMCI: 76.92%	AD vs. CN: 96.80%, PMCI vs. SMCI: 73.23%	AD vs. CN: 90.4%, PMCI vs. SMCI: 62.6%	AD vs. CN: 88.3%, PMCI vs. SMCI: 71.2%	AD vs. CN: 0.94, PMCI vs. SMCI: 0.75
Iman Beheshti, et al. [92]	AD vs CN: 92.48%	AD vs CN: 91.07 %	AD vs CN: 93.89%	NA	NA	AD vs CN: 0.96
Olfa Ben Ahmed, et al. [93]	AD vs. CN: 87%, MCI vs. CN: 78.22%, AD vs. MCI: 72.23%	AD vs. CN: 75.5 %, MCI vs. CN: 70.73%, AD vs. MCI: 75%	AD vs. CN: 100%, MCI vs. CN: 83.34%, AD vs. MCI: 70%	NA	NA	NA
Mingxia Liu, et al. [94]	AD vs. CN: 92.51%, PMCI vs. SMCI: 78.88%	AD vs. CN: 92.89%, PMCI vs. SMCI: 85.45%	AD vs. CN: 88.33%, PMCI vs. SMCI: 76.06%	NA	NA	AD vs. CN: 0.96, PMCI vs. SMCI: 0.81
Chen Zu, et al. [95]	AD vs CN: 95.95%, MCI vs CN: 80.26%	AD vs CN: 95.10%, MCI vs CN: 84.95%	AD vs CN: 96.54%, MCI vs CN: 70.77%	NA	NA	AD vs CN: 0.97, MCI vs CN: 0.81
Lilia Mesrob, et al. [96]	AD vs. CN: 99.60%	AD vs. CN: 99.25%	AD vs. CN: 99.95%	NA	NA	NA
Tingting Ye, et al. [97]	AD vs. CN: 95.92%, MCI vs. CN: 82.13%, PMCI vs. SMCI: 71.12%	AD vs. CN: 94.71%, MCI vs. CN: 87.68%, PMCI vs. SMCI: 67.21%	AD vs. CN: 97.12%, MCI vs. CN: 71.54%, PMCI vs. SMCI: 73.93%	NA	NA	AD vs. CN: 0.97, MCI vs. CN: 0.82, PMCI vs. SMCI: 0.68
Iman Beheshti, et al. [98]	AD vs. CN: 96.32%	AD vs. CN: 94.11%	AD vs. CN: 98.52%	NA	NA	AD vs. CN: 0.99
Feng Liu, et al. [99]	AD vs. CN: 94.37%, MCI vs. CN: 78.80%, PMCI vs. SMCI: 67.83%	AD vs. CN: 94.71%, MCI vs. CN: 84.85%, PMCI vs. SMCI: 64.88%	AD vs. CN: 94.04%, MCI vs. CN: 67.06%, PMCI vs. SMCI: 70%	NA	NA	AD vs. CN: 0.97, MCI vs. CN: 0.83, PMCI vs. SMCI: 0.70
Manhua Liu, et al. [100]	AD vs. CN: 92%, MCI vs. CN: 85.3%	AD vs. CN: 90.9%, MCI vs. CN: 82.3%	AD vs. CN: 93%, MCI vs. CN: 88.2%	NA	NA	AD vs. CN: 0.95, MCI vs. CN: 0.91
I. Beheshti, et al. [101]	AD vs. CN: 89.65%	AD vs. CN: 87.73%	AD vs. CN: 91.57%	NA	NA	AD vs. CN: 0.95
Yubraj Gupta, et al. [102]	AD vs. CN: 99.34%, MCI vs. CN: 99.2%, AD vs. MCI: 97.77%	AD vs. CN: 98.14%, MCI vs. CN: 99.02%, AD vs. MCI: 100%	AD vs. CN: 100%, MCI vs. CN: 100%, AD vs. MCI: 95.23%	NA	NA	NA
Prashanthi Vemuri, et al. [103]	NA	AD vs. CN: 88.4(±1.6) %	AD vs. CN: 88.6 (±1.3) %	NA	NA	NA
Yubraj Gupta, et al. [104]	AD vs. CN: 93.06%, PMCI vs. SMCI: 86.95%, MCI vs. CN: 95.23%	AD vs. CN: 87.87%, PMCI vs. SMCI: 77.77%, MCI vs. CN: 95.77%	AD vs. CN: 95.58%, PMCI vs. SMCI: 92.85%, MCI vs. CN: 92.30%	NA	NA	AD vs. CN: 0.94, PMCI vs. SMCI: 0.87, MCI vs. CN: 0.96
F.J. Martínez Murcia, et al. [105]	AD vs. CN: 92.7%	AD vs. CN: 97.6%	AD vs. CN: 92.6%	NA	NA	NA
Andres ORTIZ, et al. [106]	AD vs. CN: 94%	NA	NA	NA	NA	NA
Andrea Chincarini, et al. [107]	NA	AD vs. CN: 89%, MCI vs. CN: 89%, PMCI vs. SMCI: 72%	AD vs. CN: 94%, MCI vs. CN: 80%, PMCI vs. SMCI: 65%	NA	NA	AD vs. CN: 0.97, MCI vs. CN: 0.92, PMCI vs. SMCI: 0.74
Qi Zhou, et al. [108]	AD vs. CN: 92.4%, PMCI vs. CN: 92.4%, SMCI vs. CN: 92.4%	AD vs. CN: 84.0%, PMCI vs. CN: 84.0%, SMCI vs. CN: 84.0%	AD vs. CN: 96.1%, PMCI vs. CN: 96.1%, SMCI vs. CN: 96.1%	NA	NA	NA
Tong Tong, et al. [109]	PMCI vs. SMCI: 84.1%	PMCI vs. SMCI: 88.7%	PMCI vs. SMCI: 76.5%	NA	NA	PMCI vs. SMCI: 0.92

TABLE 3. *(Continued.)* **Performance comparison of the AD classification techniques using Voxel Morphometry (VM) based feature extraction approaches.**

Mingxia Liu, et al. [110]	AD vs. CN: 93.83%, PMCI vs. SMCI: 80.90%	AD vs. CN: 92.78%, PMCI vs. SMCI: 85.95%	AD vs. CN: 95.69%, PMCI vs. SMCI: 78.41%	NA	NA	NA
Daniel Schmitter, et al. [111]	AD vs. CN: 89%, AD vs. MCI: 68%, PMCI vs. SMCI: 71%	AD vs. CN: 86%, AD vs. MCI: 69%, PMCI vs. SMCI: 75%	AD vs. CN: 91%, AD vs. MCI: 67%, PMCI vs. SMCI: 66%	AD vs. CN: 88%, AD vs. MCI: 49%, PMCI vs. SMCI: 75%	AD vs. CN: 89%, AD vs. MCI: 82%, PMCI vs. SMCI: 66%	NA

classifier. The proposed classification method produced a convincing classification outcome.

WT still has some drawbacks, such as it can't deal with the shift in-variance, and also, it Can't detect edges of a region accurately. Moreover, it offers partial information along with all the directions of a 3D image [116]. Some of the AD classification approaches where wavelet transform is used for feature extraction are discussed below.

For automated detection of Alzheimer's disease, a novel AD classification approach is proposed in the literature [117]. Initially, Contourlet Transform (CoT) is used for feature extraction from the pre-processed brain MRIs. Later on, for performance comparison, the authors have used some more commonly used feature extraction techniques in the same pre-processed brain MRIs, including, Curvelet Transform (CuT), Complex Wavelet Transform (CWT), Dual Tree Complex Wavelet Transform (DTCWT), Discrete Wavelet Transform (DWT), Empirical Wavelet Transform (EWT), and Shearlet Transform (ST). To select the most appropriate features, a student's t-test is used.

A classification framework to identify AD vs CN subjects is proposed in the literature [118]. The concept of wavelet transform is used for feature extraction. Inter-Class Variance (ICV) in the axial direction is determined for all the slices of a 3D image, and then a slice having the maximum ICV value is picked up for further processing. After comparing several wavelets, biorthogonal wavelet (bior4.4) is used because transforms of bior4.4 are similar to the gray value changes in brain images. Entropy S (Shanon entropy), is defined to find out the degree of its randomness.

A twin SVM-based classification of Alzheimer's disease, using complex dual-tree wavelet principal coefficients and Linear Discriminant Analysis (LDA) is discussed in the literature [119]. Firstly, the authors have proposed the algorithm for extracting the 5-level Dual Tree Complex Wavelet Transform (DTCWT) coefficients from all the input MR images, where, features of the 5^{th} resolution scale are used. Secondly, selected coefficients are used as the inputs in Principal Component Analysis (PCA), to map the features onto lower-dimensional space. For getting the most discriminative features, PCA coefficients are projected onto an LDA projection axis.

A novel approach for the classification of AD from MRIs by using the fuzzy neural network is proposed in the literature [120]. The pre-processed images are represented as 2D histogram signals and performed the intensity enhancement operation. This step aims to take all the features, in the form of vector approximation at every level of the wavelet decomposition. Image decomposition is done using the wavelet transformation (WT) and then output coefficient vectors are determined by using the discrete wavelet transform. The derived features are used to train the Fuzzy Neural Network (FNN).

For diagnosing AD, a classification methodology using dual-tree complex wavelet transform, Principal Component Analysis (PCA), and feed-forward neural network is proposed in the literature [121]. Dual-Tree Complex Wavelet Transform (DTCWT) develops two Discrete Wavelet Transform (DWT)s processing. The first DWT signifies the real module, and the second DWT represents the imaginary module of the transform. The DTCWT produced 6 directionally selective sub-bands, oriented in ± 15, ± 45, and ± 75 directions, for real and imaginary parts. By using DTCWT, the coefficients from each pre-processed image are extracted. Some supplementary features, such as age, gender, handedness, education, Socio Economic Status (SES), and clinical examination are also used in classification. The PCA analytically projects the input data to a lower-dimensional space, known as the principal subspace. The operation is done over an orthogonal alteration by conserving the data dissimilarities. In the group of correlated variables, the alteration that results in a group of linearly uncorrelated variables is known as the principal components (PCs). The reason for implementing PCA is to reduce the dimensionality of the DTCWT coefficients for correct classification.

For the classification of AD, some methods, based on the Wavelet Transform (WT) based feature extraction techniques proposed by several researchers are discussed. The performance of the discussed methods is presented in table 4.

From table 4, it can be observed that among all the discussed AD classification methods using Wavelet Transform (WT) based feature extraction approaches, the article by Alam *et al.* [119] provides the highest accuracy ($96.68 \pm 1.44\%$), highest sensitivity ($97.72 \pm 2.34\%$), as well as the highest specificity ($95.61 \pm 1.67\%$).

D. GRAPH/NETWORK-BASED FEATURE EXTRACTION

A graph is a collection of some connected nodes, where each node represents the entities and each connection represents relationships between the connected nodes. Graph based feature extraction method uses the supervised information while creating the neighboring relationship matrix of the graph

TABLE 4. Performance comparison of the AD classification techniques using wavelet transform-based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
U. Rajendra Acharya, et al. [117]	AD vs. CN: 94.54%, PMCI vs. SMCI: 69.3%	AD vs. CN: 96.3%, PMCI vs. SMCI: 66.7%	AD vs. CN: 93.64%, PMCI vs. SMCI: 71.2%	AD vs. CN: 88.33%, PMCI vs. SMCI: 62.6%	AD vs. CN: 92.17%, PMCI vs. SMCI: 71.2%	NA
Shui-Hua Wang, et al. [118]	AD vs. CN: 92.40±0.83%	AD vs CN: 92.14±4.39 %	AD vs CN: 92.47±1.23%	NA	NA	NA
Saruar Alam, et al. [119]	AD vs. CN: 96.68±1.44%	AD vs. CN: 97.72±2.34%	AD vs. CN: 95.61±1.67%	NA	NA	NA
Geetha C, et al. [120]	AD vs. CN: 95.5%	AD vs. CN: 88%	AD vs. CN: 86%	NA	NA	NA
Debesh Jha, et al. [121]	AD vs. CN: 90.06±0.01%	AD vs. CN: 92.00±0.04%	AD vs. CN: 87.78±0.04%	NA	NA	NA

and determining the appropriate features by computing their competence of conserving geometrical construction of the graph. However, graph-based approaches determine the features independently, hence, sometimes it is unable to handle the redundant features accurately [122]. Some of the literature where the concept of graph/network construction is used to extract the most feasible features are discussed below.

For classification of dementia using brain MRI, a novel method is discussed in the literature [123]. A graph-based multiple instances learning method is used to train the bag level classifier. In the proposed method, a graph is constructed for each image. In the graph, patches are treated as nodes, and edges between different nodes are established according to the relationships between the patches. The graphs can represent the appearances of patches, and reflect the relationships among the extracted patches from the same subject. Some patches are extracted from AD subjects, and some are from Progressive MCI (PMCI), Stable MCI (SMCI), and CN subjects. The resulting graphs are expected to be different for each subject group. A graph kernel is defined for distinguishing the positive and negative bags. Finally, a bag-level classifier is adopted using a kernel machine, influenced by the support vector machine (SVM). By using the computed graph kernels, SVM is used to train the classifier. In the test stage, labels of unseen images are estimated using the trained classifier.

A classification approach to classify AD vs CN, and MCI vs CN subjects, is proposed in the literature [124]. Total 239,391 features are extracted from all the subjects, which include, 83 nos. of MRI volume features, 239,304 nos. of PET images intensity features, 3 cerebrospinal fluid measures, and 1 genetic categorical feature. The brain MRI is segmented into 83 anatomical regions by Multi-Atlas Propagation with Enhanced Registration (MAPER) approach, which helps in extracting the region-wise features. All Fluorodeoxyglucose (FDG)-PET images are motion-corrected, and associated with their corresponding MRIs, which are distorted to the Montreal Neurological Institute (MNI) template space using Statistical Parametric Mapping (SPM)8 toolbox, and then images are smoothened to an isotropic spatial resolution of 8 mm full-width-at-half-maximum (FWHM). Intensity normalization is performed, and the voxel wise intensities are extracted as features. Cerebrospinal fluid of all the subjects is extracted by lumbar puncture, and then, the Aβ42 level, $T - tau$, and $P - tau$ are measured. The APOE genotype information, which is determined from the blood samples of all the subjects, is obtained from ADNI Biomarker data information. A graph, $G^i = (V^i; E^i)$ has constructed by the authors for each subject, where V^i corresponds to n subjects of i^{th} modality, and E^i for i^{th} modality are weighted by similarities of the subjects. To determine similarities, the random forest method is used between pairs of subjects. Authors have used the normalization operation to all the similarity matrices in order to fuse the graphs, based on an assumption that, local contacts, having the most similarities, are more reliable than non-local connections. The authors have proposed to concatenate features from all modalities to a solo feature vector for classification, to provide a straightforward way by using multi-modality data.

An algorithm for classification of AD and prediction of MCI Conversion using a histogram-based analysis is discussed in the literature [125]. All input images are bias refined and then segmented into 3 parts, namely grey matter, white matter, and cerebrospinal fluid, by using the Voxel-based morphometry (VBM)8 toolbox. Modified brain networks are built using the $3 \times 3 \times 3$ grey matter voxels cube of size $6 \times 6 \times 6$ mm^3. The constructed brain networks can be distinctively mapped to an $M \times M$ connectivity zero-diagonal and symmetric matrix Z, where, $x_{j,k}$ is an element of z, that signifies the maximum correlation between the (j,k) cube pair. For reducing the high dimensionality, and unreliable size of the vector, a histogram-based approach is used to map the vector in a statistical pattern. The defined vector is having lower dimensionality, which provides circulation of the repeating values, falls into the interval bins in the raw-feature vector, which is known as the Histogram Feature Vector (HFV).

Using the concept of regional saliency maps, a classification framework to classify AD vs CN, and MCI vs CN, is proposed in the literature [126]. For automatic extraction of the regions, which are associated with important pathology from each slice of the MRIs, some randomly sampled patches are extracted. To get the compact set of visual primitives (visual words), a set of training images are considered, which are characterized using a multi-scale edge analysis. Next, a probabilistic Latent Semantic Analysis (pLSA) is trained

to gather hidden information associated with those regions. Then the slices are processed using the probabilities learned with pLSA. Initially, a group of slices of each volume is selected using a set of patches, sampled randomly from the training slices. Then those patches are categorized by the Sobel edge detector. Edge information is concatenated into a single vector and collected together. Thus a visual vocabulary is formed that is used to represent each training slice by a histogram of visual words, suitable to train the pLSA. Graph-Based Visual Saliency (GBVS) concatenates the discriminatory pixels with a concept of nearness in a straight manner to obtain the saliency values, by modeling the image as a fully-connected graph and storing information at edges. The steps for calculating the saliency maps are feature extraction, activation maps, and combination. Initially, some related features are mined, then a connected graph is created for each feature, and scaled images, by storing the discriminatory and nearness information at the edge. The activation maps are determined by creating a Markov Chain in the graph and equilibrium distribution is calculated as the major eigenvector of the transition matrix. The map is standardized by focusing the mass obtained in activation step by using the same Markovian method. Then, average saliency maps per feature are determined and combined, to form the master saliency map.

A classification method to classify AD vs CN subjects using sulcal features is discussed in the literature [127]. The sulcal meshes are determined as a set of 3D vertices, and then classified two differing faces of the mesh, using a k-means clustering-based method. Lastly, a medial surface is calculated, which consists of some new vertices, between two faces. The algorithm extracted local surface features from the medial surface. The procedure is characterized by 3 boxes, compressed by a set of input cerebral sulcal meshes, and the output extraction of local surface features. For extracting sulci from input images, authors have used the BrainVISA 4.4.0 Morphologist 2013 toolbox. While extracting the sulci, a triangular mesh of the inner cortical surface of each brain hemisphere is produced, and then a graph of the cortical fold is constructed. The toolbox finally recognized, labeled, and extracted all cortical sulci automatically. By surface extraction, and extraction of sulci, sulcal features are calculated and mined from 24 sulci. A total of 10080 numbers of sulci are mined from 210 subjects from both left and the right hemisphere in each subject. For each sulcus, the depth, length, mean curvature, Gaussian curvature, and surface area features are computed.

A mechanism for classification of MCI and AD, from CN using directed graph measures of resting-state fMRI, is discussed in the literature [128]. A total of 264 Region of Interests (RoIs) of atlases are determined for parcellating the brain. The areas are designed by applying meta-analytic and functional connectivity plotting with the resting-state-fMRI data. For generating a demonstrative signal for each Region of Interests (RoI), the time series of the voxels are averaged. The authors also used the Automated Anatomical atlas (AAL) parcellation. Out of 244 Region of Interests (RoIs), 90 most effective Region of Interests (RoIs) is obtained by the AAL and averaged the signals of 90 time series for each subject. The Granger causality analysis is used to determine connectivity among all Region of Interests (RoIs) for constructing the directed brain network. The graph measures, such as the degree (in-degree and out-degree), betweenness centrality, flow coefficient, local efficiency, K-coreness centrality, page rank centrality, node strength, clustering coefficient, global efficiency, characteristic path length, range coefficient, etc., are determined by the Brain Connectivity Toolbox (BCT). Based on the discriminative properties, the Fisher algorithm is used to sort all features. After that, half of the best features with the highest discriminative properties are sorted by using the wrapper feature selection algorithm.

An AD Classification method Based on the individual hierarchical networks, is proposed in the literature [129]. For defining the Region of Interests (RoIs), a grey matter-based Automated Anatomical Labeling (AAL) atlas is used which included 90 Region of Interests (RoIs) from the cerebral regions, and 26 Region of Interests (RoIs) from the cerebellar regions which are denoted as S1. To reduce the number of Region of Interests (RoIs), some of the Region of Interests (RoIs) from S1, where the first two digits and the last digit are similar are combined. For example, Cingulum Ant L (4001), Cingulum Mid L (4011), and Cingulum Post L (4021) are combined and treated as a single Region of Interests (RoI). The newly combined Region of Interests (RoIs) is denoted as S2, where a total of 54 Region of Interests (RoIs) are selected. Similarly, some of the Region of Interests (RoIs) from S1, where the first and the last digit are similar, such as Cingulum Ant L (4001), Cingulum Mid L (4011), Cingulum Post L (4021), Hippocampus L (4101), Para Hippocampal L (4111) and Amygdala L (4201) are combined together and denoted as S3. A total of 14 Region of Interests (RoIs) are included in S3. Similarly, in S4, only one RoI is left, i.e, the whole brain. Based on the cluster of Region of Interests (RoIs), construction of a hierarchical network is done where the nodes, namely S1, S2, S3, and S4, and the link among the Region of Interests (RoIs) are denoted as edges. From every node of the network, 6 texture features are extracted by applying the Gray Level Co-occurrence Matrices (GLCM) technique. The extracted features, namely Energy (ENE), Contrast (CON), Inverse Difference Moment (IDM), Entropy (ENT), Difference Variance (DVA), and Difference Entropy (DEN). After extracting the features, mean texture property of Region of Interests (RoIs) is determined in 4 directions (i.e,. 0^0, 45^0, 90^0, and 135^0).

Using the whole brain hierarchical network, a novel classification framework for AD detection is discussed in the literature [130]. To perform anatomical parcellation of the brain, an Automated Anatomical Labeling (AAL)-based method is applied, and 90 cerebral and 26 cerebellar regions are selected. Among all the selected Region of Interests (RoIs), some RoIs have some functional dependencies. For example, Cingulum, Hippocampus, etc. have some common

properties, as they all are part of the limbic system. Based on the functional dependencies, similar RoIs are grouped together and assigned in 4 groups, namely; L1, L2, L3, and L4. L4 contains 90 RoIs, L3 contains 54 RoIs, L2 contains 14 RoIs, and L1 contains only one RoI, i.e, whole brain. From each of the RoIs, 3D texture features are extracted in 0^0, 45^0, 90^0, and 135^0 directions by applying the GLCM-based approach. For all 4 directions, the mean constraints of every RoIs are determined by a six-dimensional vector. The connectivity among all RoIs is determined with help of the Pearsons' correlation coefficients. At the end, a Whole Brain Hierarchical Network (WBHN) is constructed, where the RoIs represent the nodes and the connectivity among the RoIs, representing their edges.

An AD classification mechanism by combining multiple measures is proposed in the literature [131]. Using the Automated Anatomical Labeling (AAL)-based approach, all the images are registered. In the cortical regions, operations like the determination of the Grey Matter Volume (CGMV), Cortical Thickness (CT), Cortical Surface Area (CSA), Cortical Curvature (CC), Cortical Folding Index (CFI), and Subcortical Volume (SV) are observed with the help of FreeSurfer toolbox. Similarly, volumes of subcortical regions are determined with the help of the Fast Analysis of Sequences Toolbox toolbox. By taking reference from the nonlinear Montreal Neurological Institute (MNI)-152 template, affine registration is done in the whole brain region. An AAL-based approach is applied in the subcortical regions in order to get correct affine registration. Based on the properties of cortical and subcortical regions, two networks are created. For cortical regions, a total of 5 different networks are constructed based on 5 different measurements (i.e, CGMV, CT, CSA, CC, and CFI). For all the networks, cortical regions are represented as nodes and CGMV similarities among the nodes are represented as edges. Similarly, for the subcortical regions, a single network had constructed based on the measurement of Great Saphenous Vein (GSV).

Some of the recently published articles on AD classification using graph/network based feature extraction approaches are discussed. The performance of the discussed articles is presented in table 5.

From table 5, it can be noticed that amongst all the discussed AD classification methods by graph/network based feature extraction approaches, the maximum accuracy is achieved by Liu *et al.* [129], which is 95.37%, whereas, the maximum performance rate is achieved by Khazaee *et al.* [128], which is approximately 93.79%.

E. EIGENVECTOR-BASED FEATURE EXTRACTION

The eigenvalue or eigenvector is a transformation of the covariance matrix, that helps to reveal the primary directions of dissimilarity among all the images in a dataset of different subjects [132]. The application of eigenvector includes, image classification, object identification, etc [132]. Some of the advantages of the eigenvector value-based feature extraction approach are a) features that are not correlated can be extracted, b) determination of most suitable linear calculation, c) finding out the discrepancy in the mined features, etc [133]. The eigenvector value-based feature extraction approach has some demerits, such as a) Self-regulating variables become less explainable, b) It is difficult to extract the features when an image has a complex shape, etc [134]. One of the commonly used eigenvectors based feature extraction methods is the Laplace Beltrami. In Laplace Beltrami, the discriminative pixels after comparing with a set of images of the same subject groups are selected, and a covariance matrix helps to track the discriminative direction [135].

Principal Component Analysis (PCA) is another popular eigenvector-based technique that helps to reduce the dimensionality by eliminating non-co-related features without losing much information of the images and then extract the most feasible features [136]. PCA determines the interior structure of data, based on the variances in the information. The major advantage of PCA is that it is very less sensitive to noise. Moreover, to implement PCA, very little information is required, and hence it's computationally faster [137]. Using the concept of PCA and LDA, M.M. López, *et al.* proposed a novel AD classification method in the literature [138]. The experiments are done using the SPECT data, where The FBP along with the Butterworth noise filter is applied to reconstruct the images. By implementing a mask based approach, voxels from the most affected brain parts such as, posterior cingulate gyri, precunei, temporo-parietal, etc. are extracted which are reconstructed into the vector forms. PCA based approach is used the extract the relevant feature vectors using the polynomial kernels and RBF for non-linear vector forms. For distinguishing different roles of the variations, LDA based projection is used. Finally, the authors have used the SVM classifier and achieved a convincing result.

PCA has some disadvantages, such as, components produced by PCA are the linear mixture of actual features, which are not as legible and understandable as the original features. Moreover, during the dimension reduction, some important features may also get eliminated [139].

Partial Least Squares (PLS) is a well known feature extraction approach, that works almost similar to PCA. The main difference between PCA and PLS is that PCA is unsupervised while PLS is supervised [77]. Based on the PLS method, one of the first novels approaches for AD classification is proposed in the literature [140]. The SPM toolbox is used for spatial normalization of all the input SPECT images, and then a binary mask is designed by allowing the voxels with an average intensity of more than 50% of the highest value. To select the most relevant features, the PLS based on the regression model is applied. For final classification, the Random Forest (RF) based classifier is applied. The proposed method achieved a better classification result with sensitivity $= 100\%$, specificity $= 92.7\%$ and accuracy $= 96.9\%$.

Some of the recently published AD classification literature, where, eigenvector based feature extraction is performed, are discussed below.

TABLE 5. Performance comparison of the AD classification techniques using graph/network-based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
Tong Tong, et al. [123]	AD vs. CN: 89.2%, PMCI vs. SMCI: 69.3%	AD vs. CN: 85.1%, PMCI vs. SMCI: 66.7%	AD vs. CN: 92.6%, PMCI vs. SMCI: 71.2%	AD vs. CN: 90.4%, PMCI vs. SMCI: 62.6%	AD vs. CN: 88.3%, PMCI vs. SMCI: 71.2%	NA
Tong Tong, et al. [124]	AD vs. CN: 91.8%, MCI vs CN: 79.5%	AD vs. CN: 88.9%, MCI vs CN: 85.1%	AD vs. CN: 94.7%, MCI vs CN: 67.1%	NA	NA	AD vs. CN: 0.98, MCI vs CN: 0.81
Iman Beheshti, et al. [125]	AD vs. CN: 84.17%, MCI vs. CN: 70.38%, PMCI vs. SMCI: 61.05%, AD vs. SMCI: 67.59%, AD vs. PMCI: 62.84%	AD vs. CN: 88.83%, MCI vs. CN: 78.17%, PMCI vs. SMCI: 52.65%, AD vs. SMCI: 79.25%, AD vs. PMCI: 76.38%	AD vs. CN: 79%, MCI vs. CN: 60.22%, PMCI vs. SMCI: 70.52%, AD vs. SMCI: 45.47%, AD vs. PMCI: 39.57%	NA	NA	AD vs. CN: 0.86, MCI vs. CN: 0.72, PMCI vs. SMCI: 0.62, AD vs. SMCI: 0.70, AD vs. PMCI: 0.64
Andrea Pulido, et al. [126]	AD vs. CN: 69.85%, MCI vs. CN: 87.21%	AD vs. CN: 67.14%, MCI vs. CN: 85%	AD vs. CN: 72.73%, MCI vs. CN: 87.77%	NA	NA	NA
Maciej Plocharski, et al. [127]	AD vs. CN: 87.9%	AD vs. CN: 90.0%	AD vs. CN: 86.7%	NA	NA	AD vs. CN: 0.89
Ali Khazaee, et al. [128]	CN: 93.29%, MCI: 93.29%, AD: 93.29%	CN: 88.4%, MCI: 100%, AD: 81.8%	CN: 100%, MCI: 85.5%, AD: 100%	CN: 100%, MCI: 88.9%, AD: 100%	NA	NA
Jin Liu, et al. [129]	AD vs. CN: 95.37%, MCI vs. CN: 86.56%, PMCI vs. SMCI: 73.95%, AD vs. MCI: 90.41%	AD vs. CN: 92.49%, MCI vs. CN: 90.74%, PMCI vs. SMCI: 76.13%, AD vs. MCI: 92.83%	AD vs. CN: 96.08%, MCI vs. CN: 84.83%, PMCI vs. SMCI: 72.24%, AD vs. MCI: 88.82%	NA	NA	AD vs. CN: 0.97, MCI vs. CN: 0.85, PMCI vs. SMCI: 0.73, AD vs. MCI: 0.92
Jin Liu, et al. [130]	AD vs. CN: 94.65%, MCI vs. CN: 84.79%, PMCI vs. SMCI: 72.08%, AD vs. MCI: 88.63%	AD vs. CN: 95.03%, MCI vs. CN: 88.91%, PMCI vs. SMCI: 75.11%, AD vs. MCI: 91.55%	AD vs. CN: 91.76%, MCI vs. CN: 80.34%, PMCI vs. SMCI: 71.05%, AD vs. MCI: 86.25%	NA	NA	AD vs. CN: 0.95, MCI vs. CN: 0.83, PMCI vs. SMCI: 0.72, AD vs. MCI: 0.91
Jin Liu, et al. [131]	AD vs. CN: 95.24%, MCI vs. CN: 86.35%, PMCI vs. SMCI: 74.28%, AD vs. MCI: 90.85%	AD vs. CN: 94.26%, MCI vs. CN: 89.49%, PMCI vs. SMCI: 71.51%, AD vs. MCI: 91.77%	AD vs. CN: 95.74%, MCI vs. CN: 85.68%, PMCI vs. SMCI: 76.46%, AD vs. MCI: 89.56%	NA	NA	AD vs. CN: 0.97, MCI vs. CN: 0.91, PMCI vs. SMCI: 0.79, AD vs. MCI: 0.94

A Laplace Beltrami eigen value based classification framework to classify AD vs CN subjects is proposed in the research paper [141]. The segmentation of Corpus Callosum (CC) is performed using the Reaction Diffusion (RD) level set method. Segmentation is compared with the manually segmented ground truth images and validated by an expert radiologist. The most important discriminative features are extracted using Laplace Beltrami (LB) eigenvalue spectrum. The inherent information of segmented CC is associated with the spectrum of the Laplace operator to find out the shape variations. The segmented images are measured as the closed bounded domain $\Omega \subset R^d$ with slice wise smooth borders, and its corresponding Laplace operator. From all the extracted features, a set of most discriminative features are selected by using the Information Gain (IG) based ranking method.

For early diagnosis of AD, a classification framework based on partial least squares, principal component analysis, and support vector machine, is discussed in the literature [142]. Feature extraction is done based on Principal Component Analysis (PCA) method. Authors have used the Partial Least Squares (PLS) to exploit co-variances amongst different sets of predictors and predicted variables. It is detected using a linear regression tool by projecting predicted variables and the predictor's variables to a new space. After perceiving 'n' data samples from each block of variables, PLS is used to decompose the $n \times N$ matrix of zero mean variables X, and the $n \times M$ matrix of zero mean variables Y, into the regression models form.

For designing a computer aided Alzheimer's diagnosis system, a novel classification mechanism is proposed in the literature [143]. For extracting the features having the maximum

variances, Principal Component Analysis (PCA) method is used. PCA is mainly acted as a zero-mean data-set, which is based on a linear transformation. The resultant vectors form a new group of de-correlated variables. The eigenvalue of vectors represents the variances amongst the variables. In this work, PCA is applied to the mean image vectors to extract all uncorrelated eigenvectors. The Independent Component Analysis (ICA) is used to determine the transformation of the peak level voxels, by taking reference from original mean image sources. The latent variables, which are produced by the ICA, are used to form a subspace of input images. The class variances are boosted after projecting image vectors on the subspace. The FastICA algorithm is applied to achieve the ICA transformation.

Based on Partial Least Squares and Support Vector Machine, a classification framework to classify AD vs CN subjects is discussed in the literature [144]. For reducing computational time with no loss of information, each voxel is down-sampled by a factor of two. Voxels, with an intensity value of more than 50% of the topmost intensity value, are selected and created using a binary mask. This binary mask is then applied to all the input images to determine the decrements of input spaces. Voxels, which are nominated by the mask, are reorganized in the form of a vector. The Partial Least Squares (PLS) method is applied to get a score and loading matrices. A weight matrix is also gained to determine the score vectors for the voxels, where the PLS method is not applicable. Thus, it can be easily observed, which voxels are taken into account for further processing and which are not. The Out of Box (OOB) is an error checking-based mechanism, used for noise removal. Leave-One-Out (LOO) method is used for validating the algorithm.

A Gaussian Mixture Models (GMM) based, SPECT image classification, for diagnosis of AD is proposed in the literature [145]. For density approximation, the authors have used the GMMs. In GMMs, trials are strained according to a Probability Distribution Function (PDF) which is demonstrated by a summation of k Gaussians. For selecting the maximum likelihood features, the Expectation Maximization (EM) algorithm is used. EM is an iterative optimization technique, that works in maximum likelihood approximation, even if part of some data is missing or incomplete. In order to extract the most effective features for classification, the authors have used the GMM based approach to select the Region of Interests (RoIs). While extracting the discriminative features, the authors considered the fact that the perfusion decoration of brain image for an AD subject is more variable than the CN perfusion. For dimensionality reduction of the selected features, a PCA based technique is used.

An Association rule-based feature selection method with the collaboration of Principal Component Analysis (PCA), for Alzheimer's disease diagnosis, is proposed in the literature [146]. The images are normalized and then a 3D mask is created by averaging all the images, where a threshold intensity a^T is determined by the 50% of maximum intensity value. Voxels, which are not covered by the mask, are discarded for further processing. Finally, the voxels whose intensity value is more than a^T are considered as activated. Apriori algorithm is applied to identify relevant associations between the concurrently activated brain areas from CN subjects. The allegations for Apriori rules are established in terms of antecedents and consequents, amongst earlier determined 3D activated blocks. Moreover, ARs are extracted as per the protocols of the leave-one-out (loo) cross-validation strategy.

Some of the AD classification methods using eigenvector-based feature extraction approaches are discussed. The performance of the discussed methods is compared in table 6.

From table 6, among all the compared AD classification approaches, it can be observed that the highest performance is achieved by Ramaniharan *et al.* [141], with an accuracy rate of 93.37%, the sensitivity of 93.37%, and the specificity of 93.37%.

Apart from all the discussed eigenvector-based feature extraction techniques for AD classification, Factor Analysis (FA) also plays a major role in selecting the relevant features. The FA is a widely used statistical technique used for describing the inconsistency among experiential variables in terms of lesser overlooked variables known as the factors [147]. In a literature [147], D. Salas-Gonzalez, *et al.* described a factor analysis based feature selection for AD classification. For all the input PET images, initially, the voxels for classification are selected by using a t-test based approach. All the selected voxels are then modeled using the FA to reduce the dimensionality. Three various techniques are used for classifying the subjects are; 2 multivariate Gaussian mixture models and SVM with a linear kernel. According to the performance analysis, SVM with linear kernel produced the maximum classification accuracy.

F. HARMONIC FUNCTION BASED FEATURE EXTRACTION

For extracting the most discriminative features from images, Harmonic Analysis provides pixel-wise smooth curves in the frequency domain denoted by amplitude and the phase [148]. From frequency curves, the most discriminative features are extracted from functional associations among the spectral bands. The harmonic descriptor is a collection of information that helps to define a given shape [149]. One advantage of harmonic descriptors is that it is invariant to the luminousness. Moreover, because of the polynomial nature, harmonic descriptors can determine the smooth variations in the image signal [150]. Some of the literature where Harmonic function based feature extraction is performed for AD classification are discussed below.

A classification framework to classify AD vs CN, MCI vs CN, and AD vs MCI subjects on structural MR images is proposed in the literature [151]. The circular harmonic function descriptors on the hippocampus and posterior cingulate cortex are used. The Automated Anatomical Labeling (AAL) based atlas is used to select 2 Region of Interests (RoIs), namely, the hippocampus, and the posterior cingulate cortex. For segmenting the Region of Interests (RoIs), the Statistical Parametric Mapping (SPM)8 toolbox is used. The authors

TABLE 6. Performance comparison of the AD classification techniques using eigenvector-based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
Anandh Kilpattu Ramaniharan, et al. [141]	AD vs. CN: 93.37%	AD vs. CN: 93.37%	AD vs. CN: 93.37%	NA	NA	NA
L. Khedher, et al. [142]	AD vs. CN: 88.49%, MCI vs. CN: 81.89%, AD vs. MCI: 85.41%	AD vs. CN: 90.39%, MCI vs. CN: 82.16%, AD vs. MCI: 85.95%	AD vs. CN: 86.17%, MCI vs. CN: 81.62%, AD vs. MCI: 84.86%	NA	NA	NA
I.A. Illán, et al. [143]	AD vs. CN: 88.24%, MCI vs. CN: 70.21%	AD vs. CN: 87.80%, MCI vs. CN: 40.91%	AD vs. CN: 88.64%, MCI vs. CN: 83.51%	NA	NA	NA
F. Segovia, et al. [144]	AD vs. CN: 91.75%	AD vs. CN: 92.68%	AD vs. CN: 91.07%	NA	NA	NA
J.M. Górriz, et al. [145]	AD vs. CN: 84.17%	AD vs. CN: 89.29%	AD vs. CN: 90.24%	NA	NA	NA
R. Chaves, et al. [146]	MCI vs. CN: 90.72%	MCI vs. CN: 89.29%	MCI vs. CN: 92.68%	NA	NA	NA

have used circular harmonic function to select the contrasting patterns, and their coefficients form the descriptors of brain pattern. Moreover, a dense sampling approach is used for computing signal decomposition on the circular harmonic function. From each slice, 2D descriptors are extracted from the segmented RoIs. The signal differences inside the RoIs are represented as a set of local circular harmonic function coefficients. The features are then leveraged for differentiating normal and abnormal images. The shape of both the RoIs are different, hence a Balanced Weighted Voting (BVW) based approach is applied separately to cluster the extracted features, and built the visual vocabulary (codebook). Moreover, the RoI's shape varies from one projection (sagittal, axial, and coronal) to another. The clustering process is performed three times for different projections and produced a separate codebook for each projection and RoI. For reducing the resultant image signature dimension, authors have used the Principal Components Analysis (PCA) based approach.

A multidimensional classification approach to classify AD vs CN, and MCI vs CN subjects, based on the hippocampal shape, is proposed in the literature [152]. The authors have performed an automatic hippocampus and amygdala segmentation technique, based on a region growing approach. The approach comprises prior information about the hippocampus and the amygdala location, derived from a probabilistic atlas. With the help of Spherical Harmonics-Point Distribution Model (SPHARM-PDM) software, the hippocampus is labeled as a series of spherical harmonics. SPHARM provides a mathematical method for representing the surfaces with spherical topology. The method can be observed as a 3D analog of Fourier series expansion. The authors have determined 2 correspondences between the objects, namely i) SPHARM coefficients, used as features in SVM classifier, ii) SPHARM-PDM landmarks, applied for visualizing the localization of shape differences between groups. By SPHARM decomposition with the degree of 20, subjects are characterized by a feature vector of size 2646. The feature vector is determined by concatenating three coordinates of all coefficients, results in $(20 + 1)^2$ vector coefficients. Furthermore, there are two hippocampus and three spatial coordinates, thus a total of $2 \times 3 \times (20+1)^2 = 2646$ features are generated. For identifying, and selecting only the most discriminative features, a univariate feature selection method combined with a bagging strategy is used.

AD classification using harmonic function based feature extraction approaches are discussed. The performance of the discussed articles is shown in table 7.

From table 7, it can be observed that among all the discussed AD classification methods using harmonic function-based feature extraction approaches, the highest performance is achieved by Gerardin *et al.* [152], with an average accuracy of 88.5%, the sensitivity of 89.5%, and the specificity of 88%.

G. SCALE-INVARIANT FEATURE TRANSFORMS BASED FEATURE EXTRACTION

Scale Invariant Feature Transform (SIFT) is one of the most popular techniques, for feature extraction and matching of the prominent properties at various scales, amongst the set of input images, [153]. In SIFT, firstly the key features from objects are mined from some reference images and kept in a database. Then, the object in an input image is determined by associating a new image with the image from the database by using Euclidean distance-based feature vectors matching approach [154]. Among all the matching, subgroups of features, which are more suitable on the object, in terms of its location, scale, orientation, etc., are determined to find out the best matches. Lastly, the probability of the subgroups is calculated, which specifies the occurrence of an object, correctness, and the quantity of possible wrong matches [154]. One of the major advantages of SIFT is that it can produce a sufficient quantity of features, that can compactly cover the whole image [155]. Moreover, features extracted by SIFT are local, hence no segmentation is required. One of the disadvantages of SIFT is that, sometimes it produces lots of non-feasible features, hence, the process is computationally time consuming [156].

TABLE 7. Performance comparison of the AD classification techniques using Harmonic function based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
Olfa Ben Ahmed, et al. [151]	AD vs. CN: 83.7%, MCI vs. CN: 66.7%, AD vs. MCI: 76.5%	AD vs. CN: 78.8%, MCI vs. CN: 62%, AD vs. MCI: 78.9%	AD vs. CN: 85.7%, MCI vs. CN: 68.3%, AD vs. MCI: 52.8%	NA	NA	AD vs. CN: 0.82, MCI vs. CN: 0.65, AD vs. MCI: 0.66
Emilie Gerardin, et al. [152]	AD vs. CN: 94%, MCI vs. CN: 83%	AD vs. CN: 96%, MCI vs. CN: 83%	AD vs. CN: 92%, MCI vs. CN: 84%	NA	NA	NA

TABLE 8. Performance comparison of the AD classification techniques using scale-invariant feature transforms based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
Mohammad Reza Daliri, et al. [157]	AD vs. CN: 72%	NA	NA	NA	NA	NA

A classification framework to classify AD vs CN subjects, using the scale-invariant feature transforms in MR images, is proposed in the research article [157]. For detecting salient features from the images, a scale invariant approach is applied using the scale-space representation. The operation is performed using a Gaussian kernel, at different variances, and then convolved with the original image. The Difference-of-Gaussian (DOG) is calculated by differencing adjacent images in the scale-space. The detected prominent points in an image are represented as a feature vector, that is used for local shape description, such as location, scale, and orientation. Samples are used for creating the orientation histograms, over 4×4 sample regions. The final feature descriptor for each salient feature has 128 dimensions while using 8 directions for each histogram. For the unique representation of the salient points of all images, all the SIFT descriptors are grouped into a fixed number of clusters (K clusters) and extracted the common salient points from training images. The K- means clustering technique is used to minimize the Within-Cluster Sum of Squares (WCSS).

Some of the AD classification approaches, using SIFT based feature extraction, are discussed in this section. The performance of the discussed articles is presented in table 8.

From table 8, it can be observed that the average performance of the method is approximately 72% only, which is less compatible with respect to other methods.

H. ARTIFICIAL NEURAL NETWORK (ANN) BASED FEATURE EXTRACTION

ANN is a sequence of methodologies, that helps to determine the fundamental associations in a group of data, by following a similar procedure of the human brain [158]. Because of the influential matching determination, ANN is a commonly used feature extraction or dimensionality reduction tool. The major advantage of ANN in feature extraction is that it has the capability to learn by itself, and provides results that are not restricted to the initial information provided in [159]. In a neural network, information is stored in the networks rather than stored in a database, hereafter, the data-loss doesn't distress its operation. One of the major disadvantages of ANN is the mysterious behavior of the network. After producing a solution, ANN never gives a hint of how and why the result is, which may reduce the confidence in the network [160]. Some literature, where ANN-based feature extraction is performed for AD classification, is discussed below.

A classification method, to classify AD vs CN, and MCI vs CN subjects, is discussed in the paper [161]. For intra-slice features extraction, authors have proposed a 2D-CNN model for learning the features, invariant to the simple alterations and linear changes. The 2D-CNN model is designed for each group of slices, and slices from the same group are used for training the model. The model is composed of 5 convolutional layers, namely 2 max pooling layers, 2 full connection layers, and 1 SoftMax classification layer. For inter-slice features extraction from each group of slices, the authors have proposed a stacked Recurrent Neural Networks (RNNs). In order to achieve more inter-slice features, the Bidirectional Gated Recurrent Unit (BGRU) is applied, which has a forward GRU and a backward GRU. In the classification step, the features generated from the BGRU network layer, followed by 2 fully connected layers, and 1 SoftMax layer, are jointly optimized. The individual CNNs as well as the BGRU combination models (CNN-GRU) are trained separately for axial, sagittal, and coronal axis.

A Convolutional Neural Network-based, MR image analysis, for AD classification is proposed in the literature [162]. In this study, the hippocampus is considered as a Region of Interest (RoI). An automated patch-based separation technique with geometric coordinates of the International Consortium for Brain Mapping (ICBM) template, is applied for extracting the RoI. The Local Entropy Minimization with a bi-cubic Spline (LEMS) model is used for noise removal and intensity homogeneity correction. The 1^{st} layer of the network is the Convolution layer, which is used for extracting properties from the input images. The features are mined by preserving association amongst the pixels of learning features, using trivial squares of the input data. The main task of max pooling is to consider only major components from the rectified feature map, for reducing the unnecessary parameters. For considering major components, the algorithm also used the concept of average pooling.

A short description along with some of the research articles on AD classification using ANN based feature extraction approaches is discussed. The performance of the research articles is presented in table 9.

TABLE 9. Performance comparison of the AD classification techniques using Artificial Neural Network (ANN) based feature extraction approaches.

Authors	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under Curve (AUC)
Manhua Liu, et al. [161]	AD vs. CN: 91.2%, MCI vs. CN: 95.3%	AD vs. CN: 91.4%, MCI vs. CN: 78.9%	AD vs. CN: 91.0%, MCI vs. CN: 80.0%	NA	NA	AD vs. CN: 0.95, MCI vs. CN: 0.84
Boo-Kyeong Choi, et al. [162]	AD vs. CN: 92.3%, MCI vs. CN: 85.5%, AD vs. MCI: 78.1%	AD vs. CN: 93.3%, MCI vs. CN: 88.4%, AD vs. MCI: 77.0%	AD vs. CN: 91.1%, MCI vs. CN: 82.8%, AD vs. MCI: 79.3%	AD vs. CN: 91.4%, MCI vs. CN: 93.9%, AD vs. MCI: 80.2%	AD vs. CN: 93.2%, MCI vs. CN: 88.1%, AD vs. MCI: 76.0%	NA

From table 9, between two relevant research articles on AD classification using ANN based feature extraction approaches, the maximum performance is achieved by Liu *et al.* [161] with an average accuracy of 93%, the sensitivity of 85.15%, and specificity of 85.5%.

III. RESULTS AND DISCUSSION, AND COMPARISON ON DIFFERENT AD CLASSIFICATION METHODS

In this study, a total of 50 recently published articles, on AD classification, using different feature extraction approaches, are reviewed, and compared with their performances. It can be observed from this study that, feature extraction plays a major role in the classification of AD using brain images. A detailed comparison based on the performances of different classification methods is presented from table 2 to table 9. To make the performance comparison easier, we have analyzed the average classification performance by averaging the results of all the performance parameters (i.e, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and the area under the curve) for all subject groups. The graphical representation of the performance comparison is shown in figure 3, and figure 4.

From table 2 to table 9, as well as from the graphical representation, it can be observed that overall maximum performance is achieved by Mesrob *et al.* [96]. While classifying AD vs. CN subjects, the highest accuracy (99.60%) can be observed in the same literature by Mesrob *et al.* [96]. In this approach of AD classification, the authors have parcellated the MRIs into anatomical Region of Interests (RoIs), with the help of pre-labeled templates. Next, from each of the RoIs, Diffusion Tensor Imaging (DTI) measures, as well as the absorption of grey matter, are mined. All the subjects (AD/CN) are acquired as per the guidelines provided by the National Institute of Neurological and Communication Disorders and Stroke/AD and Related Disorders Association (NINCDS-ADRDA). For all the subjects, T1-MRI is considered for the study, using the Spoiled Gradient echo Sequence (SPGS) and Diffusion Tensor Imaging (DTI) scans (in 23 directions). From the Diffusion Tensor Imaging (DTI) inputs, tensors are determined and then, an Apparent Diffusion Coefficient (ADC), as well as the Fractional Anisotropy (FA) maps, are mined as shown in figure 5. All the structural inputs are segmented into three parts (grey matter, white matter, cerebrospinal fluid) as shown in figure 6. Using the Montreal Neurological Institute (MNI) toolbox, a common minimum volume for all the inputs is calculated, and using that mask 73 regions in the brain are parcellated as shown in figure 7. Next, the average of the Apparent Diffusion Coefficient (ADC) is determined for every RoIs and the multimodal features are determined from the proportion of Apparent Diffusion Coefficient (ADC) to grey matter absorption in each voxel. Thus, the most discriminative features are extracted.

Though the performance claimed by Mesrob *et al.* [96] is convincing, but the experiment is performed by taking only 32 subjects, also the authors have excluded many important cortical regions from the study which may impact in overall classification performances [163].

Xiao *et al.* [80] have claimed that, while classifying AD vs. MCI, the method achieved 100% of sensitivity, MCI vs. CN with 100% specificity, MCI vs. CN with 100% of PPV, and AD vs. MCI with 100% of NPV. Altaf *et al.* [81] have claimed that the classification mechanism achieved a 100% of sensitivity while classifying AD vs. CN. One of the common issues in these two articles [80], [81] is that they have used GLCM based feature extraction technique, which is computationally expensive due to the presents of many zero elements which may not be necessary for further processing [78]. Gupta *et al.* [102] have claimed that, while classifying AD vs. MCI subjects, the method achieved a 100% of sensitivity. Also, the method classified AD vs. CN, and MCI vs. CN, with 100% specificity. One of the major problems with the method [102] is that, for atrophy measurement in brain cells, they have used a free-surfer toolbox, which introduces some unnecessary biases [164]. Khazaee *et al.* [128] have claimed that they achieved 100% of sensitivity while classifying MCI, and 100% specificity as well as 100% PPV, while classifying AD vs. CN. One of the major issues in the study [128] is that the authors have used an atlas-based Region of Interests (RoIs) extraction approach, where no boundary information is present, which may affect the overall performance in classification [165]. In similar research, Ahmed *et al.* [93], while classifying AD vs. CN, have claimed that the proposed method can achieve a 100% of classification specificity. One of the main issues in the approach [93] is that the authors have extracted the hippocampus using an atlas based method, where no boundary information is present [165]. While classifying AD vs. CN subjects, the highest Area Under Curve (99.93%) is achieved by Beheshti *et al.* [98]. One major limitation of the approach is that the authors have used a

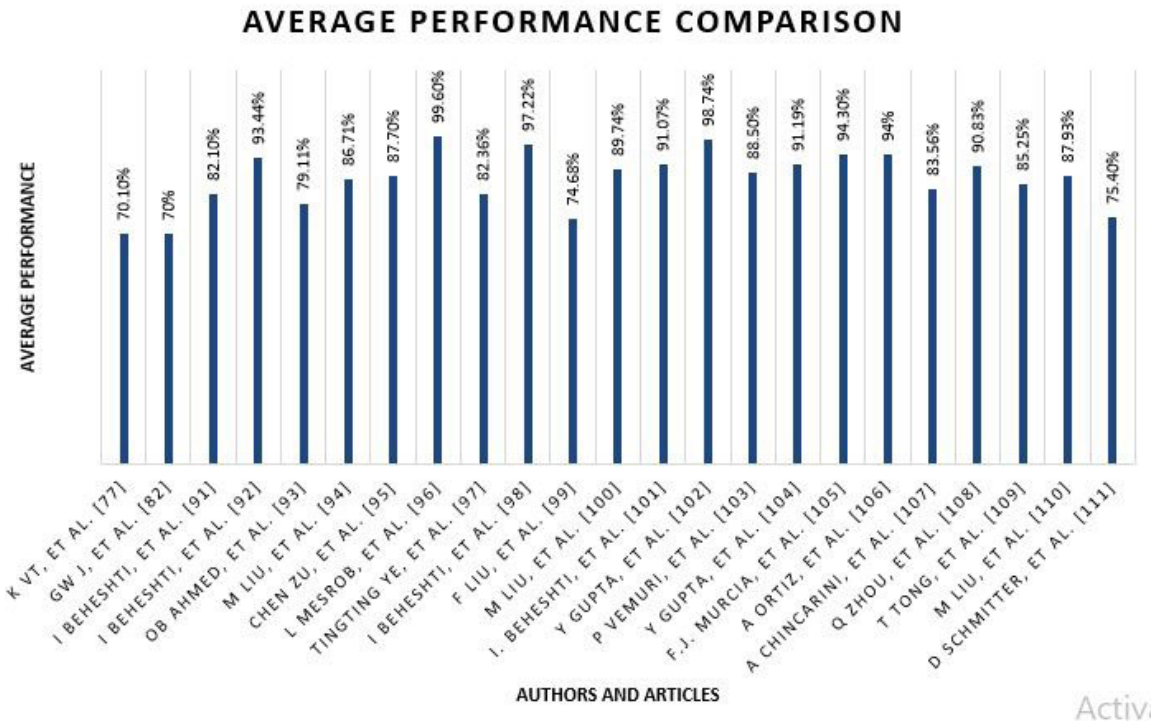

FIGURE 3. Average performance comparison of 50 AD classification approaches (Part 1).

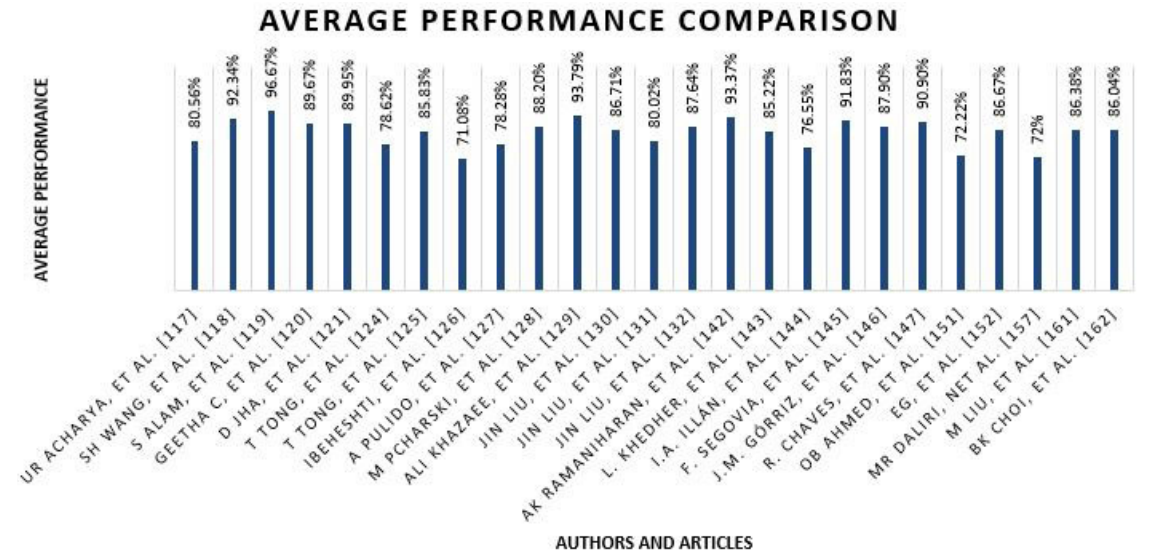

FIGURE 4. Average performance comparison of 50 AD classification approaches (Part 2).

PCA-based feature selection technique, which cannot explore the spatial information [166].

For all the discussed literature, we have analyzed the detailed observations. The detailed observations are summarized in table 10.

From table 10, it can be observed that, although all the research articles perform well in their own way, still there lies few scopes for further improvements. From the observations, it can be noticed that, among all the popular feature extraction/selection approaches, Principal Component Analysis (PCA) and GLCM have been used in several literature [80], [81], [91], [98], [104], [119], [121], [142], [143], [145]. Although these two techniques are more popular, they have some limitations which may affect on the final classification outcomes. One limitation of GLCM is that, it is a sparse matrix with many elements as zero, which are unnecessary for texture features calculation, hence, it is computationally expensive [78] [79]. Similarly, one of the limitations of PCA is that, it cannot explore the spatial information from an input image [166].

TABLE 10. Summarization of all the discussed AD classification approaches.

Authors	Year	Publication	Dataset	Classifier	Observations
Prashanthi Vemuri, et al. [103]	2008	NeuroImage: Clinical	Mayo Clinic Alzheimer's Disease Research Center (ADRC)	SVM	1. The selection of RoI is done based on a manual template-based approach. Manual selection of RoI required expertized knowledge. Moreover, it is time consuming and the final RoI may contain some unwanted pixels as well [167]
E. Gerardin, et al. [152]	2009	NeuroImage	Japanese-ADNI (JADNI)	SVM	1. The authors have used the hippocampus as the RoI and used a region growing approach for segmentation. Because of the complex structure, region growing cannot segment the hippocampus accurately [168]. 2. For the experiment, the authors have taken 23 AD patients, 23 MCI patients, and 25 CN patients, which is very less in number.
Andrea Chincarini, et al. [107]	2011	NeuroImage: Clinical	ADNI	SVM	1. The authors have extracted only 9 VoIs/RoIs for the study. However, no clear justification has been given why and which RoIs/VoIs are considered for extraction.
I. A. Illa´n, et al. [143]	2011	Information Sciences	ADNI	SVM	1. The authors have used PCA and a FastICA based approach for extracting the features, but PCA sometime fails to explore the spatial information [166], and FastICA often fails to extract the local independent features [169].
J. M. Go´rriz, et al. [145]	2011	Applied Soft Computing	Virgen de las Nieves hospital in Granada (Spain)	SVM	1. The authors have used the GMM model for density estimation which is based on a false assumption that each pixel is independent of its neighbors [170]. 2. For feature reduction, authors have used the PCA method which ignores the spatial information [166]
M. Reza Daliri, et al. [157]	2011	Journal of Medical Systems	OASIS	SVM	1. The authors have compared their methodology only with one previous related work, which is very less in number. 2. The proposed method will be computationally expensive as SIFT takes more time if the dataset is large [171]. 3. The authors have not performed any noise filtering technique, and used Histogram based approach for feature extraction, which is noise sensitive [172].
L. Mesrob, et al. [96]	2012	Advances in Molecular Imaging	Research and Resource Memory Centre, France	SVM	1. Only 15 AD subjects and 17 CN subjects are used in this study which is very small in number. 2. The proposed approach excluded approximately 17 cortical RoIs. Cortical regions play an important role in finding the discriminated areas [163].
F. J. Mart´ınez Murcia, et al. [105]	2012	Expert Systems with Applications	Irgen de las Nieves hospital, in Granada (Spain)	SVM	1. The authors have used the FBP for input image reconstructions which is sensitive towards star effect, and sometimes falsely it produces the streak-like artifacts in the image [173]. 2. The authors have manually set the parameter that pixels having the intensity value lower than 70 will be discarded. However, the pixel values may differ image to image, hence an automatic threshold value selection might increase the accuracy in pixel selections. 3. The authors have used the MWW U-test for feature rankings. The basic assumption of the MWW U -test is that the sets of input data have the similar distributions [174], which may not be true for all the cases.
R. Chaves, et al. [146]	2012	Expert Systems with Applications	ADNI	SVM	1. Voxel selection is done using Apriori algorithm which sometimes selects some unnecessary voxels and hence the algorithm is time consuming [175][176].
A. Pulido, et al. [126]	2013	The International Society for Optical Engineering	OASIS	SVM	1. The authors have completely ignored the pre-processing step. Pre-processing steps like skull stripping, denoising, intensity enhancement, etc. may help to improve the classification accuracy [177].
F. Segovia, et al. [105]	2013	Expert Systems with Applications	Virgen de las Nieves hospital in Granada (Spain)	SVM	1. After performing the OOB operation, the authors have used only 10 components from each subject for further classification without a proper justification or comparison why they have chosen only 10 components and whether increasing/decreasing the components will affect the classification accuracy or not. 2. The authors have not provided any information about the number of subjects used in the experiment. However, they have mentioned that 95 SPECT images are acquired, which is too less in numbers.
G. W. Jiji, et al. [82]	2014	Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization	Radiology Research Hippocampus Segmentation Database	SVM	1. For this study, the authors have acquired only 15 subjects, which is very less for the experiment. 2. The authors have used a level set-based approach for the hippocampus segmentation, which is the one and only RoI for the study. However, one of the disadvantages of the level set based segmentation approach is that here the edge-stopping function is never exactly zero at the edges, and so the curve may eventually pass through object boundaries [178].

TABLE 10. *(Continued.)* **Summarization of all the discussed AD classification approaches.**

OB. Ahmed, et al. [93]	2014	Multimedia Tools and Applications	ADNI and Bordeaux dataset	SVM, and Bayesian classifier	1. The authors have used considered the hippocampus as a VoI, and used an atlas-based approach for extracting the VoI, and hence proper extraction is required for better performance. The main disadvantage of atlas-based approaches is the lack of boundary information [165], which may lead to an inappropriate segmentation of the VoI.
Feng Liu, et al. [99]	2014	NeuroImage: Clinical	ADNI	SVM	1. The brain is partitioned into 93 regions for feature extractions. The authors have not provided any information about the selection criteria of the 93 regions. Selection of proper RoIs plays an important role and a machine learning based approach may select the RoIs more accurately [134]
M. Liu, et al. [100]	2014	Human Brain Mapping	ADNI	SVM, and Vector basd clasifier	1. The authors have used the t-test based approach for extracting the important patches, which is sensitive towards Type I error [179].
A. ORTIZ, et al. [106]	2014	KES Innovation in Medicine and Healthcare	ADNI	Sparse Representation Classifiers (SRC)	1. The authors have manually set the parameter that pixels having the activation value below 50% will be discarded. However, the pixel values may differ image to image, hence an automatic threshold value selection may increase the accuracy in pixel selections. 2. For feature ranking, the authors have used a t-test approach, which is sensitive towards Type I error [179].
Qi Zhou, et al. [108]	2014	IEEE Transactions on Biomedical Engineering	Wien Center for Alzheimer's Disease and Memory Disorders	SVM	1. The authors have extracted only 52 VoIs/RoIs for the study. However, no clear justification has been given why and which RoIs/VoIs are considered for extraction. 2. For feature ranking, the authors have used a t-test approach, which is sensitive towards Type I error [179].
T. Tong, et al. [123]	2014	Medical Image Analysis	ADNI	SVM	1. The intensities of patches are used as features without performing denoising operation, which may lead to the alteration to intensity inhomogeneity, or other forms of noise. 3. The trained classifier used by the authors is a bag-level classifier which can only predict the label of unseen images. Moreover, the approach can't assign a probabilistic label for each voxel. To improve the classification accuracy, instance level classifiers may be used to assign a disease score for every voxel [180].
M. Liu, et al. [94]	2015	Human Brain Mapping	ADNI	SVM	1. The authors have proposed a multi atlas-based image registration approach, where multiple atlases are used, and hence this approach may computationally expensive [181]. 2. The proposed method ignored the longitudinal discriminative information. Longitudinal information can play a major role in classification of AD [182].
I. Beheshti, et al. [101]	2015	Computers in Biology and Medicine	ADNI	SVM	1. The authors have used DARTEL registration technique where some diffeomorphic formations for which velocities varies over time are impossible to get [183]. 2. The authors have used the PLS feature extractor approaches. One of the disadvantages of using the PLS is that, it overestimates and extracts some unnecessary features [184].
M. Liu, et al. [110]	2015	IEEE Transactions on Biomedical Engineering	ADNI	SVM	1. The authors have used the OVA approach for feature selection. OVA is computationally expensive [185]. 2. Authors have used the naïve multi-task sparse feature selection method with a $l_{2,1}$ norm based regularizer, where relationships among subjects are not considered at all. 3. The authors have extracted regional features in multiple template spaces, where the partitions of RoIs in different templates may be different from each other. However, it is difficult to directly compare subjects in two template spaces because of anatomical structure differences among the templates.
Daniel Schmitter, et al. [111]	2015	NeuroImage: Clinical	ADNI	SVM	1. For this study, brain morphometry is measured by several software which required expertized knowledge. An automatic brain morphometry measurement technique can help in better classification accuracy.
Jin Liu, et al. [130]	2015	IEEE/ACM Transactions on Computational Biology and Bioinformatics	ADNI	SVM	1. The authors have used AAL method for s electing the RoIs that have some disadvantages, such as: Poor boundary information which may lead to a false selection of a region [186]. 2. In this study, Cerebellar regions is totally ignored which may contain some important features for classification.

TABLE 10. *(Continued.)* **Summarization of all the discussed AD classification approaches.**

L. Khedher, et al. [142]	2015	Neurocomputing	ADNI	SVM	1. The authors have used the PCA, and the PLS feature extractor approaches, but PCA cannot explore the spatial information from an input image [166], and PLS overestimates and extracts some unnecessary features [184].
OB. Ahmed, et al. [151]	2015	Computerized Medical Imaging and Graphics	ADNI	SVM	1. The authors have used Automated Anatomical Labeling (AAL) for selecting the RoIs which has a poor boundary information, hence sometimes it may lead to a false selection of a region [186].
I. Beheshti, et al. [92]	2016	computer methods and programs in biomedicine	ADNI	SVM	1. The authors have used DARTEL registration technique which uses the concept of the fixed velocity field, hence some diffeomorphic formations for which velocities varies over time are impossible to get [183]. 2. The authors have used 7 feature ranking techniques to rank the best features which may leads the process to become complex and take more time to execute. Instead of that the authors could have used a single ranking-based feature selection approach such as the deformation-based analysis [187].
C. Zu, et al. [95]	2016	Brain Imaging and Behavior	ADNI	SVM	1. The method proposed by the authors requires equal number of features from all the modalities, but some modalities in ADNI, such as Cerebrospinal Fluid and genetic data, have dissimilar feature numbers which may have important pathological information. 2. The proposed method ignores the longitudinal data which may have some important information for classification.
T. Ye, et al. [97]	2016	Brain Imaging and Behavior	ADNI	SVM	1. The authors have selected 93 RoIs using atlas wrapping. Selection of proper RoIs plays an important role and an automatic approach can select the RoIs more accurately [134].
I. Beheshti, et al. [98]	2016	Magnetic Resonance Imaging	ADNI	SVM	1. For feature selection, the authors have used the PCA method which cannot explore the spatial information from an input image accurately [166], and for feature ranking the t-test method which is sensitive towards Type I error [179].
Shui-Hua Wang, et al. [118]	2016	Multimed Tools Appl	OASIS	Multilayer Perceptron (MLP)	1. The authors have used BBO for network training, where no provision to select the best members from each generation to improve the classification accuracy [188]. 2. For single slice feature selection, the authors have used the ICV in only one direction (axial). Since the image is in 3D, the authors could have been used the ICV for all the 3 directions axial, sagittal and coronal for better performance.
T. Tong, et al. [124]	2016	Pattern Recognition	ADNI	Random Forest	1. The authors have focused only in the cross-sectional study. Longitudinal data information may help to increase the performance accuracy [189].
Maciej Plocharski, et al. [127]	2016	Computer Methods and Programs in Biomedicine	ADNI	SVM	1. The authors have ignored the pre-processing steps like skull stripping, denoising, intensity enhancement, etc. which may help to improve the classification accuracy [177].
A. Khazaee, et al. [128]	2016	Behavioural Brain Research	ADNI	SVM	1. The authors have used 264 RoIS from each subject for further classification without a proper justification or comparison for considering those particular RoIs. 2. The authors have used an atlas-based approach for extracting the RoIs which has the lack of boundary information [165], that may lead to an inappropriate segmentation of the RoIs.
Anandh Kilpattu Raman-iharan, et al. [141]	2016	Expert Systems with Applications	OASIS	KNN, SVM, Naïve Bayes	1. For the 3D images, among 176 slices, the authors have selected manually slice number 90 for further processing. Proper selection of slice is important and required advance expert knowledge [190]. 2. The authors have acquired the data of only 60 subjects (30 for AD, 30 for CN), which is very less in number.
Zhe Xiao, et al. [80]	2017	Computational and Mathematical Methods in Medicine	ADNI	Suport Vector Machine (SVM)	1. The authors have used GLCM for texture feature extraction, but one of the drawbacks of GLCM is that it is a sparse matrix with many elements as zero which are unnecessary for texture features calculation hence it is computationally expensive [78] [79]. 2. The authors have not used any noise removal techniques and then used SVM-REF feature selection method which is: i) Noise sensitive [191] [192]. ii) While selecting the best features, it can't avoid the highly correlated discriminative features effectively [191]. iii) Can't effectively deal with group-based feature selection [191].

TABLE 10. *(Continued.)* **Summarization of all the discussed AD classification approaches.**

I. Beheshti, et al. [91]	2017	Computers in Biology and Medicine	ADNI	SVM	1. Spatial information plays an important role in cluster and classification analysis of an image. For feature reduction, authors have used the PCA method. One of the disadvantages of PCA is that, it cannot explore the spatial information from an input image [166]. 2. For feature ranking, the authors have used a t-test approach, which is sensitive towards Type I error, i.e. if the variances or the group sizes are paired negatively then it may lead to incorrect results [179].
Tong Tong, et al. [109]	2017	IEEE Transactions on Biomedical Engineering	ADNI	SVM	1. The authors have not used any noise filtering approach and then used the EN approach for feature selection which has several demerits such as: it is noise sensitive, sometimes it selects some unnecessary features, etc [193].
Saruar Alam, et al. [119]	2017	Journal of Healthcare Engineering	ADNI, and OASIS	Twin SVM	1. For feature reduction, authors have used the PCA method which cannot explore the spatial information from an input image [166]. 2. For converting the 3D images to 2D, authors have analysed the information of the slices manually. An automatic slice selection technique may help to select more appropriate slices.
D. Jha, et al. [121]	2017	Journal of Healthcare Engineering	OASIS	Feed Forward Neural Network (FFNN)	1. The authors have converted the 3D images to 2D, and have taken 32 slices manually for each subject for further processing which required expertise knowledge. 2. For feature reduction, authors have used the PCA method that cannot explore the spatial information [166].
I. Beheshti, et al. [125]	2017	Journal of Alzheimer's Disease	Japanese-ADNI (JADNI)	SVM	1. The authors have proposed to construct the brain network but avoid to extract the complex network measures from individual male and female subjects, which may lead for an incorrect network construction, as male and female subjects have different network structures [194].
Jin Liu, et al. [129]	2017	IEEE Transactions on Nano Bioscience	ADNI	SVM	1. The authors have used F-Score to select the classification features which may not always select the best features as it fails to expose the mutual information among features. 2. The authors have used Automated Anatomical Labeling (AAL) for selecting the RoIs which have the Poor boundary information so it may lead to a false selection of a region [186].
Jin Liu, et al. [131]	2017	IEEE/ACM Transactions on Computational Biology and Bioinformatics	ADNI	SVM	2. The authors have used AAL for selecting the RoIs where no provision is there to define the number of voxels in a region [186]. 3. The authors have used F-Score to select the classification features that fail to expose the mutual information among features.
Tooba Altaf, et al. [81]	2018	Biomedical Signal Processing and Control	ADNI	SVM, ensemble, KNN, and decision tree	1. The authors have used the GLCM, SIFT, LBP, HOG, and BOW based feature extraction techniques. The proposed method maybe computationally expensive as SIFT takes more time if the dataset is large [171], GLCM contains many zero elements [78] [79], and LBP extracts more irrelevant features as the number of neighbours increased [195].
Geetha C, et al. [120]	2018	Biomedical Research	Harvard Medical School, OASIS, and ADNI	Fuzzy Neural Network (FNN)	1. The authors have not provided information of the number of subjects taken in this study, but revealed that a total of 160 MR images are acquired for this study, which is very less in number.
M. Liu, et al. [161]	2018	Frontiers in Neuroinformatics	ADNI	CNN-GRU (CNN-Gated Recurrent Units)	1. The authors have used GRU based inter-slice feature selection as well as in the final classification technique. However, GRU is computationally expensive as the model has problems with its slow convergence rate and low learning efficiency [196]
K.Vaith-inathana, et al. [77]	2019	Journal of Neuroscience Methods	ADNI	Random forest, SVM, and KNN	1. The authors have used 3 feature selection approaches namely, Fisher score, Elastic net regularization, and the SVM Recursive Feature Elimination (SVM-RFE) technique. The proposed classification technique may be computationally expensive as all the 3-feature selection algorithms have their own disadvantages, such as: - The fisher score fails at creating low-dimensional classifiers with decent predictive performance and low variable acquisition costs [197]. - The SVM-RFE is noise sensitive [191] [192], and can't effectively deal with group-based feature selection [191].

TABLE 10. *(Continued.)* **Summarization of all the discussed AD classification approaches.**

Y. Gupta, et al. [102]	2019	Journal of Healthcare Engineering	National Research Center for Dementia (NRCD)	SVM, KNN, Naïve Bayes, and SoftMax classifier	1. For cortical thickness atrophy measurement, the authors have used the FreeSurfer toolbox. One of the disadvantages of FreeSurfer while measuring the cortical thickness is that it introduces biases that could require statistical adjustments [164].
Yubraj Gupta, et al. [104]	2019	PLOSONE	NRCD	SVM, Random Forest, and KNN	1. For feature reduction, authors have used the PCA method. One of the disadvantages of PCA is that, it cannot explore the spatial information from an input image [166]
U. R. Acharya, et al. [117]	2019	Journal of Medical Systems	University of Malaya Medical Centre & the Harvard Brain Atlas.	KNN	1. For feature selection, the authors have used a t-test approach, which is sensitive towards Type I error [179]. 2. The authors have used KNN method which has some challenges such as, to determine the optimum k-value [198]. KNN follows a distance-based learning, where it is not clear that which type of distance can give a better performance [199].
Boo-Kyeong Choi, et al. [162]	2020	Current Medical Imaging	ADNI	CNN	1. The authors have considered only hippocampus as the RoI for classification and performed a manual segmentation using 3D Slicer toolbox. Manual segmentation required expertized knowledge [200], and hence, a machine learning based automatic segmentation of hippocampus might have been helped in getting better classification results.

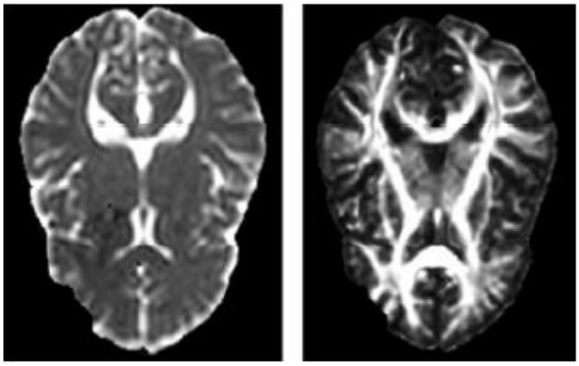

FIGURE 5. Sample Apparent Diffusion Coefficient (ADC) and Fractional Anisotropy (FA) images [96].

A. CURRENT TRENDS IN THE FIELD OF AD CLASSIFICATION USING BRAIN IMAGES

In this paper, we have discussed about the different feature extraction based AD classification techniques and brief comparisons amongst the techniques are presented in several tables. However, in recent years, multi-modal techniques have been used widely in AD classification [201]. One of the major advantages of using the multi-modal approaches is that these methods are competent in wrapping multimodal neuro-imaging features together which requires fewer labeled data as well [202]. Some of the recently published articles on multi-modal techniques for AD classification are discussed below.

For MCI Diagnosis, a multiview feature learning with multiatlas based functional connectivity networks model is proposed in the literature [203]. The authors first implemented a 3-step transformation-based technique on Automated Anatomical Labeling (AAL) template for generating a personalized atlas. For the transformation of resting-state fMRI (rs-fMRI) data into the MNI template space, a dynamical registration approach is used. The authors have used the deformation-based field to differentiate the dynamic mapping and then applied the Affinity-Propagation (AP) clustering mechanism to produce a group of personalized atlas instances for extracting the regional mean-time series and constructing the numerous Functional Connectivity Networks (FCNs) for all the subjects. For feature extraction, the graph-theory-based technique is used. For feature selection, Sparse Group Lasso (SGL) based approach is used. The authors have proposed a multi-task learning mechanism to optimize the multi-view characteristics and to train the SVM for a proper NC vs MCI classification and achieved a convincing result.

A multimodal AD classification approach is proposed in the literature [204]. The authors proposed a hyper graph-based multi-task attribute assortment model. The hyper graph-based regularization for the proposed method is designed for unambiguous illustration of the association in all the modalities, such as MRI and PET. The proposed multimodal classification model is followed by two major steps, namely hyper graph creation, and the hyper graph-based multi-task feature learning. A separate hyper graph for all the modalities is constructed by using the concept of multiple hyper-edges which imitate the associations among all the subjects. The l2,1 normalization method is applied for selecting the features jointly from the same brain area at the same moment. Finally, the authors used a multi-kernel SVM for combining the selected characteristics to perform the classification.

For the classification of MCI disease, strength and similarity guided group-level brain functional network based approach is proposed in the literature [205]. For preserving accurate analytical group dissimilarity, the authors have proposed to explore the functional connectivity (FC) properties into a group sparse representation (GSR) based network model. For reducing inter-subject inconsistency, a population-based prior-constrained graphical Lasso is designed for which the sparsity formation is imposed across every subject. The inter regional couple-wise FC is computed by determining the temporal synchronization of blood-oxygen-level-dependent (BOLD) signals using Pearson's

FIGURE 6. Sample structural MRI segmented in grey matter, white matter and cerebrospinal fluid [96].

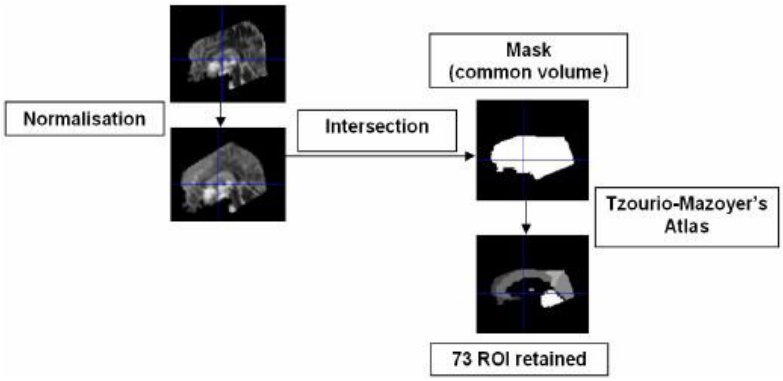

FIGURE 7. The common volume mask and brain parcellation in 73 RoIs [96].

correlation (PC) approach for all the individuals. The PC-based FC information is then used for guiding the group-level brain network for all the subject groups. Additional to the PC-based network, which is referred to as low-order FC (LOFC), the authors have also proposed to find out high-order FC (HOFC) by estimating the LOFC correlation. This will be further helpful for the GSR-based network model which is capable of incorporate LOFC and HOFC together into the same GSR-based network model, namely Strength- and Similarity-Guided GSR (SSGSR).

For early dementia diagnosis, a multi-modal latent space is proposed in the literature [206]. The authors have proposed a new AD classification mechanism by using the concept of multi-modal latent space and ensemble SVM classifier. For exploiting the association among all the modalities, the authors anticipated the ROIs-based features into a latent space. Different modalities of neuroimages are acquired such as MRI and PET data, and then all the features are projected to a regular latent space. The latent representations are mapped into several label spaces to learn numerous diversified classifiers, and an ensemble approach is used to deal with the heterogeneity of AD progression. Latent space learning as well as the classifier training is then incorporated into an integrated framework to make all the components work together.

By considering multi-model images, a brain connectivity based model for the prediction of AD is proposed in the literature [207]. The authors developed a brain connectivity model based on different modalities of images such as the MRI and PET to determine the morphological as well as the metabolic relations for all the subjects. The brain areas with remarkable variances are marked for each modality and trained to categorize the subjects by training the model from a large dataset consisting of CN and AD subjects. Brain areas having vigorous structural as well as metabolic relations with target regions are identified and then a multi-task sparse-based regression framework is used for determining the connectivity while treating the connectivity mining of all the image modalities in the target dataset as a single assignment.

A summary of the discussed multi-modal techniques is presented in table 11.

Apart from multi-modal techniques, deep learning based AD classification approaches is also using widely. Some of the major advantages of using deep learning based approach are; 1) it can extract the hidden features from the data, 2) it can produce convincing results, even if the data are unstructured, 3) it allows parallel processing, etc [208]. Some of the recently published articles on the deep learning based AD classification are discussed below.

Using rs-fMRI and Residual Neural Networks, a deep learning based AD classification approach is proposed in the literature [209]. Upon the pre-processed images, the authors have performed the training operation based on 3 ResNet-18 networks, namely the 1-Channel ResNet (1CR), Off-the-Shelf (OTS), and the Fine-Tuning (FT) for classifying the various stages of AD. All the input images are resized to 224 $\times$ 224 pixels in order to match with the pre-trained network's input size. Initially, the learning rate for the model is set to 0.001, and then in each iteration (up to 25,000 iterations), the rate is decremented by 10%. The Gamma value,

TABLE 11. Summarization of some recent studies on multi-modal brain network-based studies in AD classification.

Authors	Year	Publication	Dataset	Methods
Zhang, et al. [203]	2020	IEEE TRANSACTIONS ON CYBERNETICS	ADNI	Multiview feature learning method with multi atlas-based Functional connectivity networks
Wei Shao, et al. [204]	2020	Computerized Medical Imaging and Graphics	ADNI	Hypergraph based multi-task feature selection method
Yu Zhang, et al. [205]	2019	Pattern Recognition	ADNI	Individual functional connectivity (FC) information and the Group Sparse Representation (GSR) -based network
Tao Zhou, et al. [206]	2020	Medical Image Analysis	ADNI	Multi-modality latent space inducing ensemble SVM classifier-based method.
Weihao Zheng, et al. [207]	2019	Frontiers in Human Neuroscience	ADNI	Principal component-based analysis on in both structural and metabolic information

TABLE 12. Summarization of some recent studies on deep learning-based AD classification.

Authors	Year	Publication	Dataset	Methods
Farheen Ramzan, et al. [209]	2020	Journal of Medical Systems	ADNI	Using Resting-State fMRI and Residual Neural Networks
Manhua Liu, et al. [210]	2020	NeuroImage	ADNI	Multi-task deep CNN and 3D Densely Connected Convolutional Networks (3D DenseNet based method.
Nguyen Thanh Duc, et al. [211]	2020	Neuroinformatics	Chosun University National Dementia Research Center (Gwangju, South Korea)	Functional 3D independent component spatial maps, and 3D CNN-based method.
Kanghan Oh, et al. [212]	2019	scientific reports	ADNI	Volumetric convolutional neural network (CNN) model based method.
Xin Bi, et al. [213]	2020	Cognitive Computation	ADNI	Convolutional and recurrent learning along with the Extreme learning machine (ELM) based method.

momentum, and the weight decay factor for the model are initialized to 0.1, 0.9, and 0.0005 respectively. The authors have introduced a Stochastic-gradient-descent (SGD) based solver having a batch size of 32 images. The experimental results indicated that the OTS based network model produced the most convincing results.

A multi-model CNN model for joint learning hippocampus segmentation and AD classification is proposed in the literature [210]. To get quick convergence, a deep-CNN model is proposed for learning the residual functions at each of the convolution stages. Two residual blocks are constructed, where each block comprising of the 3-D convolutional layers (CLs), batch normalization (BN), Parametric Rectified Linear Unit (PReLU) activation, as well as the dropout layers. In the first residual block, a residual-function is trained by a short correlation, whereas the second residual block comprising of 2 CLs. The kernel dimension for the model is initialized as $3 \times 3 \times 3$ for each of the convolutions. All the trained filters are then convolved with the input images and a non-linear PReLU activation, where separate feature maps are produced for all the filters. For the proposed mechanism, multi-task deep CNN carries the information of the multi-level characteristics, and a deep 3D for DenseNet is introduced to learn the characteristics from the hippocampus.

Using the rs-fMRI, a novel 3D-Deep Learning based AD diagnosis framework is discussed in the literature [211]. The linear-regression is used along with support-vector-regression, bagging-based ensemble regression model, and the tree regression model by using the concept of the group-independent-component (gIC) analysis mechanism to predict the Mini Mental State Examination (MMSE) scores for all the subjects. The rs-fMRI data are used to determine the functional 3D self-regulating module spatial maps which are then used as the characteristics to classify the subjects as well as in the regression process. For the identification of the most useful gICs and to discard noisy ICs, the automatic clustering toolbox (FSLNets) is used. For extracting the subject-specific IC time-courses as well as the IC analysis spatial maps, the concept of the 2 stages dual regression is used. The authors have applied the 10-fold cross validation algorithm to validate the performance. To classify the subject groups, subjects-specific ICA maps are used in a 3D CNN.

Based on the concept of the Volumetric Convolutional Neural Network (VCNN) and the Transfer Learning (TL), a novel AD classification method is discussed in the literature [212]. Initially, the concept of the traditional and inception module-based convolutional auto-encoder method is used for pre-training of the MRI data for all the subjects, and then the concept of the fine-tuning-based algorithm is used for building the classifier. The proposed auto-encoder model consists of the convolution layers, dropout layers, the ReLU, as well as the pooling layers. For reducing the dependency,

the GoogLeNet inception module is used. To determine the approximate spatial influences, the class saliency visualization (CSV) method is used. Since, it is more difficult to classify pMCI and sMCI, hence, to train the model more accurately, the authors used the transfer learning algorithm, for the visual presentations of the AD vs. CN classification.

For AD classification, a novel deep learning (DL) and extreme learning based method is proposed in the literature [213]. For functional brain-network categorization, 2 DL models are considered along with the extreme learning machine (ELM) boosted structure for the learning of deep regional-connectivity features as well as the deep neighboring positional characteristics. While constructing the brain network, the concept of the Pearson correlation coefficient is applied. The deep learning model comprises the convolutional layer, ReLU activation function, pooling layer, fully connected layer, and the decision layer.

The summary of the discussed deep learning-based AD classification approaches is presented in table 12.

IV. CONCLUSION AND FUTURE SCOPE OF WORK

Alzheimer's disease is one of the major death causing neurological disorders in the world. The number of AD patients is increasing significantly all over the world. The manual diagnosis system of AD by the neurologist is time consuming, and may not provide accurate results all the time. The research of AD classification using brain images has been showing promising outcomes, which is less time consuming too. Many researchers have been trying to develop a classification mechanism using brain images with fewer research issues. Feature extraction is one of the major steps for AD classification using brain images. In this paper, we have discussed and summarized the performance of several AD classification methods using brain images, based on different feature extraction approaches. Firstly, details about 8 commonly used feature extraction approaches, along with their pros and cons have been discussed, then their classification performances are presented and compared. It is observed from the performance comparison is that the average performance of AD classification using Wavelet transform-based feature extraction approaches (89.84%) is better amongst all the 8 feature extraction approaches, followed by the Voxel Morphometry (VM) based feature extraction approaches (88.26%), Eigenvector-Based feature extraction approaches (87.63%), Neural Network (NN) based feature extraction approaches (87.21%), Graph/Network-based feature extraction approaches (83.35%), Texture-based feature extraction approaches (81.38%), Harmonic function based feature extraction approaches (80.45%), and the Scale-Invariant Feature Transforms based feature extraction approaches (72%). Overall the classification approach discussed in the article cites mesrob2012dti provides the maximum performance rate (99.60%). It can also be observed from the study that, feature extraction plays a major role in AD classification, hence a proper feature extraction technique is necessary in order to achieve a better classification result.

It is observed that the AD classification method proposed in the article [96] provides the maximum performance rate among all the discussed articles. But one of the issues with the classification framework is that the authors have acquired a very less number of samples (15 AD and 17 CN subjects) from the ''Research and Resource Memory Centre of the Pitié-Salpêtrière hospital (Pitié- Salpétrière Hospital, Paris, France''. As a scope of future work, a large number of data samples from other data sources such as ADNI. OASIS, etc. can be acquired and compared the performance again. Moreover, the proposed approach excluded approximately 17 cortical Region of Interests (RoIs). Cortical regions play an important role in finding discriminated areas. A machine learning based technique can be designed which can help to select the most appropriate cortical regions. The selection of appropriate cortical regions may help to get more accurate results.

From this study, it is also observed that the Wavelet transform-based feature extraction approaches help the classifiers to achieve better performances than any other feature extraction approaches discussed. Among all Wavelet transform-based feature extraction approaches discussed, the classification method discussed in the article [119] provides the maximum performance results. The authors used a Principal Component Analysis (PCA) based approach for feature selection. One of the disadvantages of PCA is that it cannot explore the spatial information from an input image. In the future, some mechanisms for selecting spatial information can be added to get more accurate classification results. Moreover, for converting the 3D images to 2D, authors have analyzed and selected the information of the slices manually, which requires high expertise knowledge. An automatic slice selection technique may be designed in the future to select more appropriate slices.

Moreover, in the future, more recent articles on AD classification, based on some more feature extraction approaches can be compared which may help the researchers to choose a proper feature extraction technique for AD classification.

A. CHALLENGING ISSUES IN THE STUDY OF BRAIN IMAGES FOR ad CLASSIFICATION

Though researchers have been achieving promising results for AD classification using brain images, still several challenging issues lie in this field of study. One of the major issues is to get sufficient data for the study. For acquiring data, most of the researchers depend only on two online data-sets namely OASIS, and ADNI. Though these data-sets provide a large number of data, but if we do the gender-wise and age-wise distribution, which may play an important role while classifying AD from other subject groups, the data may not be sufficient to train a machine learning based model accurately.

One more major issue is to determine the proper bio-markers in AD. Sometimes, similar kind of changes in brain structure may occur due to some other neurological disorders, and consideration of those data in training or testing

sets can produce a wrong result showing 100% accuracy in the model. So, finding the exact bio-markers in brain studies which is only related to AD is a challenging task.

As we discussed, feature extraction is one of the most important steps in AD classification. The human brain is very complex in structure, and it contains a lot of information. One of the major challenges in extracting features from brain images is scalability. Since brain images contain lots of features, it is challenging to design a feature extraction technique that can handle it feasibly.

B. AUTHOR CONTRIBUTIONS

All authors are responsible for analysis, conceptualization, and writing the original manuscript. All authors have read and agreed to the published version of the manuscript.

C. FINANCIAL DISCLOSURE

This research received no external funding.

D. CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

REFERENCES

[1] I. O. Korolev, "Alzheimer's disease: A clinical and basic science review," *Med. Student Res. J.*, vol. 4, no. 1, pp. 24–33, 2014.

[2] NIH. *Alzheimer's Disease: A Clinical and Basic Science Review*. Accessed: Jul. 13, 2020. [Online]. Available: https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet

[3] A. Association. *Alzheimer's Disease Fact Sheet*. Accessed: Jul. 13, 2020. [Online]. Available: https://www.alz.org/in/dementia-alzheimers-en.asp#diagnosis

[4] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, and H. Chertkow, "Mild cognitive impairment," *Lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.

[5] National Institute on Aging (NIH). *What Is Mild Cognitive Impairment?* Accessed: Jul. 20, 2020. [Online]. Available: https://www.nia.nih.gov/health/what-mild-cognitive-impairment

[6] J. B. Pereira, M. Mijalkov, E. Kakaei, P. Mecocci, B. Vellas, M. Tsolaki, I. Kłoszewska, H. Soininen, C. Spenger, S. Lovestone, A. Simmons, L.-O. Wahlund, G. Volpe, and E. Westman, "Disrupted network topology in patients with stable and progressive mild cognitive impairment and Alzheimer's disease," *Cerebral Cortex*, vol. 26, no. 8, pp. 3476–3493, Aug. 2016.

[7] K. M. Langa and D. A. Levine, "The diagnosis and management of mild cognitive impairment: A clinical review," *Jama*, vol. 312, no. 23, pp. 2551–2561, 2014.

[8] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, Jr., C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, and R. C. Mohs, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 263–269, 2011.

[9] R. Sivera, H. Delingette, M. Lorenzi, X. Pennec, N. Ayache, and Alzheimer's Disease Neuroimaging Initiative, "A model of brain morphological changes related to aging and Alzheimer's disease from cross-sectional assessments," *NeuroImage*, vol. 198, pp. 255–270, Sep. 2019.

[10] L. L. Beason-Held, J. O. Goh, Y. An, M. A. Kraut, R. J. O'Brien, L. Ferrucci, and S. M. Resnick, "Changes in brain function occur years before the onset of cognitive impairment," *J. Neurosci.*, vol. 33, no. 46, pp. 18008–18014, Nov. 2013.

[11] A. D. Smith, "Imaging the progression of Alzheimer pathology through the brain," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 7, pp. 4135–4137, Apr. 2002.

[12] M. Querol-Vilaseca, M. Colom-Cadena, J. Pegueroles, R. Nuñez-Llaves, J. Luque-Cabecerans, L. Muñoz-Llahuna, J. Andilla, O. Belbin, T. L. Spires-Jones, and E. Gelpi, "Nanoscale structure of amyloid-β plaques in Alzheimer's disease," *Sci. Rep.*, vol. 9, no. 1, pp. 5181–5191, 2019.

[13] L. I. Binder, A. L. Guillozet-Bongaarts, F. Garcia-Sierra, A. W. Berry, and Alzheimer's Disease, "Tau, tangles, and Alzheimer's disease," *Biochimica et Biophysica Acta (BBA)-Molecular Basis Disease*, vol. 17, no. 2, pp. 216–223, 2005.

[14] H. Yang, H. Xu, Q. Li, Y. Jin, W. Jiang, J. Wang, Y. Wu, W. Li, C. Yang, and X. Li, "Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls," *Gen. Psychiatry*, vol. 32, no. 2, pp. 32–41, 2019.

[15] S. W. Moon, B. Lee, and Y. C. Choi, "Changes in the hippocampal volume and shape in early-onset mild cognitive impairment," *Psychiatry Invest.*, vol. 15, no. 5, pp. 531–537, May 2018.

[16] K. Juottonen, M. Lehtovirta, S. Helisalmi, P. J. Riekkinen, Sr., and H. Soininen, "Major decrease in the volume of the entorhinal cortex in patients with Alzheimer's disease carrying the apolipoprotein e ε4 allele," *J. Neurol., Neurosurgery Psychiatry*, vol. 65, no. 3, pp. 322–327, Sep. 1998.

[17] J. Barnes, J. L. Whitwell, C. Frost, K. A. Josephs, M. Rossor, and N. C. Fox, "Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration," *Arch. Neurol.*, vol. 63, no. 10, pp. 1434–1439, 2006.

[18] M. Symms, H. R. Jäger, K. Schmierer, and T. A. Yousry, "A review of structural magnetic resonance neuroimaging," *J. Neurol., Neurosurgery Psychiatry*, vol. 75, no. 9, pp. 1235–1244, Sep. 2004.

[19] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nature Rev. Neurol.*, vol. 6, no. 2, pp. 67–77, Feb. 2010.

[20] C. Ledig, A. Schuh, R. Guerrero, R. A. Heckemann, and D. Rueckert, "Structural brain imaging in Alzheimer's disease and mild cognitive impairment: Biomarker analysis and shared morphometry database," *Sci. Rep.*, vol. 8, no. 1, pp. 11258–11284, Dec. 2018.

[21] K. Biju, S. Alfa, K. Lal, A. Antony, and M. K. Akhil, "Alzheimer's detection based on segmentation of MRI image," *Procedia Comput. Sci.*, vol. 115, no. 6, pp. 474–481, 2017.

[22] A. B. Rabeh, F. Benzarti, and H. Amiri, "Segmentation of brain MRI for detecting Alzheimer's disease," *Current Med. Imag. Rev.*, vol. 14, no. 2, pp. 263–270, Feb. 2018.

[23] J. A. Kaye, "Diagnostic challenges in dementia," *Neurology*, vol. 51, no. 1, pp. S45–S52, Jul. 1998.

[24] P. Coupé, J. V. Manjón, E. Lanuza, and G. Catheline, "Lifespan changes of the human brain in Alzheimer's disease," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Dec. 2019.

[25] R. Peters, "Ageing and the brain," *Postgraduate Med. J.*, vol. 82, no. 964, pp. 84–88, 2006.

[26] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, p. 2, Dec. 2018.

[27] ADNI. *Alzheimer's Disease Neuroimaging Initiative: ADNI*. Accessed: Jul. 13, 2020. [Online]. Available: http://adni.loni.usc.edu/data-samples/access-data

[28] OASIS Brains. *Open Access Series of Imaging Studies*. Accessed: Jul. 13, 2020. [Online]. Available: https://www.oasis-brains.org

[29] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Boston, MA, USA: Cengage Learning, 2014.

[30] S. Robila, "An investigation of spectral metrics in hyperspectral image preprocessing for classification," in *Proc. Geospatial Goes Global, Your Neighborhood Whole Planet. ASPRS Annu. Conf.*, Baltimore, MD, USA, 2005, pp. 7–11.

[31] R. A. Hazarika, K. Kharkongor, S. Sanyal, and A. K. Maji, "A comparative study on different skull stripping techniques from brain magnetic resonance imaging," in *Proc. Int. Conf. Innov. Comput. Commun.* Singapore: Springer, 2020, pp. 279–288.

[32] M. Goto, O. Abe, S. Aoki, N. Hayashi, T. Miyati, H. Takao, T. Iwatsubo, F. Yamashita, H. Matsuda, H. Mori, A. Kunimatsu, K. Ino, K. Yano, and K. Ohtomo, "Diffeomorphic anatomical registration through exponentiated lie algebra provides reduced effect of scanner for cortex volumetry with atlas-based method in healthy subjects," *Neuroradiology*, vol. 55, no. 7, pp. 869–875, Jul. 2013.

[33] P. K. Mandal, R. Mahajan, and I. D. Dinov, "Structural brain atlases: Design, rationale, and applications in normal and pathological cohorts," *J. Alzheimer's Disease*, vol. 31, no. 3, pp. S169–S188, Sep. 2012.

[34] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and A. E. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Amsterdam, The Netherlands: Elsevier, 2011.

[35] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[36] J. Ashburner and K. J. Friston, "Why voxel-based morphometry should be used," *NeuroImage*, vol. 14, no. 6, pp. 1238–1243, Dec. 2001.
[37] S. Smith, M. Woolrich, T. Behrens, C. F. Beckmann, D. Flitney, M. Jenkinson, P. Bannister, S. Clare, M. De Luca, P. Hansen, and H. Johansen-Berg, "Fmrib software library," Big Healthcare Challenges Chronic Disease, Oxford Centre Funct. Magn. Reson. Imag. Brain Softw. Library, Oxford, U.K., Tech. Rep., 2014.
[38] A. W. Toga, A. W. Toga, J. C. Mazziotta, and J. C. Mazziotta, *Brain Mapping: The Methods*, vol. 1. New York, NY, USA: Academic, 2002.
[39] F. Farokhian, I. Beheshti, D. Sone, and H. Matsuda, "Comparing CAT12 and VBM8 for detecting brain morphological abnormalities in temporal lobe epilepsy," *Frontiers Neurol.*, vol. 8, p. 428, Aug. 2017.
[40] B. Sheehan, "Assessment scales in dementia," *Therapeutic Adv. Neurol. Disorders*, vol. 5, no. 6, pp. 349–358, Nov. 2012.
[41] S. E. O'Bryant, M. M. Mielke, R. A. Rissman, S. Lista, H. Vanderstichele, H. Zetterberg, P. Lewczuk, H. Posner, J. Hall, and L. Johnson, "Blood-based biomarkers in Alzheimer disease: Current state of the science and a novel collaborative paradigm for advancing from discovery to clinic," *Alzheimer's Dementia*, vol. 13, no. 1, pp. 45–58, Jan. 2017.
[42] C. Humpel, "Identifying and validating biomarkers for Alzheimer's disease," *Trends Biotechnol.*, vol. 29, no. 1, pp. 26–32, Jan. 2011.
[43] C. R. Jack, D. A. Bennett, K. Blennow, M. C. Carrillo, H. H. Feldman, G. B. Frisoni, H. Hampel, W. J. Jagust, K. A. Johnson, D. S. Knopman, R. C. Petersen, P. Scheltens, R. A. Sperling, and B. Dubois, "A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers," *Neurology*, vol. 87, no. 5, pp. 539–547, Aug. 2016.
[44] K. A. Q. Cousins, D. J. Irwin, D. A. Wolk, E. B. Lee, L. M. J. Shaw, J. Q. Trojanowski, F. Da Re, G. S. Gibbons, M. Grossman, and J. S. Phillips, "ATN status in amnestic and non-amnestic Alzheimer's disease and frontotemporal lobar degeneration," *Brain*, vol. 143, no. 7, pp. 2295–2311, Jul. 2020.
[45] P. Janicak and M. E. Dokucu, "Transcranial magnetic stimulation for the treatment of major depression," *Neuropsychiatric Disease Treatment*, vol. 11, p. 1549, Jun. 2015.
[46] M. Hallett, R. Di Iorio, P. M. Rossini, J. E. Park, R. Chen, P. Celnik, A. P. Strafella, H. Matsumoto, and Y. Ugawa, "Contribution of transcranial magnetic stimulation to assessment of brain connectivity and networks," *Clin. Neurophysiol.*, vol. 128, no. 11, pp. 2125–2139, Nov. 2017.
[47] R. Cassani, M. Estarellas, R. San-Martin, F. J. Fraga, and T. H. Falk, "Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment," *Disease Markers*, vol. 2018, pp. 1–26, Oct. 2018.
[48] R. S. Turner, T. Stubbs, D. A. Davies, and B. C. Albensi, "Potential new approaches for diagnosis of Alzheimer's disease and related dementias," *Frontiers Neurol.*, vol. 11, p. 496, Jun. 2020.
[49] Z. A. Dastgheib, B. Lithgow, and Z. Moussavi, "Diagnosis of Parkinson's disease using electrovestibulography," *Med. Biol. Eng. Comput.*, vol. 50, no. 5, pp. 483–491, May 2012.
[50] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications*, vol. 207. Physica-Verlag, 2008.
[51] I. Guyon and A. Elisseeff, "An introduction to feature extraction," in *Feature Extraction*. Berlin, Germany: Springer, 2006, pp. 1–25.
[52] E. A. A. Maksoud, S. Barakat, and M. Elmogy, "Medical images analysis based on multilabel classification," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 209–245.
[53] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
[54] P. V. Balachandran, D. Xue, J. Theiler, J. Hogden, J. E. Gubernatis, and T. Lookman, "Importance of feature selection in machine learning and adaptive design for materials," in *Materials Discovery and Design*. Cham, Switzerland: Springer, 2018, pp. 59–79.
[55] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, vol. 26. New York, NY, USA: Springer, 2013.
[56] N. El Aboudi and L. Benhlima, "Review on wrapper feature selection approaches," in *Proc. Int. Conf. Eng. MIS (ICEMIS)*, Sep. 2016, pp. 1–5.
[57] K. Yan and D. Zhang, "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," *Sens. Actuators B, Chem.*, vol. 212, pp. 353–363, Jun. 2015.
[58] N. Sánchez-Marono, A. Alonso-Betanzos, and M. Tombilla-Sanromán, "Filter methods for feature selection—A comparative study," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Berlin, Germany: Springer, 2007, pp. 178–187.
[59] M. Shinozuka and B. Mansouri, "Synthetic aperture radar and remote sensing technologies for structural health monitoring of civil infrastructure systems," in *Structural Health Monitoring of Civil Infrastructure Systems*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 113–151.
[60] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. Amsterdam, The Netherlands: Elsevier, 2006.
[61] K. L. Kvamme, E. G. Ernenwein, and J. G. Menzer, "Putting it all together: Geophysical data integration," in *Innovation in Near-Surface Geophysics*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 287–339.
[62] J. Lever, M. Krzywinski, and N. Altman, "Logistic regression," Nature Methods, London, U.K., Tech. Rep. 13, 2016, pp. 541–542.
[63] S. McCann and D. G. Lowe, "Local naive Bayes nearest neighbor for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3650–3656.
[64] E. M. Dogo, O. J. Afolabi, N. I. Nwulu, B. Twala, and C. O. Aigbavboa, "A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks," in *Proc. Int. Conf. Comput. Techn., Electron. Mech. Syst. (CTEMS)*, Dec. 2018, pp. 92–99.
[65] J. Kim[1], B. Kim, and S. Savarese, "Comparing image classification methods: K-nearest-neighbor and support-vector-machines," in *Proc. 6th WSEAS Int. Conf. Comput. Eng. Appl., Amer. Conf. Appl. Math.*, vol. 1001, 2012, pp. 2122–48109.
[66] B. Shepherd, "An appraisal of a decision tree approach to image classification," in *Proc. IJCAI*, 1983, pp. 473–475.
[67] H. Liu, M. Cocea, and W. Ding, "Decision tree learning based feature evaluation and selection for image classification," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 2, Jul. 2017, pp. 569–574.
[68] N. Abdullah, U. K. Ngah, and S. A. Aziz, "Image classification of brain MRI using support vector machine," in *Proc. IEEE Int. Conf. Imag. Syst. Techn.*, May 2011, pp. 242–247.
[69] R. Bala, "Survey on texture feature extraction methods," *Int. J. Eng. Sci. Comput.*, vol. 7, no. 4, pp. 10375–10377, 2017.
[70] S. Xiaoming, Z. Ning, W. Haibin, Y. Xiaoyang, W. Xue, and Y. Shuang, "Medical image retrieval approach by texture features fusion based on Hausdorff distance," *Math. Problems Eng.*, vol. 2018, pp. 1–12, Aug. 2018.
[71] J. Zhang, C. Yu, G. Jiang, W. Liu, and L. Tong, "3D texture analysis on MRI images of Alzheimer's disease," *Brain Imag. Behav.*, vol. 6, no. 1, pp. 61–69, Mar. 2012.
[72] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[73] L. S. Davis, "A survey of edge detection techniques," *Comput. Graph. Image Process.*, vol. 4, no. 3, pp. 248–270, Sep. 1975.
[74] D.-S. Huang, D. C. Wunsch, D. S. Levine, and K.-H. Jo, *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues: Fourth International Conference on Intelligent Computing, ICIC 2008 Shanghai, China, September 15-18, 2008 Proceedings*, vol. 5226. Shanghai, China: Springer, 2008.
[75] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, I. A. Illán, and C. G. Puntonet, "Texture features based detection of Parkinson's disease on DaTSCAN images," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.* Berlin, Germany: Springer, 2013, pp. 266–277.
[76] F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, M. Moreno-Caballero, and M. Gómez-Río, "Parametrization of textural patterns in[123]I-ioflupane imaging for the automatic detection of parkinsonism," *Med. Phys.*, vol. 41, no. 1, Dec. 2013, Art. no. 012502.
[77] K. Vaithinathan, L. Parthiban, and Alzheimer's Disease Neuroimaging Initiative, "A novel texture extraction technique with T1 weighted MRI for the classification of Alzheimer's disease," *J. Neurosci. Methods*, vol. 318, pp. 84–99, Apr. 2019.
[78] T. A. Pham, "Optimization of texture feature extraction algorithm," Delft Univ. Technol., Delft, The Netherlands, Tech. Rep., 2010.
[79] D. Gadkari, "Image quality analysis using GLCM," Univ. Central Florida, Orlando, FL, USA, Tech. Rep. 187, 2004.
[80] Z. Xiao, Y. Ding, T. Lan, C. Zhang, C. Luo, and Z. Qin, "Brain MR image classification for Alzheimer's disease diagnosis based on multifeature fusion," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–13, May 2017.
[81] T. Altaf, S. Anwar, N. Gul, N. Majeed, and M. Majid, "Multi-class Alzheimer disease classification using hybrid features," in *Proc. Future Technol. Conf. (FTC)*, 2017, pp. 264–267.
[82] G. W. Jiji, G. E. Suji, and M. Rangini, "An intelligent technique for detecting Alzheimer's disease based on brain structural changes and hippocampal shape," *Comput. Methods Biomech. Biomed. Eng., Imag. Visualizat.*, vol. 2, no. 2, pp. 121–128, Apr. 2014.
[83] D. Chyzhyk and A. Savio, "Feature extraction from structural MRI images based on VBM: Data from OASIS database," Univ. Basque Country, Internal Res. Publication, Basque, Spain, Tech. Rep., 2010.
[84] J. Ashburner and K. J. Friston, "Voxel-based morphometry—The methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.

[85] G. F. Busatto, G. E. Garrido, O. P. Almeida, C. C. Castro, C. H. Camargo, C. G. Cid, C. A. Buchpiguel, S. Furuie, and A. M. Bottino, "A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease," *Neurobiol. Aging*, vol. 24, no. 2, pp. 221–231, 2003.

[86] G. B. Frisoni, C. Testa, A. Zorzan, F. Sabattoli, A. Beltramello, H. Soininen, and A. Laakso, "Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry," *J. Neurol., Neurosurgery Psychiatry*, vol. 73, no. 6, pp. 657–664, Dec. 2002.

[87] G. Fung and J. Stoeckel, "SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information," *Knowl. Inf. Syst.*, vol. 11, no. 2, pp. 243–258, Feb. 2007.

[88] I. Álvarez, M. López, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, C. G. Puntonet, and F. Segovia, "Automatic classification system for the diagnosis of Alzheimer disease using component-based SVM aggregations," in *Proc. Int. Conf. Neural Inf. Process.* Berlin, Germany: Springer, 2008, pp. 402–409.

[89] J. Ramírez, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, and C. G. Puntonet, "Early detection of the Alzheimer disease combining feature selection and kernel machines," in *Proc. Int. Conf. Neural Inf. Process.* Berlin, Germany: Springer, 2008, pp. 410–417.

[90] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. Álvarez, F. Segovia, and C. G. Puntonet, "Computer aided diagnosis of Alzheimer disease using support vector machines and classification trees," in *Proc. Int. Conf. Neural Inf. Process.* Berlin, Germany: Springer, 2008, pp. 418–425.

[91] I. Beheshti, H. Demirel, H. Matsuda, and Alzheimer's Disease Neuroimaging Initiative, "Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Comput. Biol. Med.*, vol. 83, pp. 109–119, Apr. 2017.

[92] I. Beheshti, H. Demirel, F. Farokhian, C. Yang, H. Matsuda, and Alzheimer's Disease Neuroimaging Initiative, "Structural MRI-based detection of Alzheimer's disease using feature ranking and classification error," *Comput. Methods Programs Biomed.*, vol. 137, pp. 177–193, Dec. 2016.

[93] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, G. Catheline, and Alzheimer's Disease Neuroimaging Initiative, "Classification of Alzheimer's disease subjects from MRI using hippocampal visual features," *Multimedia Tools Appl.*, vol. 74, no. 4, pp. 1249–1266, Feb. 2015.

[94] M. Liu, D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "View-centralized multi-atlas classification for Alzheimer's disease diagnosis," *Hum. Brain Mapping*, vol. 36, no. 5, pp. 1847–1865, May 2015.

[95] C. Zu, B. Jie, M. Liu, S. Chen, D. Shen, D. Zhang, and Alzheimer's Disease Neuroimaging Initiative, "Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment," *Brain Imag. Behav.*, vol. 10, no. 4, pp. 1148–1159, Dec. 2016.

[96] L. Mesrob, M. Sarazin, V. Hahn-Barma, L. C. de Souza, B. Dubois, P. Gallinari, and S. Kinkingnéhun, "DTI and structural MRI classification in Alzheimer's disease," *Adv. Mol. Imag.*, vol. 2, no. 2, p. 12, 2012.

[97] T. Ye, C. Zu, B. Jie, D. Shen, D. Zhang, and Alzheimer's Disease Neuroimaging Initiative, "Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease," *Brain Imag. Behav.*, vol. 10, no. 3, pp. 739–749, 2016.

[98] I. Beheshti, H. Demirel, and Alzheimer's Disease Neuroimaging Initiative, "Feature-ranking-based Alzheimer's disease classification from structural MRI," *Magn. Reson. Imag.*, vol. 34, no. 3, pp. 252–263, Apr. 2016.

[99] F. Liu, C.-Y. Wee, H. Chen, and D. Shen, "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification," *NeuroImage*, vol. 84, pp. 466–475, Jan. 2014.

[100] M. Liu, D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis," *Hum. Brain Mapping*, vol. 35, no. 4, pp. 1305–1319, Apr. 2014.

[101] I. Beheshti, H. Demirel, and Alzheimer's Disease Neuroimaging Initiative, "Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease," *Comput. Biol. Med.*, vol. 64, pp. 208–216, Sep. 2015.

[102] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and A.-R. Kwon, "Alzheimer's disease diagnosis based on cortical and subcortical features," *J. Healthcare Eng.*, vol. 2019, Mar. 2019, Art. no. 2492719.

[103] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack, "Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies," *NeuroImage*, vol. 39, no. 3, pp. 1186–1197, Feb. 2008.

[104] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, G. R. Kwon, and Alzheimer's Disease Neuroimaging Initiative, "Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images," *PLoS ONE*, vol. 14, no. 10, Oct. 2019, Art. no. e0222446.

[105] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, C. G. Puntonet, D. Salas-Gonzalez, and Alzheimer's Disease Neuroimaging Initiative, "Computer aided diagnosis tool for Alzheimer's disease based on Mann–Whitney–Wilcoxon U-test," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9676–9685, 2012.

[106] A. Ortiz-Garcia, D. Fajardo, J. M. Górriz, J. Ramírez, and F. J. Martínez-Murcia, "Multimodal image data fusion for Alzheimer's disease diagnosis by sparse representation," in *Proc. KES*, 2014, pp. 1–18.

[107] A. Chincarini, P. Bosco, P. Calvini, G. Gemme, M. Esposito, C. Olivieri, L. Rei, S. Squarcia, G. Rodriguez, R. Bellotti, P. Cerello, I. De Mitri, A. Retico, and F. Nobili, "Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease," *NeuroImage*, vol. 58, no. 2, pp. 469–480, Sep. 2011.

[108] Q. Zhou, M. Goryawala, M. Cabrerizo, J. Wang, W. Barker, D. A. Loewenstein, R. Duara, and M. Adjouadi, "An optimal decisional space for the classification of Alzheimer's disease and mild cognitive impairment," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 8, pp. 2245–2253, Aug. 2014.

[109] T. Tong, Q. Gao, R. Guerrero, C. Ledig, L. Chen, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative, "A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 155–165, Jan. 2017.

[110] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1473–1482, Jul. 2016.

[111] D. Schmitter, A. Roche, B. Maréchal, D. Ribes, A. Abdulkadir, M. Bach-Cuadra, A. Daducci, C. Granziera, S. Klöppel, P. Maeder, R. Meuli, and G. Krueger, "An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease," *NeuroImage, Clin.*, vol. 7, pp. 7–17, Jan. 2015.

[112] P. Moulin, "Multiscale image decompositions and wavelets," in *The Essential Guide to Image Processing*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 123–142.

[113] R. Merry, "Wavelet theory and applications: A literature study," DCT Rapporten, Eindhoven Univ. Technol., Eindhoven, The Netherlands, Tech. Rep., 2005.

[114] D. S. R. Grace and M. I. Sheela, "A study on asphyxiating the drawbacks of wavelet transform by using curvelet transform," *Int. J. Comput. Sci. Mobile Comput*, vol. 4, no. 9, pp. 318–323, 2015.

[115] P. Padilla, J. Gorriz, J. Ramirez, R. Chaves, F. Segovia, I. Alvarez, D. Salas-González, M. López, and C. Puntonet, "Alzheimer's disease detection in functional images using 2D Gabor wavelet analysis," *Electron. Lett.*, vol. 46, no. 8, pp. 556–558, 2010.

[116] V. Mani, S. Arivazhagan, and J. J. Braino, "Multimodal image fusion using multiresolution techniques," *Elixir Adv. Engg. Info. A*, vol. 55, pp. 13160–13163, Feb. 2013.

[117] U. R. Acharya, S. L. Fernandes, J. E. WeiKoh, E. J. Ciaccio, M. K. M. Fabell, U. J. Tanik, V. Rajinikanth, and C. H. Yeong, "Automated detection of Alzheimer's disease using brain MRI images—A study with various feature extraction techniques," *J. Med. Syst.*, vol. 43, no. 9, p. 302, Sep. 2019.

[118] S.-H. Wang, Y. Zhang, Y.-J. Li, W.-J. Jia, F.-Y. Liu, M.-M. Yang, and Y.-D. Zhang, "Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10393–10417, May 2018.

[119] S. Alam, G.-R. Kwon, J.-I. Kim, and C.-S. Park, "Twin SVM-based classification of Alzheimer's disease using complex dual-tree wavelet principal coefficients and LDA," *J. Healthcare Eng.*, vol. 2017, pp. 1–12, Aug. 2017.

[120] C. Geetha and D. Pugazhenthi, "Classification of Alzheimer's disease subjects from MRI using fuzzy neural network with feature extraction using discrete wavelet transform," *Biomed. Res.*, vol. 29, pp. s14–s21, Jan. 2018.

[121] D. Jha, J.-I. Kim, and G.-R. Kwon, "Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feed-forward neural network," *J. Healthcare Eng.*, vol. 2017, pp. 1–13, Jun. 2017.
[122] H. Cheng, W. Deng, C. Fu, Y. Wang, and Z. Qin, "Graph-based semi-supervised feature selection with application to automatic spam image identification," in *Proc. Int. Workshop Comput. Sci. Environ. Eng. EcoInformat.* Berlin, Germany: Springer, 2011, pp. 259–264.
[123] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain MRI," *Med. Image Anal.*, vol. 18, no. 5, pp. 808–818, Jul. 2014.
[124] T. Tong, K. Gray, Q. Gao, L. Chen, and D. Rueckert, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognit.*, vol. 63, pp. 171–181, Mar. 2017.
[125] I. Beheshti, N. Maikusa, M. Daneshmand, H. Matsuda, H. Demirel, and G. Anbarjafari, "Classification of Alzheimer's disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks," *J. Alzheimer's Disease*, vol. 60, no. 1, pp. 295–304, Aug. 2017.
[126] A. Pulido, A. Rueda, and E. Romero, "Classification of Alzheimer's disease using regional saliency maps from brain MR volumes," in *Proc. Med. Imag., Comput.-Aided Diagnosis, Int. Soc. Opt. Photon.*, vol. 8670, 2013, Art. no. 86700R.
[127] M. Plocharski, L. R. Østergaard, and Alzheimer's Disease Neuroimaging Initiative, "Extraction of sulcal medial surface and classification of Alzheimer's disease using sulcal features," *Comput. Methods Programs Biomed.*, vol. 133, pp. 35–44, Sep. 2016.
[128] A. Khazaee, A. Ebrahimzadeh, A. Babajani-Feremi, and Alzheimer's Disease Neuroimaging Initiative, "Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI," *Behavioural Brain Res.*, vol. 322, pp. 339–350, Mar. 2017.
[129] J. Liu, J. Wang, B. Hu, F.-X. Wu, and Y. Pan, "Alzheimer's disease classification based on individual hierarchical networks constructed with 3-D texture features," *IEEE Trans. Nanobiosci.*, vol. 16, no. 6, pp. 428–437, Sep. 2017.
[130] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimer's disease using whole brain hierarchical network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 2, pp. 624–632, Mar. 2018.
[131] J. Liu, J. Wang, Z. Tang, B. Hu, F.-X. Wu, and Y. Pan, "Improving Alzheimer's disease classification by combining multiple measures," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 5, pp. 1649–1659, Sep. 2018.
[132] V. H. Gaidhane, Y. V. Hote, and V. Singh, "A new approach for estimation of eigenvalues of images," *Int. J. Comput. Appl.*, vol. 26, no. 9, pp. 1–6, Jul. 2011.
[133] A. Tsymbal, S. Puuronen, M. Pechenizkiy, M. Baumgarten, and A. W. Patterson, "Eigenvector-based feature extraction for classification," in *Proc. FLAIRS Conf.*, 2002, pp. 354–358.
[134] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinf.*, vol. 2015, pp. 1–13, Jun. 2015.
[135] M. Reuter, F.-E. Wolter, M. Shenton, and M. Niethammer, "Laplace–Beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis," *Comput.-Aided Des.*, vol. 41, no. 10, pp. 739–755, Oct. 2009.
[136] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
[137] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *J. Signal Inf. Process.*, vol. 4, no. 3B, p. 173, 2013.
[138] M. M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, and R. Chaves, "SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA," *Neurosci. Lett.*, vol. 464, no. 3, pp. 233–238, Oct. 2009.
[139] C. Li, Y. Diao, H. Ma, and Y. Li, "A statistical PCA method for face recognition," in *Proc. 2nd Int. Symp. Intell. Inf. Technol. Appl.*, vol. 3, Dec. 2008, pp. 376–380.
[140] J. Ramírez, J. M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, I. Álvarez, and P. Padilla, "Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification," *Neurosci. Lett.*, vol. 472, no. 2, pp. 99–103, Mar. 2010.
[141] A. K. Ramaniharan, S. C. Manoharan, and R. Swaminathan, "Laplace Beltrami eigen value based classification of normal and Alzheimer MR images using parametric and non-parametric classifiers," *Expert Syst. Appl.*, vol. 59, pp. 208–216, Oct. 2016.
[142] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139–150, Mar. 2015.
[143] I. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. M. López, F. Segovia, R. Chaves, M. Gómez-Rio, and C. G. Puntonet, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," *Inf. Sci.*, vol. 181, no. 4, pp. 903–916, 2011.
[144] F. Segovia, J. Górriz, J. Ramírez, D. Salas-Gonzalez, and A. I. Álvarez, "Early diagnosis of Alzheimer's disease based on partial least squares and support vector machine," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 677–683, 2013.
[145] J. M. Górriz, F. Segovia, J. Ramírez, A. Lassl, and D. Salas-Gonzalez, "GMM based SPECT image classification for the diagnosis of Alzheimer's disease," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2313–2325, Mar. 2011.
[146] R. Chaves, J. Ramírez, J. M. Górriz, C. G. Puntonet, and Alzheimer's Disease Neuroimaging Initiative, "Association rule-based feature selection method for Alzheimer's disease diagnosis," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11766–11774, Oct. 2012.
[147] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, M. López, F. Segovia, R. Chaves, P. Padilla, C. G. Puntonet, and Alzheimer's Disease Neuroimage Initiative, "Feature selection using factor analysis for Alzheimer's diagnosis using F18-FDG PET images," *Med. Phys.*, vol. 37, no. 11, pp. 6084–6095, Nov. 2010.
[148] B. Gaikwad, V. Musande, and Alzheimer's Disease Neuroimaging Initiative, "Hyperspectral image classification using harmonic analysis integrated with BFO optimized SVM," *Int. J. Comput. Sci. Netw.*, vol. 4, no. 4, pp. 2277–5420, 2015.
[149] F. Nina-Paravecino and V. Manian, "Spherical harmonics as a shape descriptor for hyperspectral image classification," in *Proc. 16th Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imag., Int. Soc. Opt. Photon.*, vol. 7695, May 2010, Art. no. 76951E.
[150] D. V. Sorokin, M. M. Mizotin, and A. S. Krylov, "Gauss-Laguerre keypoints extraction using fast Hermite projection method," in *Proc. Int. Conf. Image Anal. Recognit.* Berlin, Germany: Springer, 2011, pp. 284–293.
[151] O. Ben Ahmed, M. Mizotin, J. Benois-Pineau, M. Allard, G. Catheline, C. Ben Amar, and Alzheimer's Disease Neuroimaging Initiative, "Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex," *Comput. Med. Imag. Graph.*, vol. 44, pp. 13–25, Sep. 2015.
[152] E. Gerardin, G. Chételat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, M. Niethammer, B. Dubois, S. Lehéricy, L. Garnero, F. Eustache, and O. Colliot, "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging," *NeuroImage*, vol. 47, no. 4, pp. 1476–1486, Oct. 2009.
[153] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[154] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.
[155] J. Wu, Z. Cui, V. S. Sheng, P. Zhao, D. Su, and S. Gong, "A comparative study of SIFT and its variants," *Meas. Sci. Rev.*, vol. 13, no. 3, pp. 122–131, Jun. 2013.
[156] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei, and L. He, "Local feature descriptor for image matching: A survey," *IEEE Access*, vol. 7, pp. 6424–6434, 2019.
[157] M. R. Daliri, "Automated diagnosis of Alzheimer disease using the scale-invariant feature transforms in magnetic resonance images," *J. Med. Syst.*, vol. 36, no. 2, pp. 995–1000, Apr. 2012.
[158] S. Walczak, "Artificial neural networks," in *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*. Hershey, PA, USA: IGI Global, 2019, pp. 40–53.
[159] J. Mahanta, "Introduction to neural networks, advantages and applications," *Towards Data Sci.*, vol. 13, Jul. 2017.
[160] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *J. Clin. Epidemiol.*, vol. 49, no. 11, pp. 1225–1231, Nov. 1996.
[161] 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 Neuroinform.*, vol. 12, p. 35, Jun. 2018.

[162] B.-K. Choi, N. Madusanka, H.-K. Choi, J.-H. So, C.-H. Kim, H.-G. Park, S. Bhattacharjee, and D. Prakash, "Convolutional neural network-based mr image analysis for Alzheimer's disease classification," *Current Med. Imag.*, vol. 16, no. 1, pp. 27–35, 2020.

[163] J. M. Rondina, L. K. Ferreira, F. L. de Souza Duran, R. Kubo, C. R. Ono, C. C. Leite, J. Smid, R. Nitrini, C. A. Buchpiguel, and G. F. Busatto, "Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases," *NeuroImage, Clin.*, vol. 17, pp. 628–641, Jan. 2018.

[164] M. F. Schmidt, J. M. Storrs, K. B. Freeman, C. R. Jack, S. T. Turner, M. E. Griswold, and T. H. Mosley, Jr., "A comparison of manual tracing and FreeSurfer for estimating hippocampal volume over the adult lifespan," *Hum. Brain Mapping*, vol. 39, no. 6, pp. 2500–2513, Jun. 2018.

[165] Y. Gao and A. Tannenbaum, "Combining atlas and active contour for automatic 3D medical image segmentation," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, Mar. 2011, pp. 1401–1404.

[166] N. Lin, J. Jiang, S. Guo, and M. Xiong, "Functional principal component analysis and randomized sparse clustering algorithm for medical image analysis," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0132945.

[167] J. M. R. Tavares and J. R. Natal, *Computational Vision and Medical Image Processing IV: VIPIMAGE 2013*. Boca Raton, FL, USA: CRC Press, 2013.

[168] S. Pang, J. Jiang, Z. Lu, X. Li, W. Yang, M. Huang, Y. Zhang, Y. Feng, W. Huang, and Q. Feng, "Hippocampus segmentation based on local linear mapping," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, Apr. 2017.

[169] L.-Y. Chen and C.-J. Lu, "An improved independent component analysis algorithm based on artificial immune system," *Int. J. Mach. Learn. Comput.*, vol. 3, no. 1, p. 93, 2013.

[170] T. Nguyen, "Gaussian mixture model based spatial information concept for image segmentation," Univ. Windsor, Windsor, ON, Canada, Tech. Rep. 438, 2011.

[171] Y. Yang, Y. Song, F. Zhai, Z. Fan, Y. Meng, and J. Wang, "A high-precision localization algorithm by improved SIFT key-points," in *Proc. 2nd Int. Congr. Image Signal Process.*, 2009, pp. 1–6.

[172] P. Garg and T. Jain, "A comparative study on histogram equalization and cumulative histogram equalization," *Int. J. New Technol. Res.*, vol. 3, no. 9, 2017, Art. no. 263242.

[173] M. Tseitlin, A. Dhami, S. S. Eaton, and G. R. Eaton, "Comparison of maximum entropy and filtered back-projection methods to reconstruct rapid-scan EPR images," *J. Magn. Reson.*, vol. 184, no. 1, pp. 157–168, Jan. 2007.

[174] D. S. Jones, *Pharmaceutical Statistics*. London, U.K.: Pharmaceutical Press, 2002.

[175] B. P. Patel, N. Gupta, R. K. Karn, and Y. Rana, "Optimization of association rules mining Apriori algorithm based on ACO," *Int. J. Emerg. Technol.*, vol. 2, no. 1, pp. 87–92, 2011.

[176] M. Kavitha and S. Selvi, "Comparative study on Apriori algorithm and Fp growth algorithm with pros and cons," *Int. J. Comput. Sci. Trends Technol.*, vol. 4, pp. 2016.

[177] K. K. Pal and K. S. Sudeep, "Preprocessing for image classification by convolutional neural networks," in *Proc. IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2016, pp. 1778–1781.

[178] X. Jiang, R. Zhang, and S. Nie, "Image segmentation based on level set method," *Phys. Procedia*, vol. 33, pp. 840–845, Jan. 2012.

[179] N. A. Ahad and S. S. S. Yahaya, "Sensitivity analysis of Welch's *t*-test," in *Proc. AIP Conf.*, vol. 1605, 2014, pp. 888–893.

[180] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, pp. 577–584.

[181] B. Menze, G. Langs, A. Montillo, M. Kelm, H. Müller, and A. Z. Tu, *Medical Computer Vision. Large Data in Medical Imaging: Third International MICCAI Workshop, MCV 2013, Nagoya, Japan, September 26, 2013, Revised Selected Papers*, vol. 8331. Nagoya, Japan: Springer, 2014.

[182] J. A. Yesavage and J. O. Brooks, "On the importance of longitudinal research in Alzheimer's disease," *J. Amer. Geriatrics Soc.*, vol. 39, no. 9, pp. 942–944, Sep. 1991.

[183] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, Oct. 2007.

[184] D. M. Pirouz, "An overview of partial least squares," *SSRN Electron. J.*, 2006.

[185] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *J. Mach. Learn. Res.*, vol. 5, pp. 101–141, Jan. 2004.

[186] A. F. Zubair, S. B. Aribisala, M. Manca, and M. Mazzara, "On the parcellation of functional magnetic resonance images," in *Proc. Int. Conf. Softw. Eng. Defence Appl.* Cham, Switzerland: Springer, 2018, pp. 325–332.

[187] C. Gaser, I. Nenadic, B. R. Buchsbaum, E. A. Hazlett, and M. S. Buchsbaum, "Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI," *NeuroImage*, vol. 13, no. 6, pp. 1140–1145, Jun. 2001.

[188] P. Ammu, K. Sivakumar, and R. Rejimoan, "Biogeography-based optimization—A survey," *Int. J. Electron Comput. Sci. Eng.*, vol. 2, no. 1, pp. 154–160, 2013.

[189] F. Rajulton, "The fundamentals of longitudinal research: An overview," *Can. Stud. Population Archives*, vol. 28, no. 2, pp. 169–185, Dec. 2001.

[190] S. Blüml and A. Panigrahy, *MR Spectroscopy of Pediatric Brain Disorders*. New York, NY, USA: Springer, 2012.

[191] L. Zhou, L. Wang, C. Shen, and N. Barnes, "Hippocampal shape classification using redundancy constrained feature selection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2010, pp. 266–273.

[192] S. Niijima and S. Kuhara, "Recursive gene selection based on maximum margin criterion: A comparison with SVM-RFE," *BMC Bioinf.*, vol. 7, no. 1, p. 543, Dec. 2006.

[193] F. Lu and E. Petkova, "A comparative study of variable selection methods in the context of developing psychiatric screening instruments," *Statist. Med.*, vol. 33, no. 3, pp. 401–421, Feb. 2014.

[194] J. Xin, Y. Zhang, Y. Tang, and Y. Yang, "Brain differences between men and women: Evidence from deep learning," *Frontiers Neurosci.*, vol. 13, p. 185, Mar. 2019.

[195] B. Yang and S. Chen, "A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image," *Neurocomputing*, vol. 120, pp. 365–379, Nov. 2013.

[196] X. Wang, J. Xu, W. Shi, and J. Liu, "OGRU: An optimized gated recurrent unit neural network," in *Proc. J. Phys., Conf.*, vol. 1325. Bristol, U.K.: IOP Publishing, 2019, Art. no. 012089.

[197] S. Maldonado, J. Pérez, and C. Bravo, "Cost-based feature selection for support vector machines: An application in credit scoring," *Eur. J. Oper. Res.*, vol. 261, no. 2, pp. 656–665, Sep. 2017.

[198] V. Prasatha, H. A. A. Alfeilate, A. Hassanate, O. Lasassmehe, A. S. Tarawnehf, M. B. Alhasanatg, and H. S. E. Salmane, "Effects of distance measure choice on KNN classifier performance—A review," 2017, *arXiv:1708.04321*. [Online]. Available: https://arxiv.org/pdf/1708.04321

[199] H. Parvin, H. Alizadeh, and B. Minaei-Bidgoli, "MKNN: Modified k-nearest neighbor," in *Proc. World Congr. Eng. Comput. Sci.*, vol. 1. San Francisco, CA, USA: Citeseer, 2008, pp. 1–4.

[200] A. W. Toga, *Brain Mapping: An Encyclopedic Reference*. New York, NY, USA: Academic, 2015.

[201] R. J. Perrin, A. M. Fagan, and D. M. Holtzman, "Multimodal techniques for diagnosis and prognosis of Alzheimer's disease," *Nature*, vol. 461, no. 7266, pp. 916–922, Oct. 2009.

[202] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, and M. J. Fulham, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015.

[203] Y. Zhang, H. Zhang, E. Adeli, X. Chen, M. Liu, and D. Shen, "Multiview feature learning with multiatlas-based functional connectivity networks for MCI diagnosis," *IEEE Trans. Cybern.*, early access, Dec. 14, 2020, doi: 10.1109/TCYB.2020.3016953.

[204] W. Shao, Y. Peng, C. Zu, M. Wang, D. Zhang, and Alzheimer's Disease Neuroimaging Initiative, "Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease," *Comput. Med. Imag. Graph.*, vol. 80, Mar. 2020, Art. no. 101663.

[205] Y. Zhang, H. Zhang, X. Chen, M. Liu, X. Zhu, S.-W. Lee, and D. Shen, "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," *Pattern Recognit.*, vol. 88, pp. 421–430, Apr. 2019.

[206] T. Zhou, K.-H. Thung, M. Liu, F. Shi, C. Zhang, and D. Shen, "Multi-modal latent space inducing ensemble SVM classifier for early dementia diagnosis with neuroimaging data," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101630.

[207] W. Zheng, Z. Yao, Y. Li, Y. Zhang, B. Hu, D. Wu, and Alzheimer's Disease Neuroimaging Initiative, "Brain connectivity based prediction of Alzheimer's disease in patients with mild cognitive impairment based on multi-modal images," *Frontiers Human Neurosci.*, vol. 13, p. 399, Nov. 2019.

[208] P. Gavali and J. S. Banu, "Deep convolutional neural network for image classification on CUDA platform," in *Deep Learning and Parallel Computing Environment for Bioengineering Systems*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 99–122.

[209] F. Ramzan, M. U. G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman, and Z. Mehmood, "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks," *J. Med. Syst.*, vol. 44, no. 2, pp. 1–16, Feb. 2020.
[210] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, M. Xu, and Alzheimer's Disease Neuroimaging Initiative, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, Mar. 2020, Art. no. 116459.
[211] N. T. Duc, S. Ryu, M. N. I. Qureshi, M. Choi, K. H. Lee, and B. Lee, "3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, Jan. 2020.
[212] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, 2019.
[213] X. Bi, X. Zhao, H. Huang, D. Chen, and Y. Ma, "Functional brain network classification for Alzheimer's disease detection with deep features and extreme learning machine," *Cognit. Comput.*, vol. 12, no. 3, pp. 513–527, May 2020.

RUHUL AMIN HAZARIKA received the B.Tech. and M.Tech. degrees in information technology from North Eastern Hill University (NEHU), Shillong, India, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree in information technology. His research interests include AI and image processing.

ARNAB KUMAR MAJI received the B.E. degree in information science and engineering from Visvesvaraya Technological University (VTU), in 2003, the M.Tech. degree in information technology from Bengal Engineering and Science University, Shibpur (currently IIEST, Shibpur), in 2005, and the Ph.D. degree from Assam University Silchar (A Central University of India), in 2016. He has 16 years of professional experience. He is currently working as an Assistant Professor (Stage-3) with the Department of Information Technology, North Eastern Hill University, Shillong (A Central University of India). He has published more than 30 papers in different reputed international journals and conferences. He has published more than 20 articles as a book chapter and coauthored one book with several international publishers like Elsevier, Springer, IGI Global, and McMilan International; and four Ph.D. scholars are currently pursuing Ph.D. degrees under his active supervision. He has also guided 14 M.Tech. thesis and three Ph.D. scholars successfully. His research interests include machine learning, image processing, and natural language processing.

SAMARENDRA NATH SUR (Senior Member, IEEE) was born in Hooghly, West Bengal, India, in 1984. He received the B.Sc. degree in physics (Hons.) from the University of Burdwan, in 2007, the M.Sc. degree in electronics science from Jadavpur University, in 2007, the M.Tech. degree in digital electronics and advanced communication from Sikkim Manipal University, in 2012, and the Ph.D. degree from NIT, Durgapur. Since 2008, he has been associated with the Sikkim Manipal Institute of Technology, India, where he is currently an Assistant Professor with the Department of Electronics and Communication Engineering. His publications include more than 60 articles, including SCI/Scopus indexed journal and conference papers. His current research interests include broadband wireless communication (MIMO and spread spectrum technology), advanced digital signal processing, and radar (signal processing). He is also a Senior Member of the IEEE-IoT, the IEEE Signal Processing Society, the Institution of Engineers, India (IEI), and the International Association of Engineers (IAENG). He was a recipient of the University Medal and the Dr. S. C. Mukherjee Memorial Gold Centered Silver Medal from Jadavpur University, in 2007. He has been serving as a Guest Editor in Topical collection, *SN Applied Sciences* (Springer) since 2019. He also serves as a Reviewer for *International Journal of Electronics* (Taylor and Francis), *IET Communication*, *Ad Hoc Networks* (Elsevier), and IEEE TRANSACTIONS ON SIGNAL PROCESSING, *SN Applied Sciences*, *Transactions on Emerging Telecommunications Technologies* (Wiley), and *Telecommunication Systems* (Springer).

BABU SENA PAUL (Member, IEEE) received the B.Tech. and M.Tech. degrees in radiophysics and electronics from the University of Calcutta, West Bengal, India, in 1999 and 2003, respectively, and the Ph.D. degree from the Department of Electronics and Communication Engineering, IIT Guwahati, India. He was with Philips India Ltd., from 1999 to 2000. From 2000 to 2002, he has served as a Lecturer with the Electronics and Communication Engineering Department, SMIT, Sikkim, India. He has attended and published more than 60 research papers in international and national conferences, symposiums, and peer reviewed journals. He has successfully supervised several postgraduate students and postdoctoral research fellows. He joined the University of Johannesburg in 2010. He has served as the Head of the Department at the Department of Electrical and Electronic Engineering Technology, University of Johannesburg, from April 2015 to March 2018. He has been serving as an Associate Professor and the Director of the Institute for Intelligent Systems, University of Johannesburg. He is a Life Member of IETE. He was awarded the IETE Research Fellowship.

DEBDATTA KANDAR was born in Deulia, Kolaghat, Purba Medinipur, West Bengal, India, in 1977. He received the Ph.D. (Engg.) degree from the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, West Bengal, in 2011. He has more than 18 years of professional experience. He is currently an Associate Professor with the Department of Information Technology, North-Eastern Hill University, Shillong, India. Before joining NEHU, he has worked with the S. K. P. Engineering College, Tiruvannamalai, Tamil Nadu, and the Sikkim Manipal Institute of Technology, Sikkim, India. He has worked in a DRDO sponsored project and he has several years of industry experience too. He has successfully guided several Ph.D., M.Tech., and B.Tech., students. Four of his students already awarded the Ph.D. degree. He has national and international collaborations for carrying out research works. He has delivered talks in different workshop, seminar, and FDP. He has successfully organized several national and international seminar and conferences. He has published approximately 90 research papers in different national and international journals, conference proceedings, and book chapters. He also published edited book. He has completed his postdoctoral research from the Department of Electrical and Electronic Engineering Technology, University of Johannesburg, Johannesburg, South Africa. His research interests include wireless mobile communication, artificial intelligence, soft computing, and radar operation. He has been awarded the Young Scientist Award from Union Radio Science International (URSI GA-2005) at Vigyan Bhaban, Delhi, for his research work. President of India, Dr. A. P. J. Abdul Kalam invited him at his residence on that occasion. He also chaired several sessions in different conferences.