Brain Imaging and Behavior
https://doi.org/10.1007/s11682-018-9831-2

ORIGINAL RESEARCH

Alzheimer disease detection from structural MR images using FCM based weighted probabilistic neural network

Baskar Duraisamy[1] · Jayanthi Venkatraman Shanmugam[2] · Jayanthi Annamalai[3]

Abstract

An early intervention of Alzheimer's disease (AD) is highly essential due to the fact that this neuro degenerative disease generates major life-threatening issues, especially memory loss among patients in society. Moreover, categorizing NC (Normal Control), MCI (Mild Cognitive Impairment) and AD early in course allows the patients to experience benefits from new treatments. Therefore, it is important to construct a reliable classification technique to discriminate the patients with or without AD from the bio medical imaging modality. Hence, we developed a novel FCM based Weighted Probabilistic Neural Network (FWPNN) classification algorithm and analyzed the brain images related to structural MRI modality for better discrimination of class labels. Initially our proposed framework begins with brain image normalization stage. In this stage, ROI regions related to Hippo-Campus (HC) and Posterior Cingulate Cortex (PCC) from the brain images are extracted using Automated Anatomical Labeling (AAL) method. Subsequently, nineteen highly relevant AD related features are selected through *Multiple-criterion* feature selection method. At last, our novel FWPNN classification algorithm is imposed to remove suspicious samples from the training data with an end goal to enhance the classification performance. This newly developed classification algorithm combines both the goodness of supervised and unsupervised learning techniques. The experimental validation is carried out with the *ADNI* subset and then to the *Bordex-3 city* dataset. Our proposed classification approach achieves an accuracy of about 98.63%, 95.4%, 96.4% in terms of classification with *AD vs NC, MCI vs NC* and *AD vs MCI*. The experimental results suggest that the removal of noisy samples from the training data can enhance the decision generation process of the expert systems.

Keywords Alzheimer's disease · Structural MRI · FCM · WPNN · Hippocampus · Posterior cingulate cortex · Multiple criterion · AAL

Introduction

Alzheimer's disease (AD) is a popularly known neuro degenerative disease that generates an attempt to cause various variations in the 'cognitive' function. The patients with AD commonly suffer from memory loss and in turn causes major health concerned issues in the society. In order to receive benefits from new medical treatments an early detection of AD is important. It should be noted that the new medical treatments can suppress the neuro degenerative disease by early intervention of AD in patients. More probably, dementia is diagnosed by a neuro-imaging tool and an important question is raised whether early intervention of AD is possible alone with the value of MRI (Magnetic Resonance Imaging) data. Notably, the progression of brain atrophies can be evaluated and detected finely by a bio medical imaging modality sMRI (structural MRI); hence this modality plays a core part in the clinical assessments of brain functionalities. Considering this goodness, human brains several morphological characteristics are quantified by introducing certain techniques and methodologies for image analysis. The anatomical relevancies of brain structures were distinguished by the whole brain mor-phometric techniques proposed by (Toga et al. 2001). To compare the anatomical relevancies of brain structures the authors (Toga et al. 2001) considers

✉ Baskar Duraisamy
dbaskarphd@gmail.com; baskardr@gmail.com

[1] Hindusthan College of Engineering and Technology, Coimbatore, India

[2] Rajagiri School of Engineering and Technology, Cochin, India

[3] Sri Ramakrishna Institute of Technology, Coimbatore, India

Published online: 19 February 2018

one to one correspondence between the class labels, for example, an automatic 'volumetric' technique commonly known as VBM (Voxel based Morphometry) is utilized to examine the variations between the white and gray matters local concentration. Also, from the deformations field gradients, the local structural variations are identified by the TBM (Tensor Based Mor-phometric) technique proposed by (Studholme et al. 2006). Moreover, different complex shapes related to brains anatomical structures were analyzed by the OBM (Object Based Morphometry) technique (Magnin et al. 2009). Nonetheless, the feature based techniques tries to analyze subject variability's and this technique express their valuable thanks to their selectivity and statistical redundancy in the detection of salient image domains. The features based mor-phometric technique established by (Toews et al. 2010) completely depends on the SIFT (Scale Invariant Feature Transform) techniques promotable features. In addition to this, features involved in the brain region commonly affected by the disease are extracted by the ROI (Region of Interest) based techniques. In order to determine the most accurate measurements of brain atrophies the manual segmentation of ROI's are specifically performed by the most flexible software or by an expert. However, this manual segmentation of ROI's done by software also faces difficulties and challenges in the 'boundary' detection of brain regions; hence it generates poor results and in turn intakes more time for execution. Consequently, in brains MR images the ROI's are labeled automatically by the atlas based methods, which makes use of both the automated and standard techniques. It should be noted that, rather than this effectiveness of atlas based methodology, yet it lacks from minimum inter-subject variability's and fails to delineate information regarding brain atrophies. In recent studies, most of the researchers focused on the extraction of ROI involved in the hippocampus region and was analyzed using large number of structural analysis. The extracted ROIs are considered as an effective evidence for the detection of Alzheimer's disease. Moreover to classify the AD, the SVM (Support Vector Machine) classification approach utilized the features derived from the SPHARM (spherical harmonics) proposed by (Gutman et al. 2009; Gerardin et al. 2009) and this features includes the shape based information. In order to enhance the classification of AD and NC (Normal Controls), the SSMs (Statistical Shape Models) were utilized and this model holds the surface regions particular morphological variations (Shen et al. 2012). However, patterns with large discriminative influence are constructed potentially by the biomarkers or by the fused measurements obtained from numerous distinctive regions and this potentially effective discriminative influence can be utilized to enhance the AD related decision making process. Furthermore, for the earlier intervention of Alzheimer's disease PCC 'hypo-metabolism' is regarded as an effective tool rather than the HC (Hippocampus) Atrophy. Based on this studies (Chételat et al. 2007; Nestor et al. 2003) there raised a question, whether the earlier intervention of AD is possible to be done effectively by utilizing both the PCC and Hippocampus based atrophy rather than using HC atrophy alone for diagnosis. Moving towards the clinical diagnosis, the ROI atrophy related to brain is represented effectively through visual information provided by the structural MRI; hence the ROI atrophy related to brain emerges due to the neurodegenerative process.

An adaptive method for classification of brain soft tissues using *MRI* was developed by (Cocosco et al. 2003). In this approach, the training set was customized by utilizing the pruning criterion. By this performance, the brain *MRIs* pathology and anatomical variations can be accommodated by the classification. The prior tissues probability map generates the inaccurate sample which was reduced by minimum spanning tree. The KNN classifier makes use of these samples to classify the tissues located in *MR* images of the brain. The main limitation faced by this classifier is that it can't classify the disease affected region accurately. In order to withstand this issue, a hybrid technique was developed by (El-Dahshan et al. 2009) to classify the brain images into two categories (normal and abnormal). With the help of "Discrete Wavelet Transform" (DWT) the features were extracted from the brain *MR* images using the methods of (Frisoni et al. 2005).

Numerous pattern classification techniques were introduced by various researchers to perform diagnosis accurately in individual form. Beyond this, SVM (Support Vector Machine) is a frequently used classification approach by various researches for the discrimination of AD (Fan et al. 2008; Frisoni et al. 2005). In order to enhance the classification performance, the neuro imaging data dimensionality is ought to be gradually reduced by utilizing few specific techniques which have been illustrated in some literal works. Also, this dimensionality reduction technique should be applied prior before employing the classification algorithm in order to choose the top level discriminative features.

By the way, certain inadequacies are identified in the aforementioned studies. They are as follows a) initially the classification algorithms performance largely depends on the features which are extracted from the brain image modality and in turn, these extracted features were utilized to train the 'classifier' model. It is to be noted that, if more number of features are embedded, the classifiers may experience congestion due to the deployment of redundant information by the embedded features, therefore the computational complexity is maximized. Most of the literal works depends on the feature selection approach that relays not more than two evaluation criteria for dimensionality reduction. More probably dimensionality reduction is done by a single criterion technique yet it shows limited ability in the decision making framework. With this in mind, in this work, we employed the multiple criterion approach for the appropriate selection of an optimal set of features; (b) the above described reported works fails to focus on the suspicious samples in the training data and ignores to discuss about the removal of this suspicious samples. However, the effective removal of suspicious samples from the training data is essential to enhance the classification performance and the above discussed reported works diagnostic accuracy is gradually decreased due to poor training performance (i.e. presence of more ambiguous samples in the training data). Thus, it should be careful that prior to the construction of classifier model it is necessary to verify that the adopted classifier should possess the ability to identify and to remove the suspicious (doubtful) samples in the training data. In addition to this, the adopted classifier model must possess the ability to work well during classification of data by minimizing false negatives and false positives, respectively. Therefore, in this study, we employed a new combined classification approach referred to as FWPNN, which combines both the advantageous of supervised and unsupervised learning techniques to handle suspicious samples in the training data.

The major contribution of this paper is to enhance the classification performance by developing a most effective classification method which can identify and remove the suspicious samples in the training data. Initially, in this work, we make use of the brain images from structural *MRI* with the end goal for better discrimination of subjects namely, *NC* (Normal Control), *MCI* (Mild Cognitive Impairment) and *AD* (Alzheimer's Disease). A tremendous amount of training data employed for the discrimination of subjects may mislead the classifier by taking incorrect decisions and in turn degrades the classification performance. However, class labels when explored manually increases the time consumption and leads to highly expensive process. Thus to withstand this issue, it is essential to impose an automated technique to enhance the decision-making process by identifying and eliminating noisy samples from the training data. With this in mind, our novel FWPNN classification algorithm is developed by combining Fuzzy c-means (unsupervised learning technique) and Weighted Probabilistic Neural Network (supervised learning technique) to categorize the class labels. In order to identify suspicious samples in training data, we employed the unsupervised learning technique. Figure 1 depicts a block-diagram of the proposed classification approach. The suggested framework consists of the following steps:

- The visual information based on the Alzheimer's disease is successfully retrieved by analyzing the brain images of structural *MRI*. In the brain image 'normalization' stage, the ROI regions related to the hippocampus and posterior cingulate cortex from the brain images are extracted using Automated Anatomical Labeling *(AAL)* method.
- Next, to this normalization stage, most important texture and shape features are extracted from HC and PCC regions involved in 3 brain planes (axial, sagittal and coronal) related to each slice, which are severely affected by the disease. Approximately, for about 19 highly relevant *AD* related features are selected through *multiple-criterion* feature selection method.
- Finally, a novel classification algorithm FWPNN is proposed by combining "Fuzzy C-Means Clustering" (unsupervised learning technique) and "Weighted Probabilistic Neural Network" (supervised learning technique) to categorize NC, MCI, and AD from structural *MRI*. The discrimination of class labels at the instance of numerous training patterns turns to be tedious and time consuming process while working alone with the supervised learning technique. Therefore, the supervised learning technique combined with unsupervised learning technique can enhance the classification performance by removing the suspicious training samples in the training data.
- We employed the technique on an *ADNI* subset and then to a small portion of "French subsets" related to AD subjects of *Bordex-3 City* dataset. The empirical experimental results demonstrate that removal of suspicious (doubtful) samples in the training data can enhance the expert systems decision generation process. The proposed classification approach exhibits significant progression in classification accuracy when it is contrasted with few existing classification algorithms.

The rest of the paper is organized as follows. In the Materials and methods section, we describe the data acquisition and methodologies used in this work for better discrimination of patients with/without the AD. In the Results and discussion section, we analyzed the effectiveness of our proposed classification approach in the discrimination of AD-affected patients via comparison with the other conventional techniques and then discuss about the importance of ROI extraction from brain images for AD diagnosis. Finally, we conclude this paper and described about the further future scope of this work.

Materials and methods

Data acquisition

Initially, the *ADNI* scans were utilized to conduct the experiments and then we move on to the Bordex-3 city datasets structural *MRI* data. In the literal works, many of the authors worked on distinctive data by utilizing distinctive *MMSE* measures and with distinctive subjects. Furthermore, the authors (Cuingnet et al. 2011) conducted an experiment by comparing about 10 classification techniques submitted to

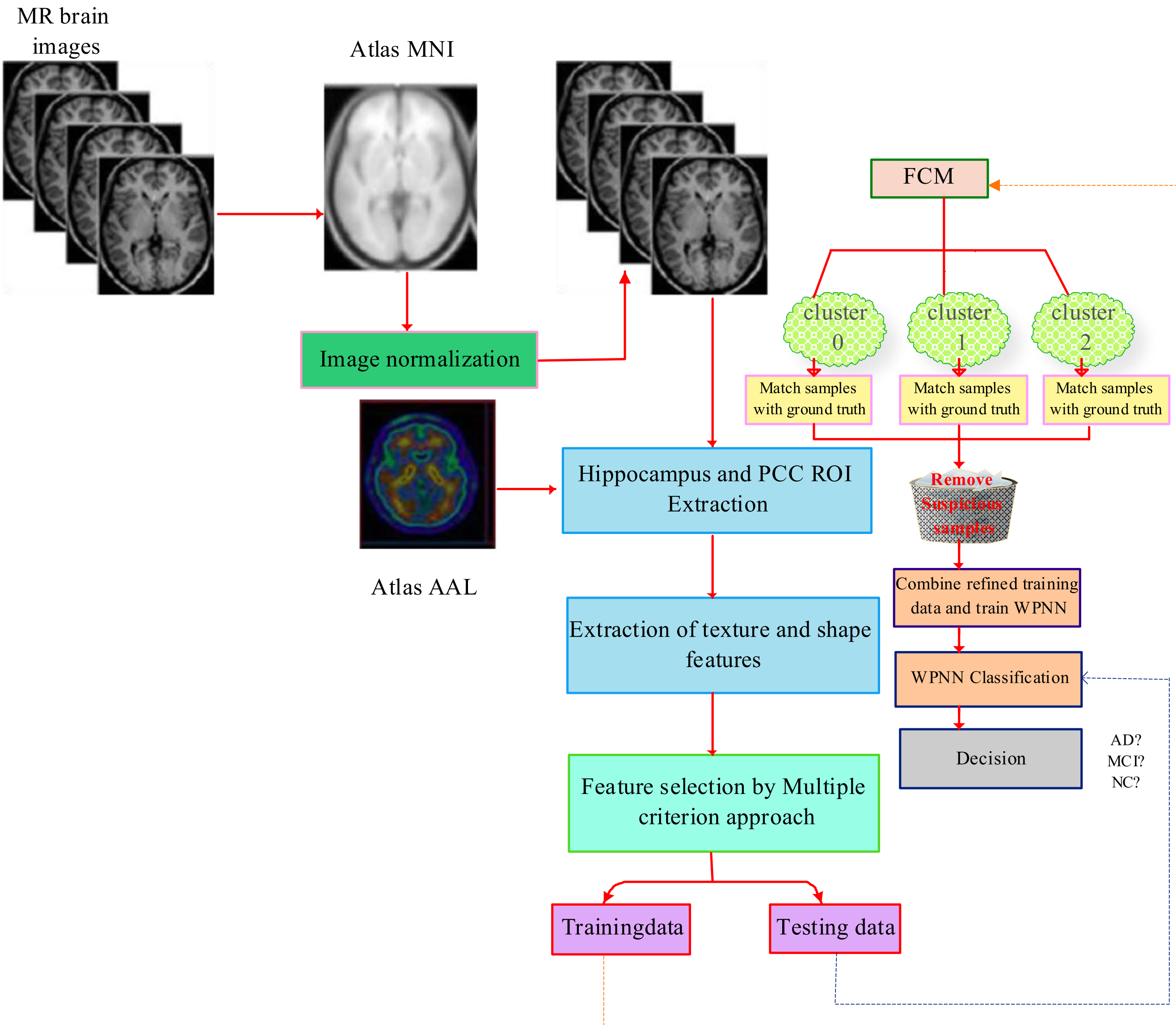


Fig. 1 Overview of the proposed classification framework

structural *MRIs* baseline obtained from the *ADNI* database. Therefore, in this work, we planned to select such kind of data from *ADNI* database. It includes about 137 AD patients, 210 MCI patients and 162 NC patients. The detailed description of the procedures of *MRI* acquisition is obtainable on the *ADNI* website.2 (http://adni.loni.ucla.edu/). The National Institute of Biomedical Imaging and Bioengineering (NIBIB) as well as National Institute on Aging (NIA) launched the *ADNI* in the year 2003. Additionally, *ADNI* was also commenced by the companies of "private pharmaceuticals" and "non-profitable firms" as 60 million, private partnership for about 5 years. The main concern behind *ADNI* is to verify that the combination of serial *MRI*, biological markers, Positron Emission Tomography (*PET*) and the assessment of clinical and neurophysiologic conditions provides the details about the brain disease such as MCI and AD. The progression behind very earlier detection of the AD by the identification of specific and sensitive markers creates interest for the clinicians and investigators to construct effective treatments. The standardized images such as 1.5T screening baseline T1 weighted are acquired by utilizing the "Volumetric protocol" 3*D MPRAGE*. After this process, our technique has been practiced to one of the "real cohort" dataset referred to as Bordex-3 city. To compute the performance of our technique towards the clinical practices we will make use of this database. It includes about 16 AD patients, 37 MCI patients, and 21 NC patients. This database shows the performance of our technique after being computed with a small amount of data as well as limitation in the availability of former knowledge about the development of the

Table 1 Demographic behaviors related to ADNI subset

Gender (male/female)	Age (range)	Number	MMSE (range)	Diagnosis
m-76/f-86	{60–90}	162	{25–30}	NC
m-127/f-83	{55–88}	210	{23–30}	MCI
m-67/f-70	{55–91}	137	{18–27}	AD

Table 2 Demographic behaviors related to Bordex-3 city dataset

Gender (male/female)	Age (range)	Number	MMSE (range)	Diagnosis
m-9/f-12	82.7 ± 4.5	21	27 ± 1	NC
m-9/f-7	77.6 ± 9.7	16	203 ± 3	AD

The measures are represented by mean ± standard deviation

disease. The investigation of MR images was functioned by utilizing the system 3 T Achieva. This system was developed by the "Philips Medical systems" in Netherland. Moreover, this system was constructed with "SENSE" head coil comprising of 8 channels. The high resolution *MRI* volumes utilized for the demonstration of anatomical structures are obtained from the "traverse plan". This plan utilized a3*D MPRAGE* T1 weighted sequence. It yields about 180 slices of 1 mm together with TR/TE8.2/3.5 ms, 256 × 256 matrix size and flipangle of 7. Approval for this study was obtained from the review board of "Institutional Human Ethics". The selected subject's demographic behaviors are summarized in Tables 1 and 2. The subjects are depicted on the basis of the examination carried out in "mini mental state" (MMSE), gender, age, and number.

Methods

We are focusing on the brain images of patients with Alzheimer's disease (AD), Mild Cognitive Impairment (MCI) and Normal Control (NC). These three categories show certain variations in the brain structure. The basic step of our proposed approach lies on normalization of brain image, which is a major step required for the comparison of the brain image. Next to this, by using a brain template, the ROI regions related to HC and PCC from the brain images are extracted using *AAL* method. Subsequently, the most important texture and shape features are extracted from HC and PCC involved in 3 brain planes (axial, sagittal and coronal). Then by utilizing *multiple-criterion* feature selection approach about 19 highly relevant AD related features are selected to reduce the dimensionality issue. Finally, we employed our novel classification algorithm FWPNN to categorize the subjects NC, MCI and AD from structural MRI modality. The major goal of our proposed classification algorithm is to enhance the classification performance by eliminating the suspicious samples in the training data. It should be noted that the discrimination of class labels at the instance of numerous training patterns turns to be tedious and time consuming process while working alone with the 'supervised learning' technique. In order to handle this issue, our proposed classification algorithm is developed by combining both the supervised and unsupervised learning technique, this, in turn, removes the suspicious training samples and improves the classification performance. Consequently, our proposed classification algorithm (FWPNN) can classify the subjects NC, MCI and AD accurately from the structural MR images of the brain. Moreover, our proposed framework executes well without interrupting the clinician activities at the time of diagnosing the disease and it operates with less time consumption. Figure 1 depicts a block-diagram of the proposed approach. In the following subsection, first, we detailed about brain image 'normalization'. Next, we detailed about the extraction of highly relevant features based on multiple criterion selection approach. Finally, we elucidate about the proposed classification algorithm which generates an attempt to remove the suspicious samples for better determination of patients with the AD.

Normalization

Initially, MRI scans are aligned based on the standard brain template to extract the visual features. Notably, for extraction of Region of Interests (ROIs) from brain image, alignment turns to be a mandatory process. According to common practices, the two types of alignment referred to as linear/nonlinear can be used. For global geometrical variations, a coarse registration is applicable with the linear transform, which may be affine or a rigid body. Example for this is rotation and magnification. Moreover, with the linear transform, precise alignment of anatomical structures is not possible due to "inter-subject" anatomical variances. Alternatively, brain structures can be aligned preciously through non-linear deformable registration. In fact, it is not possible to ensure that the brain images are aligned precisely; hence some of the solitary patterns comprised in brain structures are missed and over-alignment of images may occur. The authors (Ridha et al. 2007) presented the clear depiction for this issue. They, in turn, illustrated several limitations of the 'non-linear registration' that relates to the VBM (Voxel Based Morphometry) approach. Basically, in terms of feature based techniques, the 'deformable registration' is not so much preferable. It is not appropriate because we are in demand to conserve the particular patterns of the brain structures along with its features, yet the 'deformable registration' approach only preserves the patterns of the brain structure. Subsequently, our approach mainly focused on the extraction of *ROIs* from each slices. Figure 2 shows the three forms of brain slices, which corresponds to the sMRI scans. Hence, we analyzed each and every slice of the brain to extract *ROIs*. The slice dependent methodologies have been clearly utilized (Akgül et al. (2009). Additionally, the particular local patterns of the brain are preserved with the 'affine registration'. In this work, the entire scans that correspond to the brain template MNI 152 (Frisoni et al. 2005) are functioned with an 'affine registration'. The MNI 152 brain template was constructed by the Montreal Neurological Institute (*ICBM, NIH*). They make use of VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm/), which is available without any restraint. It diminishes the brain template and subject images 'least square' distance as well as it works with 12 degrees of freedom.

ROIs selection from Hippo-Campus and posterior cingulate cortex

In 3*D* space, the brain images are registered affinely by utilizing a digital Atlas. Depending upon the functionalities of 'Atlas', the brain images are sliced again. Hence, applying the mass of 2*D* slices over the Atlas it is possible to detect ROI (Region of Interest). It is to be notified that here in this work we utilized a small region of 3*D* brain volume. Furthermore, our medical partners suggested these regions for investigation. We, in turn, utilized this region for examination due to the fact that these regions are highly relevant for the classification of disease from particular *MR* scannings. The Hippo-Campus and PCC's Region of Interests (ROIs) are selected by utilizing a brain Atlas commonly referred to as Automated Anatomical Labeling (AAL) (Tzourio-Mazoyer et al. 2002). Figure 3 depicts the selection of ROIs from brain images. Thus to diminish the brain tissue processing, we, in turn, eliminate the skulls voxels by generating a mask. To the MNI standard space, both the labeled template and the resultant mask are being registered. This was operated with the "SPM8 software" (Institute of Neurology, London UK, UCL, Welcome trust centre for neuro-imaging). The brain subjects are classified based on the features, which is described in the subsequent sections of this work.

Extraction of visual features

The image features are computed after ROIs selection and brain alignment. It is essential to be pointed out that the features that comprises out the visualized information notifying the existence or non-existence of *AD* are preferably extracted. Moreover, T1 weighted MRI provides the high tissues contrast which facilitates us to acquire the exact analysis of structural MRI (Fig. 4). This analysis is utilized as a potential 'biomarker' for the prediction of the AD.

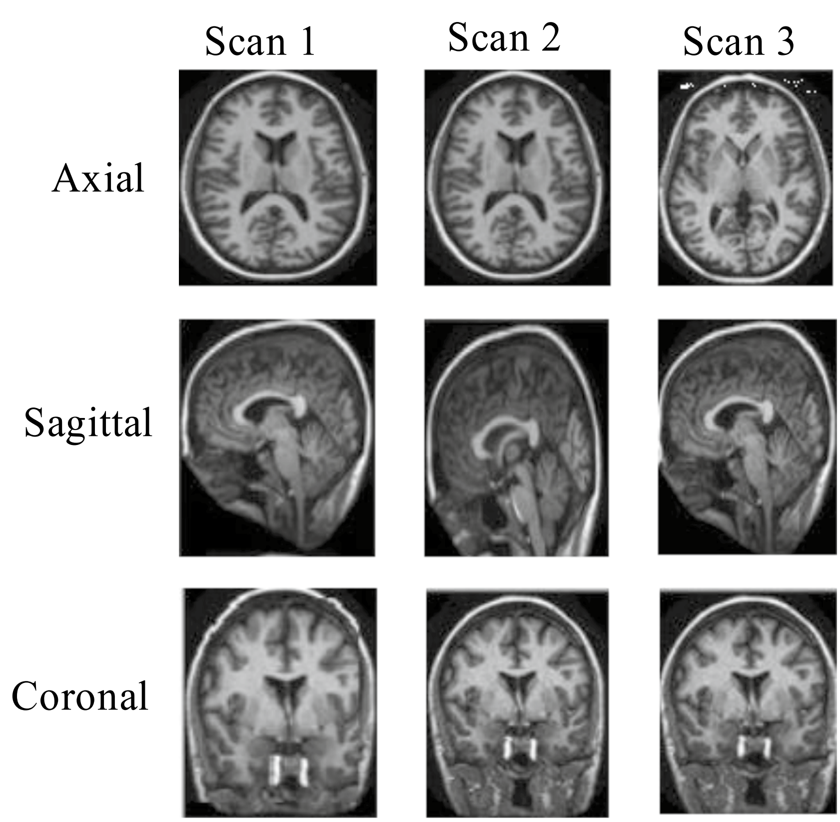


Fig. 2 Brain slices embedded for ROIs extraction

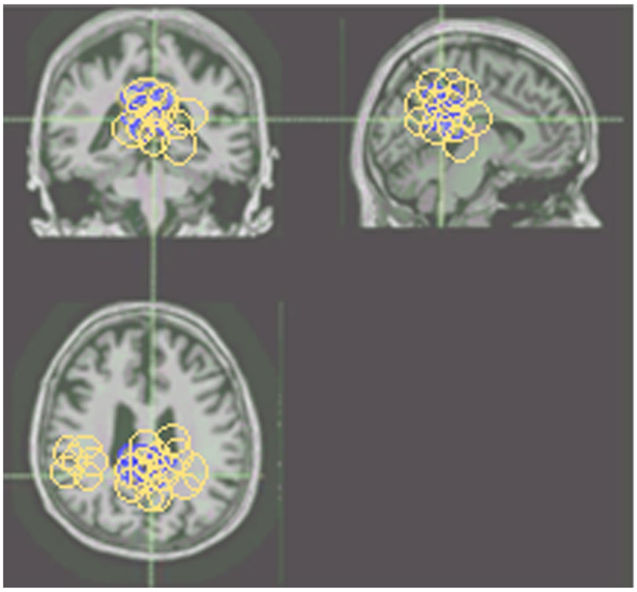


Fig. 3 ROIs detection from brain images of structural MRI

Feature extraction is commonly termed as a transformation of images into feature set. From an image, only certain essential features regarding the presence or absence of AD are extracted and these extracted features are utilized to perform classification.

One of the competing tasks behind classification is the extraction of a superior set of features. Here in this work, some of the essential texture and shape features are extracted from HC and PCC involved in three brain planes (axial, sagittal, and coronal) of each slice, which is mostly affected by the disease. In order to classify the brain diseases, the feature extraction and analysis takes a core part. Approximately, about 457 features from the brain images which utilize the brain template were extracted. More probably, 447 texture as well as 10 shape measures altogether forms 457 features. It is well depicted in Table 3 along with its respective references. Thus with the support of linear scaling the entire extracted features is normalized in the limit of [0–1].

Dimensionality reduction

Feature selection is considered as an imperative task in the recovery of dimensionality issues which may arise due to

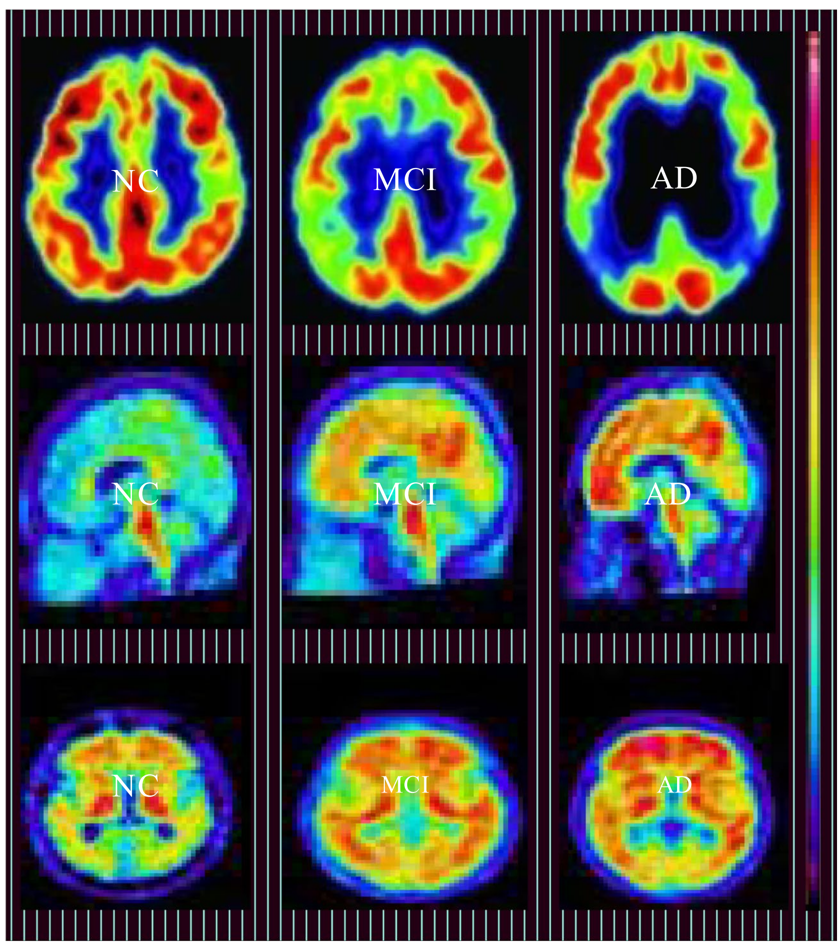


Fig. 4 High tissue contrast in the brain slices

Table 3 Summary of features utilized for classification

Category	Name	Number	Related references
Run length texture	GLNU, LRLGE, LRHGE, SRHGE, SRLGE, HGRE, LGRE, RPC, RLNU, LRE, SRE	f[1] to f [11] (i.e.11 features)	(Galloway 1975; Chu et al. 1990; Dasarathy & Holder 1991)
Spectral texture	Angular features (180) and (199) radial features	f[12] to f[191] and f[192] to f[390) (i.e. 379 features)	(Dyer & Rosenfeld 1976; Gonzalez & Woods 2010)
Fractal texture	Hurst coefficient	f[391] to f[394] (i.e. 4 features)	(Wu et al. 1992)
Laws texture	LS, ES, LE, SS, EE, LL	f[395] to f[400] (i.e. 6 features)	(Laws 1980)
Statistical feature matrix	Roughness, periodicity, contrast, coarseness	f[401] to f[404] (i.e. 4 features)	(Wu et al. 1992)
Neighbourhood Gray Tone Difference [NGTD]	Strength, complexity, busyness, contrast, coarseness	f[405] to f[409] (i.e. 5 features)	(Amadasun & King 1989; Stoitsis et al. 2006)
Gray level difference statistics	Mean, entropy, moment, angular second, contrast	f[410] to f[413] (i.e. 4 features)	(Weszka et al. 1976)
Haralick texture	The subsequent features range and mean are estimated. Correlation-1& 2 information values, Entropy difference, variance difference, Entropy, sum of variance and entropy, moment difference, sum of squares sum, correlation, contrast, angular moment	f[414] to f[439] (i.e. 26 features)	(Haralick et al. 1973; Weszka et al. 1976)
First order statistics	Entropy, uniformity measure, smoothness, kurtosis, skewness, standard deviation, variance, mean	f[440] to f[447] (i.e. 8 features)	(Srinivasan & Shobha 2008)
Moment invariants	$\mu1$ to $\mu7$	f[448] to f[454] (i.e. 7 features)	(Hu 1962)
Regional features	$\frac{Perimeter*2}{Area}$, perimeter, area	f[455] to f[457] (i.e. 3 features)	(Gonzalez & Woods 2010)

the increased number of data. The dimensionality issue may degrade the classifier performance and in turn, maximizes the utility of time required for computation. The classifier performance can be enhanced by the appropriate selection of essential features. The techniques behind the feature selection approach are widely classified into two types. They are "Wrapper" and "Filter" techniques. The Feature Ranking methodology is basically utilized by the Filter technique and is regarded as an important strategy (criterion) in the selection of significant features. The significant features were selected just by ordering them. Alternatively, the feature subsets are generated by the Wrapper technique. The Filter based techniques state of art work is delineated as follows: Relief F (RLF), Symmetrical Uncertainty (SU), Chi-Square Score (CHI2), Gain Ratio (GR) and Information Gain (IG). Moreover, the Wrapper approaches are delineated as follows: Recursive Feature Elimination (RFE), Random Forest (RF), Consistency Measure (C), and Person's Coefficient (R). The literal works of both 'Filter' and 'Wrapper' techniques are analyzed and utilized in this work. Hence, these techniques are broadly analyzed in the literal works of (Chandrashekar et al. 2014; Tang et al. 2014). Towards the classification and pattern recognition, the conventional feature selection technique faces more inadequacies. This limitation faced by the conventional feature selection technique is due to the selection of features with a single evaluation strategy (criterion). This single strategy (criterion) is derived by the (Kim et al. 2003). In this work, we introduced a novel *multiple criterion* feature selection technique to overcome the inadequacies faced by the conventional *single criterion* feature selection technique established by (Kim et al. 2003). The relevancy of features is evaluated by considering the majority of the vote obtained by the features. The majority voting acquired for the relevant features are estimated by the novel *multiple criterion* feature selection technique just by gathering ten distinctive "Filter" and "Wrapper" methodologies. The Working procedures of *multiple criterion* feature selection approach are as follows, (i) Initially, about 457 features comprised in the feature vector were extracted from the regions of *HC* and *PCC* involved in the brain images and these extracted features are provided as an input to the distinctive modules *(RFE, RF, C, R, RLF, IR, SU, CHI2, GR, IG)* of feature selection technique. (ii) Followed by this, about 19 topmost or best features are generated by each and every module of the feature selection technique. (iii) Finally, maximum votes gained for each *f*[1] to *f*[457] features are estimated and this measure is termed as 'Voting' score.

Consider an instance that, *f[3]* feature is identified in the selection 'list' of entire modules of feature selection techniques, then for *f*[3] the 'Voting' score gained is 10.

Depending upon the 'voting' score gained by the features of $f[1]$.....$f[457]$ ranking of features is carried out. In order to minimize the time required for computation only 19 better significant features have been utilized in this work for classification of class subjects related to brain disease namely NC, MCI, and AD. All other exceptional features were ignored due to minimum 'Voting' score gained by the features ($Voting < 3$). At last, features selected for HC and PCC with maximum 'voting' score are applied for classification.

Classification

In order to diagnose the patients with AD from the structural MR images accurately, we imposed a novel classification algorithm named as FWPNN. Here, we combined both the 'unsupervised' and 'supervised' learning approaches for better categorization of neuro related diseases from the structural MR images. The main function of supervised learning approach (i.e. Weighted Probabilistic Neural Network (WPNN)) is to categorize the indefinite data samples only with the definite dataset knowledge. To classify the data samples with the supervised learning approach it demands for the labels from the training data. Numerous training patterns/data employed for classification at a simultaneous time will make the supervised classifier to classify the data in an incorrect manner. Thus, classification performance is degraded with the incorrect decision of data. In order to recover this issue, the supervised learning approach is combined with unsupervised learning (Fuzzy c-means clustering (FCM)) approach to enhance the classification performance. The unsupervised learning approach does not demand for large amount of details about the decisions of the dataset as much the details required by the supervised learning approach. However, we combine the supervised learning approach to the unsupervised learning approach due to one of the advantages of supervised learning approach (i.e. severe errors can be identified just by analyzing the training set). The unsupervised learning approach does not pave the way to generate a huge number of errors by the operator. The detailed description of the proposed classification algorithm FWPNN is explained below:

Weighted probabilistic neural network

Moving towards the research field of neural networks, Weighted Probabilistic Neural Network (PNN) is considered as the most famous classification approach. The brief descriptions behind the Weighted Probabilistic Neural Network (WPNN) can be found in Kusy & Kowalski (2018). The Weighted Probabilistic Neural Network (WPNN) approach instead of depending on the heuristic methodology it completely works on the basis of statistical principles. For the gradual enhancement of system performance, certain modifications were made to the system parameters by the heuristic approaches. Subsequently, based upon the probability density functions corresponding non-parametric "Kernel-based Estimators" as well as with the "Baye's decision strategy" the WPNN is computed. More probably, the "Baye's decision strategy" represents the class of probability density functions as 'smooth' and 'continuous'. Also, it is ensured to move towards this strategy due to the fact that the continuous and smooth functions are generated by following the smoothing parameter to select the appropriate measure with respect to the trial and error technique. The WPNN based on the Baye's decision rule is represented as $d(y) = c_j$, which means the vector y corresponds to the class c_j by the condition,

$$P(c_j)P(Y|c_j) \geqslant P(c_k)P(Y|c_k), \quad \forall j = 1, 2, \ldots\ldots, N \tag{1}$$

Whereas, $P(c_j)$ denotes the probability of a vector that corresponds to the class c_j in consideration with the vector identity. On the other hand, h_j represents a priori-probability and c_j be the classes, where $j = 1, 2, 3...., N$. The conditional probability density function's is denoted by $P(y|c_j)$, in which, the function y belongs to the class c_j. In other words, the function $F_j(y)$ can also be represented as a "priori-conditional probability density function" of y in terms of each class c_j.

$g_j(Y) = P(c_j)P(Y|c_j)$, in which, $g_j(Y)$ denotes the "Baye's decision function".

$g_j(Y) > g_\ell(Y) for\ \ell \neq j$, represents the "Baye's decision rule".

It is to be noted that the above defined "Baye's decision rule" can be reformulated on the basis of PNN function as follows:

$$h_j F_j(Y) > h_\ell F_\ell(Y)\ for\ \ell \neq j \tag{2}$$

Whereas, $h_j F_j(Y) = g_j(Y)$ and the term h_j represents a probability of "priori occurrence" and the probability density function is denoted as $F_j(y)$. The class of probability density function $F_j(y)$ can be estimated by the above delineated expressions. The expression can be applied directly if a priori is observed in the function; else, it turns to be necessary for the computation of parameters. The WPNNs system model is depicted in Fig. 5. It includes about three forms of layers namely, an input layer, an output layer, and a summative layer. Hence, the Weighted Probabilistic Neural Network (WPNN) is also named as 'Three-layered Feed Forward Neural Network'. Moreover, these three layers are stable (fixed), yet the connective weights and the number of nodes may vary based upon the input/output facts.

Input Layer This input layer doesn't seek for any computational operations for the transmission of data to the neurons that correspond to the pattern unit. The input unit provides

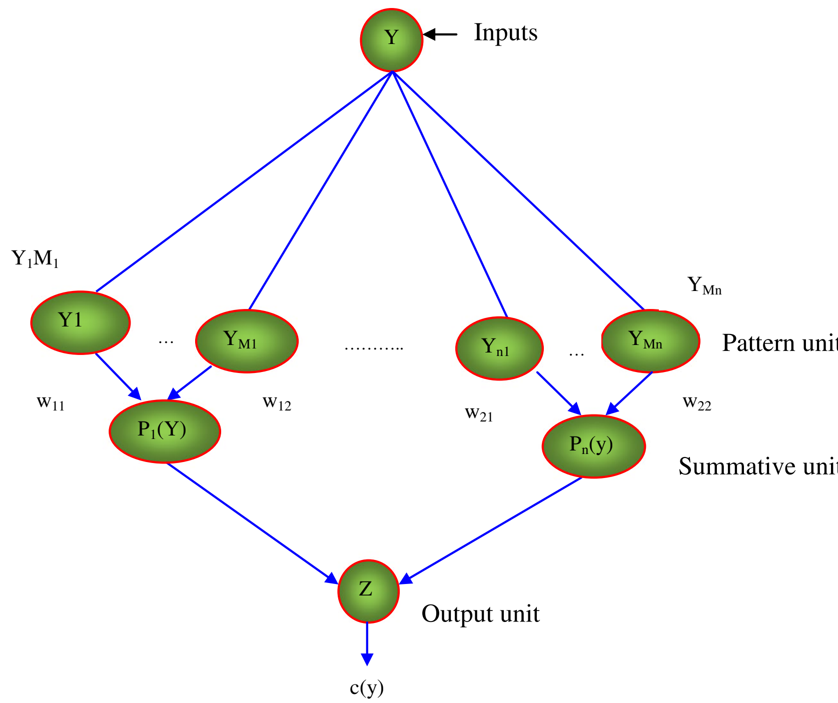


Fig. 5 Architecture of WPNN

the pattern y and the output is evaluated based upon the neuron y_{jk} embedded in the 'pattern' unit. The output from the pattern unit is computed by utilizing standard formula, which is expressed in the following equations:

$$g_j(y) = \frac{1}{(2\pi)^{P/2}\mu P}\sum_{k=1}^{n_j}\left[\exp\frac{(z_{jk}-1)}{\mu^2}\right] \Rightarrow h_j F_j(y) \tag{3}$$

Consequently, each and every sample points "bell curve" width is estimated by the smoothing parameter μ.

Summative layer From y samples, the most similar patterns are evaluated by the neurons and the similar patterns obtained are grouped into class c_j. This is done by averaging weighted outputs that corresponds to the neurons of similar class and is being computed by utilizing Eq. (4)

$$P_j(y) = \sum_{k=1}^{M_j} w_{jk}\exp\frac{(z_j-1)}{\mu^2} \tag{4}$$

The proportion between the "between-class variance" and "within-class variance" of the training pattern y_{jk} is represented by the term w_{jk}. Moreover, class c_j groups all the samples and the total sample collections are represented by the term M_j. In case, class separability of a pattern is in high range, then the ratio w_{jk} is in maximized form. On the other hand, for minimum class separability the ratio w_{jk} is in low form. Notably, the class separability of patterns is evaluated to ignore the difficulties while treating all patterns in simultaneous manner. The patterns are weighed based upon their 'class separability' and it should be noted that a high weighted pattern acquires better discrimination ability (i.e. high class separability). The patterns with maximum weight possess the ability to discriminate the class labels accurately. The weight updation is performed between the summative and output (pattern) unit. Furthermore, all the weights are fixed (constant) in Probabilistic Neural Network (PNN), but, in terms of modified PNN known as Weighted Probabilistic Neural Network (WPNN) the classification rate is highly improved with classification rate.

Output Layer In accordance with number of classes, this layer equals the number of nodes and each of them corresponds to generate one of the possible solutions (decision), whether the condition is normal or abnormal. This layer behaves as similar to "winner-take all" layer. It means the node that possesses maximum activation is highly preferred to take network decisions. Moreover, based upon the "Baye's decision rule" and with the output of all the summative neuron units, the output layer classifies the abnormal and normal conditions using the below expressed equation,

$$c(Y) = \mathrm{arg}_{\max}\left\{P_j(Y)\right\}, \quad j = 1, 2, \ldots.., M \tag{5}$$

The class that corresponds to the estimated pattern y is represented by the term $c(y)$. Moreover, the training samples in turn groups all the classes and the total collections of classes are indicated by the term M. It is to be noted that, to enhance the performance of WPNN, it turns to be crucial to select the proper measure for the smoothing parameter μ. In order to evaluate μ accurately there is no such common method, however, it can be evaluated by varying its measure minutely and by investigating its resultant classification rate. Subsequently, throughout the network, the same measure is possessed by the smoothing parameter μ and this is the only modification done to optimize the network as a 'classifier'. Furthermore, for a set of training vectors y_{jk} the finest smoothing parameter μ is identified as a part of training. This best smoothing parameter μ gradually improves the classification accuracy of other non-dependent test vectors (i.e. known set of test vectors). If there experienced the absence of test sets, then it is possible to utilize the "hold-out one testing" technique. At this condition, only one vector is taken out from the training set and this single vector is utilized to test the remaining vectors. In case, if the WPNN obtained the knowledge about the continuous variations in the decision boundary, then it approaches for a 'matched filter' at infinity. Nonetheless, the function transforms to a non-linear boundary, if the smoothing factor μ demands for a zero function. Thus, the closest neighbor classifier is represented by this non-linear boundary. More commonly, the optimized separations of class distributions can be performed by the extreme cases. It should be noted that, around the peak area a slighter variations in the smoothing factor doesn't generate a significant variations in the misclassification rate. An example for this is delineated by a learning curve and is depicted in Fig. 6.

In this work, WPNN commonly referred to as the supervised learning technique is employed to speed up the training process and to increase the convergence accuracy. In accordance with this, the Probabilistic Neural Network (PNN) (Specht DF 1990) possesses the ability to generate decision based on the Baye's strategy and obtains the probability of classification, which in turn improves the classification degree to a higher level. One of the goodness behind the conventional PNN classification approach relays on the tolerance of sudden changes experienced by employing a new set of data to the previous training vector as soon as it is available. But, it should be noted that the conventional PNN approach fails to consider their corresponding class separability and in turn it generates an attempt to manage all the patterns by assigning equal weights without the knowledge of class separability. The WPNN (Specht DF 1990) when contrasted with the conventional PNNs it embeds the weighing factors between the summative and pattern layer of the Probabilistic Neural Networks (PNNs).

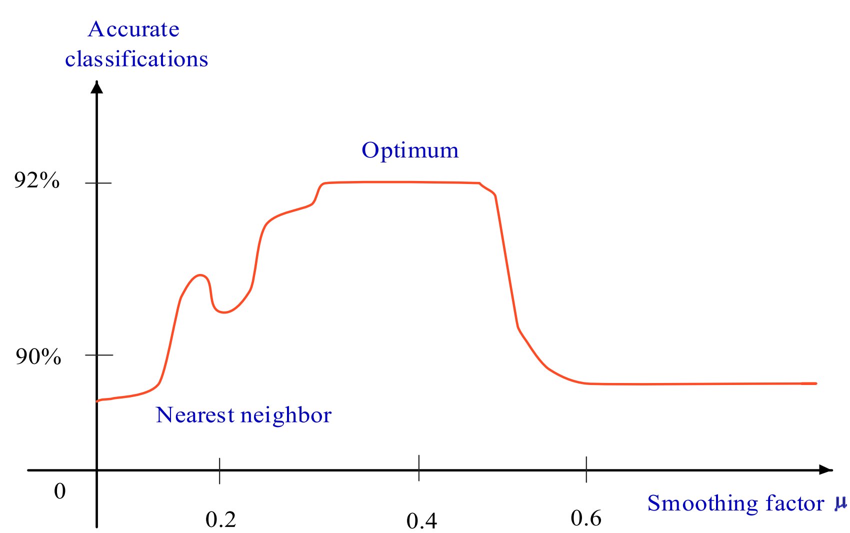
Operations of proposed FWPNN classifier

Mostly, clustering approaches are referred to as "unsupervised" technique which possesses the ability to partition a dataset into numerous subsets. The dataset Y is subdivided into m subsets which are all non-empty, pair-wise disjoint and generate Y along with the union according to Bezdek et al. (1984). One of the famous clustering techniques that tried to resolve certain difficulties of various fields are known as the FCM (Fuzzy c-means clustering) approach. More probably, this FCM approach depends on the minimization of the functionalities obtained by Bezdek et al. (1984) which means lowering the generalized "least square errors".

$$J_n(u,w) = \sum_{\ell=1}^{M}\sum_{j=1}^{C}\left(\sigma_{j\ell}\right)^n \left\| x_\ell - w_j \right\|_{A_{nr}}^2 \tag{6}$$

Whereas, $X = \{x_1, x_2, ..., x_M\}$ represents the data as well as the number of clusters in X is denoted as C at the criterion$(2 \leqslant C \leqslant m)$. The fuzziness of the resultant clusters is determined by the weighing exponent n over $(1 \leqslant n \leqslant \infty)$. The term u refers to the "fuzzy-c partition" that belongs to X, the ℓ^{th} membership of the cluster j is denoted as σ_{jk} in which

Fig. 6 Learning curve of WPNN



the centre of cluster j is represented as$w_j = (w_1, w_2, ..., w_n)$. Hence, $\|\|_{A_{nr}}$represents 'induced' A-norm of data and the notation A denotes the "positive-definite" with weight matrix $m \times m$. The A-norm for Eq. (2) is computed with respective to the 'squared distance' between x_l and w_j, which is well depicted in Eq. (3)

$$D_{j_\ell}^2 = \left\|x_\ell - w_j\right\|_{A_{nr}}^2 = \left(x_\ell - w_j\right)^T A\left(y_\ell - w_j\right) \tag{7}$$

The algorithm of Fuzzy c-means clustering is revealed in Table 2. The principle structures of data can be examined with the efficiency of the clustering algorithms. However, the knowledge acquired from the clustering algorithms can be utilized in numerous biomedical applications such as pattern recognition and classification. The performance of the classification approach is extremely closer to the training patterns. In order to train the classifier, these training patterns are highly explored. The classification algorithm turns to be highly attractive if the classification algorithm possesses the ability to detect and to remove the suspicious samples from the training dataset as soon as prior to the training. In order to tackle this problem, we proposed a novel classification algorithm that combines of FCM and WPNN algorithms. It means this proposed classification approach merges out the 'unsupervised' and 'supervised' classifiers to maximize the performance of classification technique. The schematic diagram of the proposed approach is illustrated in Fig. 1. Its operation is depicted in Fig. 7. The random selection of training data for working is depicted in Table 4. About 12training samples are depicted for illustration. It represents three distinctive classes, namely, 0 for NC, 1 for MCI, and 2 for AD.

Algorithm 1: *Fuzzy c-means clustering*

1: Initialization of parametric measures

Number of clusters C

Weighing Exponent n

Initial partition matrix u^0 such that $\sum_{C} u^0 = 1$ and $u^0 \in [0,1]$

Whereas, the termination tolerance and the number of iterations is considered as $\in > 0$

In this work, the number of iterations and the value of termination tolerance is fixed to 100 and 10^{-5} as well as the measures of C, M is prefixed as 2, 2 respectively.

2: The fuzzy cluster centers are computed by utilizing Eq. (8)

$$w_j = \frac{\sum_{\ell=1}^{M} (\sigma_{j\ell})^n \times x_\ell}{\sum_{\ell=1}^{M} (\sigma_{j\ell})^n} \tag{8}$$

3: The j^{th} cluster centre and to the ℓ^{th} data point, the distance $D_{j\ell}$ is computed by using Eq. (3)

4: Based on the obtained measure of $D_{j\ell}$ the fuzzy membership matrix u is updated. In case, $D_{j\ell} > 0$, then u is evaluated using Eq. (9)

$$u_{\ell j} = \frac{1}{\sum_{k=1}^{C} \left(\frac{D_k}{D_{\ell j}}\right)^{\frac{2}{n-1}}} \tag{9}$$

By condition if $D_{j\ell} = 0$, then complete membership measure of 1 is acquired by the data point ℓ

5: Repeat the steps from 2 to 4, until maximum number of iterations are attained (i.e. $D_{j\ell} > 0$)

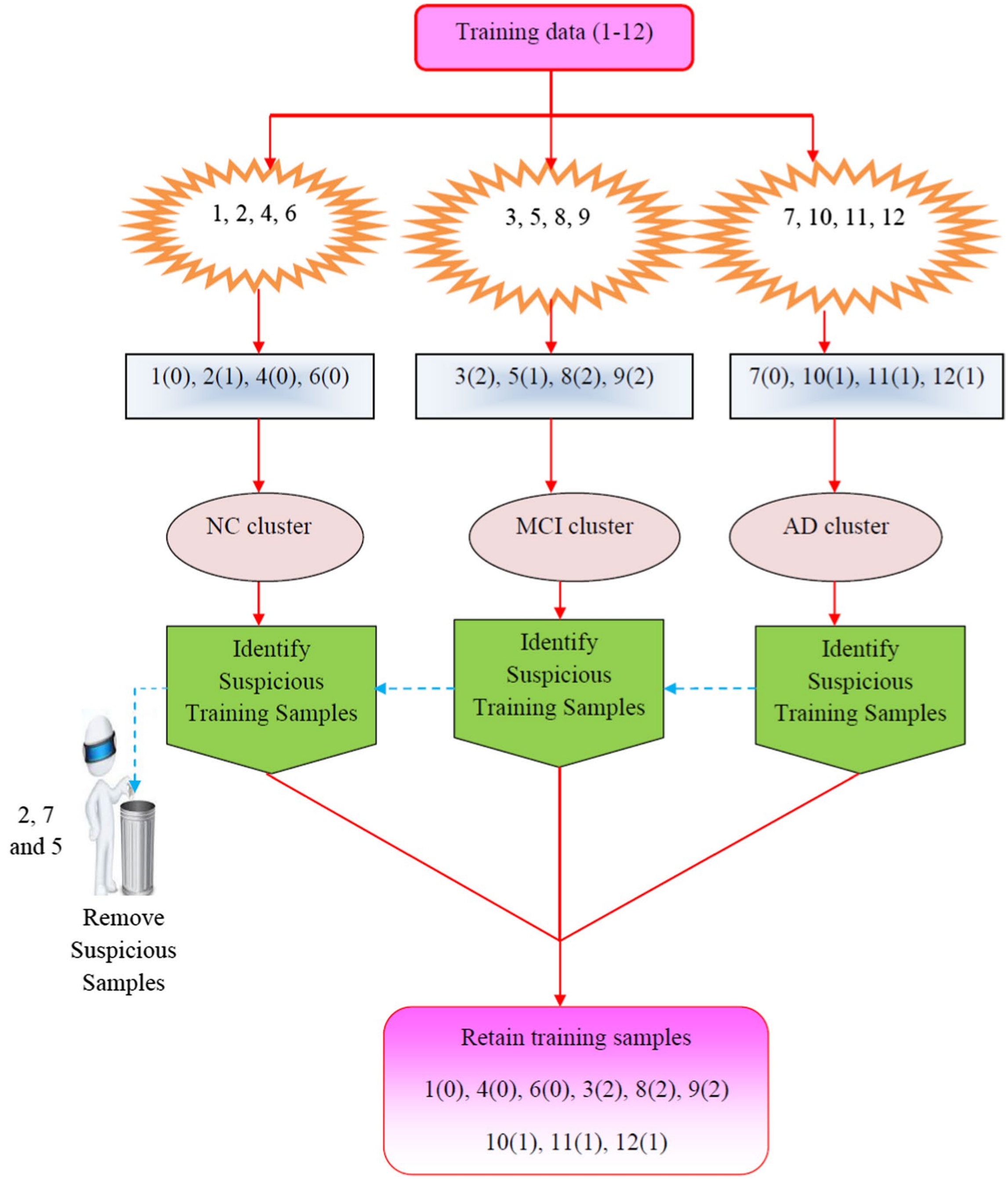


Fig. 7 Working procedures of the proposed classification algorithm

Moreover, three forms of descriptors namely Y_0,Y_1 and Y_2 are utilized to characterize each sample. Our proposed *FCM* algorithm without considering the class labels (i.e. ground truth) at the initial stage it partitions the training dataset into 3 clusters (C = 3). Figure 7 depicts the partition of clusters C_1 = (1,2,4,6), C_2=(3,5,8,9) and C_3=(7,10,11,12). Subsequently, verification of the partitioned clusters is done by keenly observing the samples in the ground truth (output). The samples (1,4,6) maintained in C_1 relate to the *Normal Control* subject, and the cluster C_1 is regarded as *NC* cluster. It is to be noted that cluster C_2 and C_3 is regarded as *MCI* and *AD* clusters, due to the fact that it takes into account most of the *MCI* and *AD* samples (i.e. 3, 8, 9 and 10, 11, 12) respectively. The clusters C_1, C_2 and C_3 include the opposite class samples (i.e. sample 2,5,7 in C_1,C_2 and C_3. Hence, it turns to be essential to remove these identified suspicious samples. In case if these suspicious samples are not eliminated, then it permits the classifier to classify the subjects *NC, MCI* and

Table 4 Ground truth generation with random selection of training data

Training samples	Attributes (input)			Ground truth (output)
	Y_0	Y_1	Y_2	
1	0.2	0.1	0.1	0
2	1	0.7	0.8	1
3	0.1	1.7	2.0	2
4	0.2	0.5	0.7	0
5	1	1.7	1.5	1
6	0.5	0.5	0.7	0
7	0.2	0.6	0.8	0
8	2.0	2.0	1.5	2
9	1.6	1.6	2.0	2
10	1.0	1.0	0.9	1
11	1.0	1.0	0.7	1
12	0.9	0.9	1	1

classification approach. The notations utilized for the illustration of steps in Algorithm 2 is delineated in Table 5.

Illustration The structural MR Images acquired from *ADNI* dataset (A_d) are preprocessed for the extraction of *ROI* from *HC* and *PCC* resulting in Bordex-3 city dataset (d_p) (step 1). More considerably, 457 features are extracted from the pre-processed structural *MR* Images and it is preserved in the dataset d (step 2). In step 3 and 4, the dataset d is partitioned up into two portions, one portion is for training (66%) and the second portion is for testing (34%). The training and the testing set are denoted by d_{train} and d_{test}, respectively. The FCM algorithm considers the clustering function to generate number of clusters in order to train and to test the data. The clustering function is applied by FCM algorithm to the training set d_{train} to generate 3 clusters (cluster 0 for NC, cluster 1 for MCI and cluster 2 for AD). In step 6, depending upon the cluster centroid and membership function values allotted by the *FCM* algorithm to distinctive samples of d_{train}, the class labels are observed and it is conserved in Z_{OCL}. In step 8, the function Match compares distinctive samples class labels which are observed in d_{train}. The class labels are matched with the ground truth Z. The samples that can be matched with the ground truth Z are preserved in d_l and at the same time, it is essential to remove the unmatched

Table 5 Acronyms used in Algorithm 2

Acronyms	Description
Z_{OCL}	Class labels observed by utilizing *FCM*
A_d	ADNI dataset
d_{train}	Dataset for training
d_{test}	Dataset for testing
d_p	Pre-processed image
d_l	Obtained training set after removal of suspicious samples form d_{train}
d	Dataset that includes of Y number of features and z ground truth (class label)
$\hat{C_l}$	Classifier relived from suspicious data
C	Fuzzy cluster center
A_j	Class label (output) evaluated from *WPNN* algorithm

Table 6 Parametric measures used to compute the classification performance of proposed classifier

Parameters	Mathematical equation	Description
Sensitivity	$\frac{TP}{(FN+TP)}$	It is used to compute the percentage of accurately (correctly) classified samples that the patient is having AD/ MCI/ NC
Specificity	$\frac{TN}{(TN+FP)}$	It is used to compute the percentage of accurately (correctly) classified samples that the patient is not having AD/MCI
False positive rate	$\frac{FP}{(TN+FP)}$	It denotes the group of patients mistakenly classified that they are having *AD*
False negative rate	$\frac{FN}{(FN+TP)}$	It denotes the group of patients mistakenly classified that they are not having *AD*
Accuracy	$\frac{TN+TP}{(FN+FP+TN+TP)}$	The term accuracy is utilized to compute the total performance of the classification framework
BAC	$0.5 \times (Specificity + sensitivity)$	Rather than these measures, the *BAC* parametric measure is utilized to compute the overall performance of both sensitivity and specificity

AD in incorrect manner. Also, it maximizes the occurrence of diagnostic mistakes (errors). Therefore, from the training set, the suspicious samples are removed and the residual good samples are utilized to train the *WPNN* classifier. Algorithm 2 depicts the operations carried out by the proposed class labels (samples) from d_{train}. In step 9, the *Construct Classifier* function constructs the classifier $C_l^{\wedge}$; hence, the classifier is constructed by training the *WPNN* utilizing d_l. In step 10, the classifier $C_l^{\wedge}$ considers the d_{test} as its input. The classifier, in turn, generates the output A_j. The performance of the proposed classification framework is evaluated by substituting the function *Performance Evaluation*. This function compares the classifiers output A_j along with the ground truth Z.

Performance evaluation

In order to compute the performance of the classifiers initially, the dataset is partitioned by using the "repeated holdout" technique. The dataset is partitioned into two subsets in a random fashion and the two partitioned subsets are referred to as training and testing sets. According to (Kohavi 1995), two third of the data is considered as the training set and the residual one-third as the test dataset and also this

Algorithm 2: *FWPNN classification algorithm*

Input: structural MR Images

Output: Specificity, Sensitivity, Accuracy, False Positive Rate, False Negative Rate and A_j

1: Preprocessing $(A_d) \rightarrow d_p$

2: Feature Extraction $d_p \rightarrow d$

3: $d_{train} \leftarrow random(d)$ /* partition the data into training $d_{train}\,(2/3)^{rd}$ and testing set $d_{test}\,(1/3)^{rd}$ portions*/

4: $d_{test} \leftarrow d - d_{train}$ /* evaluate the testing data d_{test}

5: $CLUSTERING \leftarrow \{C, M, obj-fn\}$: $(d_{train}, 3)$ //* 3 clusters namely, 0 for NC, 1 for MCI, and 2 for AD, u is a $n \times n$ fuzzy partition matrix utilizing membership values and $C \leftarrow fuzzy\ clustering\ (centre)$ using Eq. (3)

6: For each $u_j \in u$ //* for each tuple obtained from Fuzzy Partition Matrix

Evaluate [position value] $\leftarrow \max(u_j)$ //evaluate the maximum and minimum position of cluster members//

if(pos==1) then

$Z_{OCL} \leftarrow NC$ /*Training samples (observed class labels) of Z_{OCL} in d_{train} is identified as NC

else

$Z_{OCL} \leftarrow MCI$ /*Training samples (observed class labels) of Z_{OCL} in d_{train} is identified as MCI

if(pos==2)

$Z_{OCL} \leftarrow AD$ /*Training samples (observed class labels) of Z_{OCL} in d_{train} is identified as AD

else

$Z_{OCL} = \phi$

end if

end if

7: End for

8: $d_l \leftarrow match\ (d_{train}(Z_{OCL}), d_{train}(Z))$ /* The class labels which are observed are compared with the ground truth

9: $C_l^{\wedge} \leftarrow ConstructClassifier(d_l)$

10: $A_j \leftarrow C_l^{\wedge}(d_{test})$

11: *Performance Evaluation* $[A_j, Z] = \{$ Sensitivity, Accuracy, Specificity, False Positive Rate, False Negative Rate$\}$

Table 7 Selection of highly significant features by few conventional feature selection methods

Feature selection method	Category	Selected features
Recursive feature elimination (*RFE*)	Wrapper	f[140],f[35], f[13], f[19], f[50], f[52], f[38], f[20], f[56], f[37],f[42],f[55],f[53]
Random forest (*RF*)	Wrapper	f[171], f[55], f[53], f[52], f[50], f[44], f[42], f[38], f[37], f[34], f[33], f[30], f[20]
Person's correlation coefficient (*P*)	Wrapper	f[154], f[149], f[141], f[53], f[32], f[343], f[171], f[64], f[24], f[13], f[52], f[50], f[93], f[66], f[55]
Consistency (*C*)	Wrapper	f[452], f[171], f[66], f[65], f[50], f[49], f[44], f[32], f[21], f[17], f[14]
Relief F (*RFL)*	Filter	f[162], f[455], f[154], f[155], f[145], f[1339], f[158], f[448], f[163], f[140], f[141], f[146], f[457], f[11], f[453], f[49], f[50], f[53], f[52]
1*R*	Filter	F[135], f[130], f[138], f[161], f[162], f[154], f[50], f[136], f[158], f[15], f[145], f[177], f[95], f[176], f[169], f[173], f[170], f[159], f[171]
Symmetrical uncertainty (*SU)*	Filter	f[138], f[122], f[116], f[18], f[107], f[137], f[121], f[119], f[136], f[135], f[133], f[154], f[55], f[163], f[53], f[52], f[141], f[171], f[50]
Chi square score (*CHI* 2)	Filter	f[157], f[145], f[169], f[176], f[174], f[156], f[158], f[177], f[170], f[168], f[155], f[173], f[42], f[175], f[55], f[163], f[171], f[154], f[141]
Gain ratio (*GR*)	Filter	f[117], f[124], f[45], f[129], f[31], f[118], f[53],f[105], f[18], f[122], f[107], f[116], f[52], f[136], f[138], f[119], f[121], f[135], f[50]
Information gain (*IG*)	Filter	f[42], f[169], f[136], f[145], f[170], f[140], f[137], f[155], f[174], f[132], f[175], f[53], f[52], f[163], f[154], f[55], f[171], f[141]

partitioning is regarded as a universally accepted rule. In order to determine the overall performance of the classifier, the standard 'holdout' is frequently used again for about *m* times and the obtained results are averaged for each run. In this work, 15 rounds (m = 15) are carried out to minimize the unexpected variations and for each and every run the results are averaged. At each and every run, we select about 66% data samples for training and residual 34% samples for testing. Although, the training, as well as testing partition, is stabilized (fixed), and it is essential to ensure that the data samples chosen for testing and training in single run should not possess similarity with the data samples selected for next subsequent runs, this is due to the fact that few data samples in the dataset can show similarity among one another. The training and testing of data are done for each and every round of classification. At each time the evaluation is carried out with the parametric measures such as "Sensitivity", Specificity", "False Positive rate", False Negative Rate", "Accuracy" and so on. This is done to evaluate the overall performance of the proposed classification algorithm. Table 6 depicts the measures utilized to evaluate the overall performance of the proposed classification algorithm. The performance evaluation and classification experiments are done with "MATLAB®R2012a" software platform.

Whereas, *TP* (True Positive) as well as *TN* (True Negative) are referred to as accurately classified samples (subjects). Additionally, *FP* (False Positive) and *FN* (False Negative) are termed as inaccurately classified subjects. Moreover, with the support of these metrics it is possible to evaluate the level at which the features of *HC* and *PCC* are helpful at the time of predicting the subjects *NC vs AD, AD vs MCI* and *NC vs MCI*, respectively.

Results and discussion

This section evaluates the classification performance of proposed classification algorithm and presents the results based upon the comparison carried out with few conventional classification approaches. Moreover, distinctive forms of feature selection methodologies are evaluated. Furthermore, we compared the obtained results with the extracted features of mostly affected regions namely HC and PCC to the whole brain images. Subsequent to this acquired results, we made the discussion about the importance of ROI extraction from brain images for AD diagnosis.

Evaluation of classification performance with few feature selection techniques

The selection of nineteen highly relevant features for AD diagnosis is performed with "Filter" based technique. It should be noted that, we are in need to retrieve the top most signification features by a better feature selection technique

Table 8 Top most nineteen highly relevant features selected by the Multiple-criterion feature selection approach

Method	Category	Selected features
Multiple criterion	Multiple (hybrid)	f[175], f[170], f[169], f[158], f[155], f[145], f[140], f[138], f[135], f[163], f[141], f[136], f[42], f[154], f[55], f[53], f[171], f[52], **f[50]**

which can enhance the classification performance by reducing the dimensionality issues. The features sub-sets in turn are selected by the "wrapper" technique. Here, we analyzed that assignment of rank and selection of features by distinctive 'feature selection' technique which experiences slight variations. Table 7 illustrates the selection of highly relevant features from distinctive feature selection methods of the state-of art work. Consider an instance that most of the feature selection technique ranks the feature *f*[50] as highly significant feature, but this feature fails to take place in *CHI2* measures of top 19 anked features. This occurs due to the fact that selection of features is done by distinctive feature selection modules by utilizing distinctive (multiple) criterions. Thus to generate best feature set it is not so much effective to relay only on 'single criterion'. The classifier may provide the incorrect classification of subjects if the features are selected only with 'single' feature selection technique. Rather than considering the results of a single-criterion feature selection technique, the results from "multiple-criterion" feature selection technique can be considered. In order to obtain the effective feature subset, the selected features obtained by this criterion is combined by utilizing majority 'voting' obtained by the features.

Based upon the outcomes obtained with distinctive feature selection methodologies as depicted in Table 5, it is clear that it is tedious to select the most suitable feature selection technique, due to the fact that many of the feature subsets might emerge out which could possess the ability to classify the data only with certain level of accuracy. In order to rectify this issue and to enhance the robustness of feature selection procedures, we impose a *multiple-criterion* feature selection approach in this work. The outcome obtained for this feature selection approach is depicted in Table 8. From the analysis, we inferred that the feature *f[50]* is considered as a highly significant feature by this *multiple-criterion* feature selection technique. The feature *f[50]* is considered as a most significant feature because this feature gained majority voting in consideration with entire modules of the feature selection approach. Moreover, the computational burden will exceed the limit just by imposing the utility of numerous feature selection modules for the selection of top most relevant features. Therefore by considering this issue in mind, our *multiple-criterion* feature selection technique can yield better classification accuracy by developing a reliable feature set.

Moreover, numerous literal works used most popular traditional classifiers such as LDA, FLDA, k-NN, AdaBoost, bagging, SVM, etc to perform classification. However, none of the classifiers tried to adopt a mechanism to deal with the suspicious data (samples) of the training set. The classification approach employed in this work is a combination of unsupervised and supervised learning technique referred to as novel FCM based Weighted Probabilistic Neural Network (FWPNN) which possesses the ability to resolve the non-linear separable issues. Also, to detect and to remove the suspicious samples at the time of training we employed the FCM (unsupervised learning) algorithm to the supervised learning technique (WPNN) to improve the classification accuracy. Notably, a crucial process that relays on the ANN (Artificial Neural Network) is the selection of most suitable topology to solve some restraint problems in clustering (Munteanua et al. 2015). With this in mind, numerous distinctive forms of topologies have been tested and the topologies with finest generalization capability and that yields maximum accuracy are considered as the most suitable topology.

Methods taken for comparison

Furthermore, to analyze the classification performance of our proposed classification algorithm (FWPNN), we contrasted our proposed classifier model with some of the traditional classifiers developed by the researchers. The inadequacy of BP + ANN classifier proposed by (Saad et al. 2015) is that it can't operate without the prior information about training. Hence, ANN is termed as supervised learning technique and the system can be trained by applying the data to the input layer. The system is trained with pre-fixed data class sets. Preferably, two sorts of neural network based classifiers namely *KNN + FFBPNN* (E.S.A. El-Dahshan et al. 2010) were used for classification. The Feed Forward Back Propagation Network (*FFBPNN*) classifies the class labels. But this approach turns to be a time consuming process and it is so much expensive due to the fact that images were analyzed by the high resolution degrees with the help of feature extraction tool. The single classifier *BPNN* (Backpropagation Neural Network) (Zhang et al. 2011) does not provide fast convergence speed. Therefore it was combined with another algorithm Scaled Conjugate Gradient (*SCG*) to evaluate the classifier *BPNN* optimal weights. It was highly expensive and increased the computational time. The brain diseases were analyzed using *MRI* and the obtained outcomes were contrasted with the classifiers *BPNN* and *CNN* (Pan et al. 2015). Training of data was done at different layers. But the *CNN* classifier might fail by the distribution of training samples in an uneven manner. Moreover, we made an attempt to combine both the *KFCM* (Zöllner et al. 2012) and *BPANN* (Zhang et al. 2011) (i.e. KFCM + BPANN) classification algorithms to discriminate the class labels. But, it generates not so much satisfactory classification accuracy, due to the possibility of solutions trapping in the local optimum and overfitting of training data in the BPANN (Zhang et al. 2011) classifier. In order to recover this issue, in this work, we proposed novel FCM based WPNN classification algorithm, here in this approach over fitting of data is highly reduced by assigning the weighing factors between the summative and pattern layer with the knowledge of

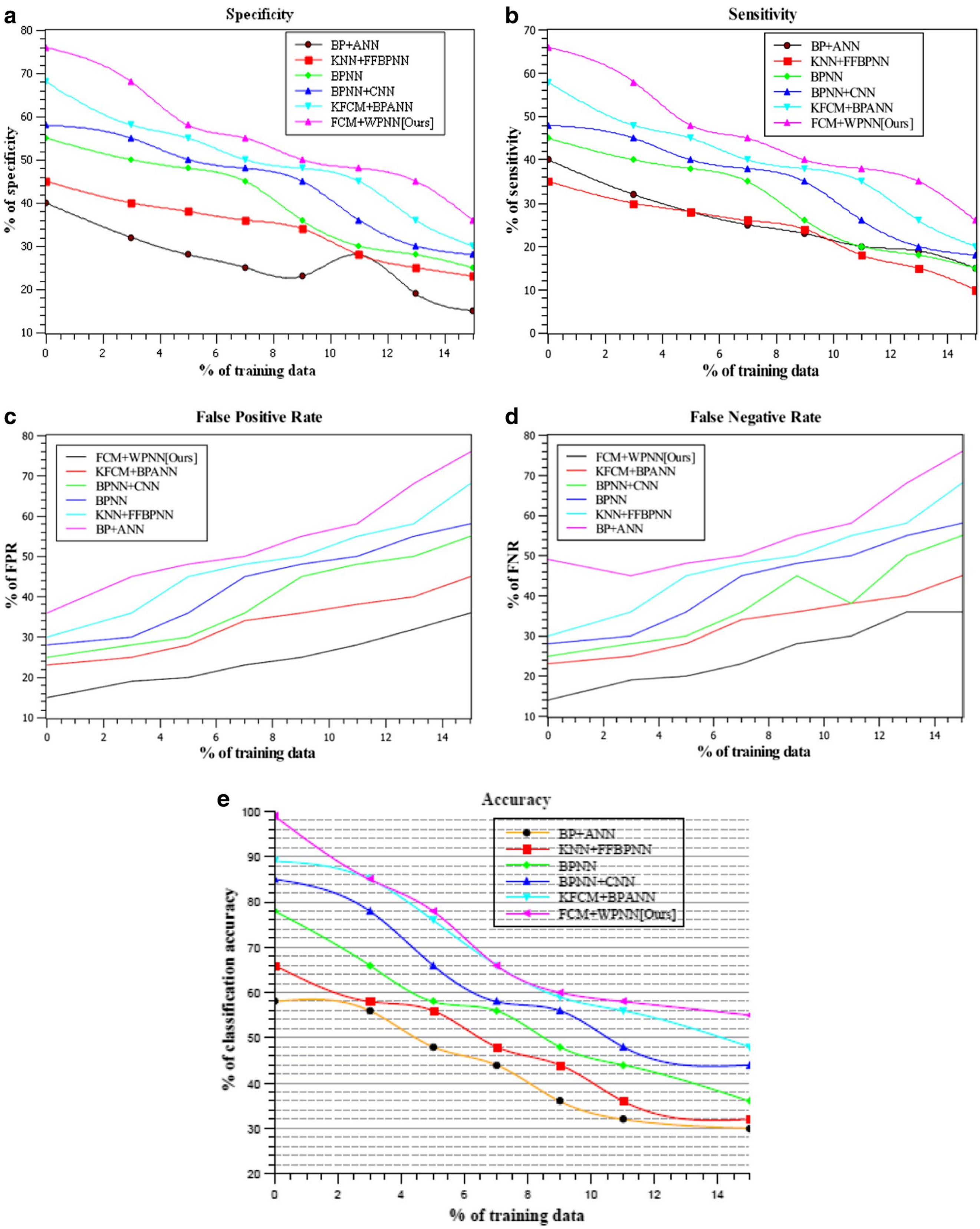


Fig. 8 Comparison of classifiers performance with parametric measures: **a** Specificity with the percentage of training data **b** Sensitivity with the percentage of training data **c** False Positive Rate with the percentage of training data **d** False Negative Rate with the percentage of training data and e) Accuracy with the percentage of training data

class separability. Moreover, our proposed classification algorithm extracts effective texture and shape features from the highly disease affected regions (i.e. from ROI extracted HC and PCC regions). Instead of retrieving entire features the *multiple criterion* feature selection approach will select only highly significant features. This approach avoids the dimensionality issues and enhances the classification accuracy. The classification performance can be further enhanced by removing unwanted suspicious training data effectively from the training set. Therefore, it makes the process more effective to classify the class labels of brain images, especially the subjects AD, MCI and AD without any doubt. To estimate the classification performance of proposed FWPNN classifier model altogether with the performance of traditional classifiers evaluation is done with few well known parametric measures namely, Specificity, Sensitivity, False Positive Rate (FPR), False Negative Rate (FNR) and Accuracy (depicted in Table 4).

Figure 8c, d delineates *FPR* and *FNR* performance of the proposed classification algorithm *FWPNN* with the conventional classification algorithms. From the analysis, it is inferred that our proposed classification algorithm attains minimum *FPR* and *FNR* when it is contrasted with *FPR* and *FNR* of the conventional classification algorithms (Fig. 8c, d). The minimum *FPR* and *FNR* are possible to be attained by the proposed system if the percentage of training data is about 66%. But if the training data is increased beyond 66%, then *FPR* and *FNR* is gradually increased subsequently by 4%. Based on the observation, we can ensure that the system performance can be enhanced if also tremendous amount of training data is administered for training. It shows signification performance with increased training data due to the removal of suspicious training data and also by the effective feature selection subsets provided for the classification. Figure 8a, b, e illustrate *Specificity, Sensitivity* and *Accuracy* of the proposed classifier model utilizing the proposed

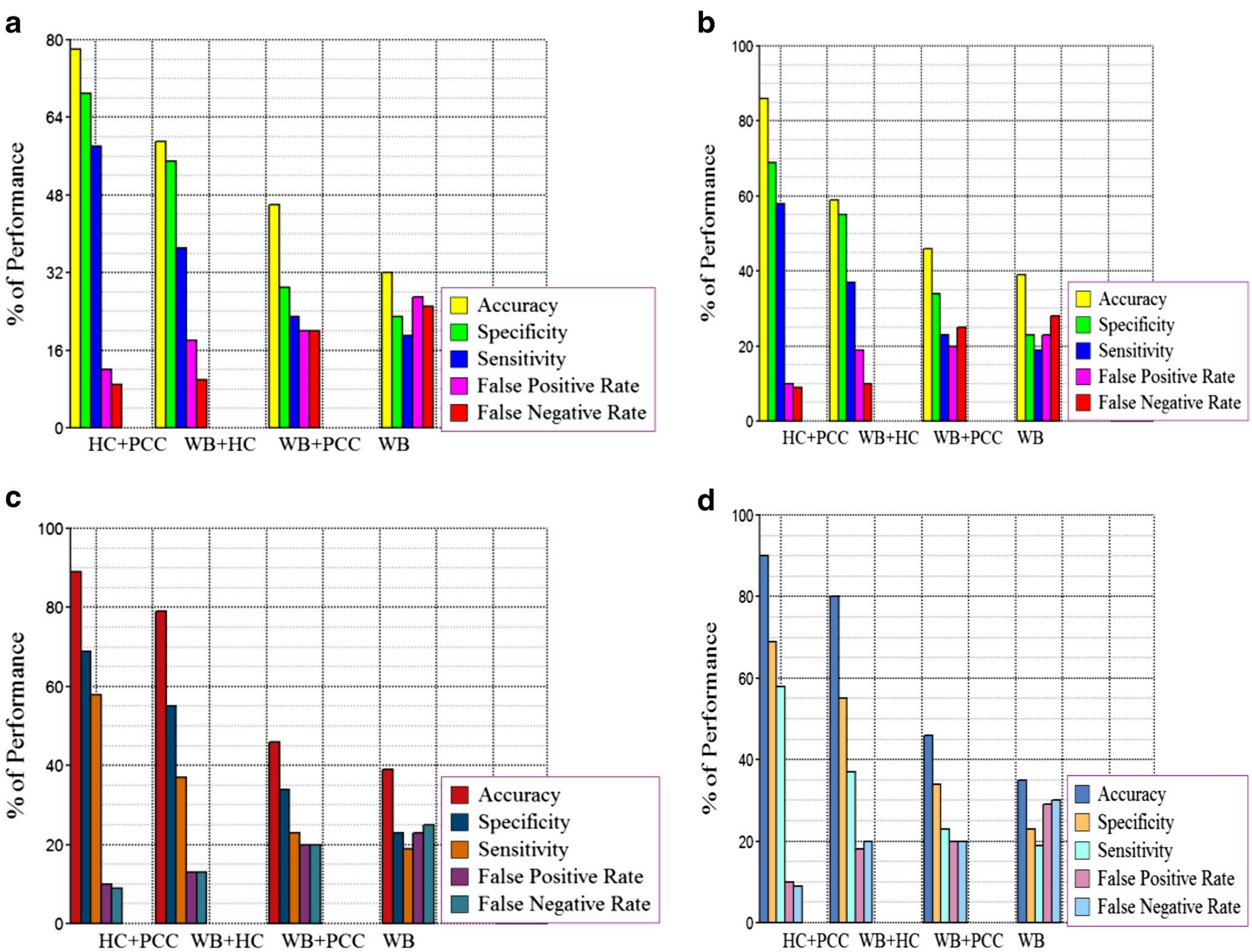


Fig. 9 Performance comparison with different subjects. **a, b** performance comparison of AD vs NC (ADNI and Bordex) groups, **c, d** performance comparison of MCI vs NC and AD vs MCI

classification algorithm with the conventional classification algorithms. From the analysis, it is observed that our proposed *FWPNN* algorithm achieves maximum *Specificity* and *Sensitivity*. Also, the accuracy of the proposed system is in maximized level when compared to the accuracy of the conventional classification algorithms even after increasing the percentage of training data. Finally, it is observed that better classification performance can be attained by the removal of suspicious samples from the training data. Also, we determined the total classification accuracy obtained with our proposed classification algorithm is 98.58%.

Comparison for AD diagnosis using different feature fusions

To further evaluate our method, comparative analysis is done with the results obtained from the extraction of features from WB (Whole Brain) images from structural MRI modality. However, we can infer that extraction of features from the scan of whole brain is a tedious and a time consuming process. This is due to the fact that, complexity arises while trying to extract huge number of features from entire images and from each slices in all projections. Moreover, we conducted an experiment with our proposed classification techniques and are analyzed with the features of the Whole Brain (WB), Whole Brain with (HC), Whole Brain with (PCC), and alone on the features of HC and PCC. Hence, determining the classification accuracy by fusing different features, we can identify the effectiveness of our proposed classification algorithm in the discrimination of class labels.

Patients with AD vs NC

Based on the *ADNI* dataset, we performed a comparison with AD patients and with the Normal Controls (NC). From the analysis, we inferred that our proposed framework achieves better results while obtaining the features of both HC and PCC alone from the whole brain for classification. Extracting *ROI* regions from the whole brain in the normalization stage will consume more time and the results obtained are not so much efficient. It may lead the classifier to classify the subjects in an inaccurate manner. From Fig. 9a, b we analyzed that extraction of *ROI* from HC and PCC features alone can increase the performance of classifier for about 85.63% specificity, 79.56% sensitivity and 98.63% accuracy. The FPR and FNR measure obtained for our proposed classification framework is lesser when compared to the extraction of *ROI* features from whole brain alone, whole brain with HC features, and whole brain with features of PCC, respectively. It should be noted that, while extracting features of the whole brain together with affected HC and PCC regions individually will gradually increases the FPR and FNR measure. This occurs due to the fact that, the classifier intakes numerous data for accurate classification of subjects and in turn it generate classification errors. From Fig. 9a, b, we inferred that our proposed approach will experience a minimum FPR and FNR of about 15.4% and 12.3% respectively. This minimized measure is obtained by our proposed classification framework due to the application of error free data in the training set.

Patients with MCI vs NC

In Fig. 9c we carried out the classification with subjects NC vs MCI based upon *ADNI* subset. The classification is done with the visual features extracted from *ROI* regions of HC and PCC, respectively. Here in this classification of subjects the proposed classification framework achieves an accuracy of about 95%, specificity, and sensitivity of about 82% and 78% respectively. The FPR and FNR measure obtained after classification of MCI vs NC subjects, alone by the features of HC and PCC is about 15% and 17% respectively (Fig. 9c).

Patients with AD vs MCI

From Fig. 9d, we analyzed that extracting *ROI* features from HC and PCC alone from the whole brain makes it more convenient for the clinician to classify the patients with AD and MCI with an accuracy of about 96.4%, specificity and sensitivity for about 85% and 76% respectively. Moreover, FPR and FNR measure obtained is about 14% and 17%. The subject MCI is commonly known as a 'transition' state in between the states of NC and AD. Finally, we can conclude that extraction of HC and PCC features alone from the whole brain decrease the number of training data and increases the performance of the classifier.

Discussion

In this paper, we proposed a multi-modal classification framework that combines of both the supervised and unsupervised learning techniques in order to improve the classification performance by removing the suspicious samples from the training set. This proposed classification framework helps the clinician to diagnose the patients with AD accurately from numerous brain images. To classify three groups of class labels namely, Normal Control (NC), Mild Cognitive Impairment (MCI) and AD (Alzheimer's disease), we used visual similarities n between the "base-line MRI". The approach was employed to the 'ADNI' subjects and then to the real cohort 'Bordex-3 city' dataset. Moreover, most of the researchers have utilized distinctive types of image analysis techniques and statistical techniques over different datasets; hence it turns to be a complicated process to perform an accurate comparison with the prior works. Additionally, to perform better discrepancy between the previous and the current work, few major information such as duration of disease, severity of disease, data size, and clinical information concerned to the subjects as well as variations in demographic might taken into account. Notably, the ADNI study is still in progression and in addition to this, certain subjects marked as MCI shows considerable progression to the AD group in future. In the following, we begin by discussing the importance of ROI selection for accurate diagnosis of AD affected patients.

Specific attention to ROI extraction and MCI subject

The ADNI database considered entire combinations related to patients classification such as AD vs MCI, NC vs MCI and AD vs NC. In order to recognize MCI category based on the structural variations turns to be a most difficult process due to the fact that the MCI category exhibits unequal structural variations in the characteristics brain regions. In recent years, AD research is moved to MCI, in the expectation of capturing the progressions of AD and defending it prior to the next progression of AD. Also, we exposed the use of ROI selection from two characteristic regions namely, HC and PCC, which shows systematic outperformance that the classification results obtained when Whole brain region is used for classification. It should be noted that, in case, if whole brain region is employed for ROI extraction it turns to be a time consuming process. Thus to recover this inadequacy, we in this proposed work we extracted ROI from the highly disease affected regions such as Hippocampus (HC) and PCC (Posterior Cingular Cortex). The essential texture and shape based features that exhibits similarity between the AD and MCI category was supported by the Multiple-criterion feature selection technique. Consequently, compared to other literal works that focused on the extraction of ROI from HC region alone by developing individual classifications depending upon the HC volume (Colliot et al. 2008), our proposed multimodal classification approach provides better performance. In fact, between AD and MCI patients about 83% of accurate classification rate is achieved by (Colliot et al. 2008), while our proposed classification framework attains 96.4% classification accuracy considerably for this state. Preferably, even though in literal works of (Fan et al. 2008), two ROIs related to HC and entorhinal cortex were used for classification, yet our proposed classification approach performs better due to the fact that the HC region is not so much spatially interrelated with PCC than the entorhinal cortex. Hence, it makes our proposed classification approach easier for the discrimination of class labels. The classification rate for AD vs MCI subjects based on cross validation accuracy obtained is about 74.3% for Voxel based approach proposed by (Fan et al. 2008). Furthermore, after employing two ROIs together to the ADNI dataset the classification rate attained is about 76.5%. In turn, if we contrast our approach with HC and PCC related ROIs altogether with the technique of (Zhang et al. 2011) which utilized the “Gray matter maps” we can provide certain justifications. In their research [60], the gray matter maps 93 ROIs were selected and classified by utilizing a single classifier SVM (in our case we multi-modeled classification algorithm referred to as FWPNN for accurate discrimination of class labels). When contrasted to the technique of (Zhang et al. 2011), our proposed classification framework achieves 98% accuracy, 83% sensitivity and 79.5% specificity (Fig. 3) and the work of (Zhang et al. 2011) attained 74.8% specificity, 62.53% sensitivity and 69.45% accuracy. Therefore, we can conclude that selection of ROIs from both the HC and PCC region is better to resolve the classification problems that we have stated. Moreover, we preferred the work of (Klöppel et al. 2008) to provide a contradiction with the Voxel based approach. In this work, for feature analysis the author used HC regions and temporal lobes. In the selection of ROIs from whole brain the classification accuracy achieved is about 63% which is lower than the classification done between the subjects MCI vs NC in the work (Zhang et al. 2011), yet our proposed attains accuracy of about 69.56%. Nevertheless, an effective classification system is required by the ROI approach hence the classification done by means of the automated software is the time consuming process and clinician can’t take a quick decision regarding the affected disease. The inadequacy of their approach emerges out due to the usage of whole brain ROIs for classification and this in turn generates large sized feature vector. Thus the approach suffers from the dimensionality issue due to large sized feature vector, therefore to withstand this issue, we in this work utilized multiple feature selection criteria to select top most highly relevant features that relates to AD.

Feature selection

The subject’s scan that corresponds to each individual is denoted as a group of discrete features that belongs to the HC and PCC ROIs characteristics. These obtained details were used to discriminate the AD and MCI subjects from NC and also among the AD patients and MCI. To provide a better contradiction for our proposed work with regard to other feature dependent approaches we prefer to refer the work of (Toews et al. 2010). Moreover our approach aims to extract texture and shape based features from ROIs of HC and PCC and the dimensionality issue is in highly minimized by imposing multiple feature selection technique to select only top most highly relevant features for classification. The authors (Toews et al. 2010), established a feature based Morphometry technique to analyze the structural changes in the characteristic of the brain regions. This method completely depends on the local SIFT descriptors which follows a learning probabilistic model and provides the details about the anatomical behaviors of the brain. The inadequacy behind

Table 9 Statistical significance obtained for classification of subjects

Subjects	Accuracy	Sensitivity	Specificity
AD vs NC	$2.755^{-5} < 0.001$	$2.330^{-12} < 0.001$	$4.932^{-9} < 0.001$
MCI vs NC	$2.31e^{-10} < 0.001$	$7.243^{-10} < 0.001$	$0.07041e^{-5} < 0.001$
AD vs MCI	$2.146^{-8} < 0.001$	$1.88e^{-9} < 0.001$	$3.755^{-5} < 0.001$

Table 10 Computational time required for proposed multi-modal classification framework

Steps	Computational time
Classification	1.3 min
Feature extraction	0.8 min
Feature selection	0.9 min
Normalization	0.5 min

this approach is that it focused only on the classification of AD vs NC subjects. The brain images utilized for this approach is obtained from OASIS dataset. Beyond this, the features retrieved from SURF (Bay et al. 2008) or from SIFT version are not so much optimal for the Magnetic Resonance Imaging (MRI); hence it lacks of the highly contrasted structures and high frequency textures. It should be noted that, this limitation occurs due to overfitting of data in the training set. To recover this issue we moved on with our proposed FWPNN classifier to generated accurate decision about the class labels. The supervised learning approach increased the convergence speed, whereas the unsupervised learning approach eliminates suspicious samples from the training set. Therefore, the classification performance is highly improved with our proposed classification algorithm.

Image normalization vs accurate classification

Our approach initially relays on the alignment of brain images to extract ROIs from HC and PCC regions, this is done automatically by the AAL method. It does not require any manual operation for the selection of ROIs from HC and PCC; hence it is a time consuming process. The AAL method adopts only selection of ROIs from disease affected regions. Consequently, different forms of structures can be modeled by the AAL approach. In contrast manual segmentation process generates imprecise classification results. Furthermore, other proposed methods demands for neuro anatomical knowledge from experts (Chupin et al. 2009) and are computationally expensive, since it takes run time of hours to days (Cuingnet et al. 2011). Therefore this method is not so much suitable in the clinical settings. With this in mind, we developed a simple classification approach that extracts the ROIs by a AAL method and classifies the subjects accurately by eliminating the suspicious samples that misleads the classifier to incorrect decisions. The AAL method captures the pathological structures, example for this is shrunken HC. Our proposed approaches overrides the difficulties in the Atlas Based segmentation process and the advantageous of our proposed work relays on the trapping of atrophy patterns related to the progressive neuro-related disorders.

Statistical evaluation

In order to evaluate the effectiveness of *ROI* extracted features, we made an assessment of statistical variation obtained for different parametric measures Sensitivity, Specificity and Accuracy. In response to each cross validation runs "Paired Student's t-tests" were experimented with distinctive classification values obtained when analyzing with features of the Whole Brain (WB), and only with extracted features HC and PCC of the whole brain. Basically, this test is performed by utilizing the outcomes acquired with an *RBF Kernel*. From Table 9, we inferred that (p-value < 0.001) for entire classification of subjects (AD vs NC, MCI vs NC and AD vs MCI). From this, we can ensure that null hypothesis can be rejected and by the extraction of HC and PCC features alone from the brain can show significant statistical improvement in the classification of AD subjects.

Computational time

The proposed techniques average computational time is depicted in Table 10. Evaluation is carried out with 2.4 GHZ Intel core i7 with 8GO memory. For about one query the average computational speed required for normalization is2.5 min, for computation of features it is about 0.8 min and for classification is about 1.3 min. It is essential to be noted that, depending upon the total number of software's, hardware's and scans the time required for computation time is evaluated. Also, from each scan, which means entire slices obtained from 3 planes (axial, sagittal and coronal) we selected only about 19 top most significant features with Multiple criterion feature approach. Therefore, the time total time required to classify the samples from features is highly diminished. Moreover, we generate an effective classification algorithm FCM based WPNN classification algorithm to improve the performance of the system.

Time efficiency

The implementation is done on the platform *C*/*C*++. The time taken for diagnosis of disease from the scan is about 11 s and 6.4 min with the image normalization procedures done by using MATLAB. Moreover, good quality results are acquired with minimized features. Also, our proposed classification framework depending upon the contrary of single point time it possesses the ability to classify the subjects.

Conclusion

In this paper, a Multi-modal classification framework is proposed for better discrimination of subjects related to brain disease, especially the patients with “Alzheimer’s disease”. This neuro-degenerative disease causes severe health concerned issues in the public. Initially, the structural *MR* images of the brain are aligned at the brain image normalization stage. The key contributions of this work were summarized as follows: (1) normalizing brain images and extracting out ROIs from HC and PCC, rather than from whole brain image by an automatic method (i.e. by AAL); (2) analyzing comprehensive set of features which are comprised with distinctive form of texture and shape based features; (3) using multiple criterion feature selection approach to estimate the top most relevant features from the feature set; (4) evaluating state of art feature selection approaches; 4) modeling a classification algorithm than can identify and remove suspicious samples from the training set which in turn enhances the classification accuracy. Hence, a novel classification algorithm FWPNN has been developed by combining “Fuzzy C-Means Clustering” (unsupervised learning technique) and “Weighted Probabilistic Neural Network” (supervised learning technique) to categorize NC, MCI and AD subjects. The unsupervised learning technique (FCM) can be effectively employed to identify faulty samples in the training set. Furthermore, these suspicious samples are essential to be removed from the training data hence it can mislead the classifier to take inaccurate decisions and in turn degrades the classification by generating errors. The experimental validation was carried out with the ADNI subset and then to the Bordex-3 city dataset. The performances are analyzed with five parameters namely, Sensitivity, Specificity, False positive rate, False negative rate and accuracy. From the analysis we inferred that our classification approach achieves accuracy of about 98.63%, 95.4%, 96.4% in terms of classification with AD vs NC, MCI vs NC and AD vs MCI. Our work is still in progression. The future scope is to consider other *MRI* characteristics for further improvement of classification performance.

References

Akgül, C. B., Ünay, D., & Ekin, A. (2009). Automated diagnosis of Alzheimer’s disease using image similarity and user feedback. *In Proceedings of the ACM International Conference on Image and Video Retrieval*, 1–34). ACM.

Amadasun, M., & King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics, 19*(5), 1264–1274.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). Computer vision and image understanding, 110(3), pp. 346–359.

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: the fuzzy c-means clustering algorithm. *Computers & Geosciences, 10*(2–3), 191–203.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16–28.

Chételat, G., Desgranges, B., Landeau, B., Mezenge, F., Poline, J. B., de La Sayette, V., … & Baron, J. C. (2007). Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer’s disease. *Brain, 131*(1), 60–71.

Chu, A., Sehgal, C. M., & Greenleaf, J. F. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters, 11*(6), 415–419.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., … & Colliot, O. (2009). Fully automatic hippocampus segmentation and classification in Alzheimer’s disease and mild cognitive impairment applied on data from ADNI. *Hippocampus, 19*(6), 579–587.

Cocosco, C. A., Zijdenbos, A. P., & Evans, A. C. (2003). A fully automatic and robust brain MRI tissue classification method. *Medical Image Analysis, 7*(4), 513–527.

Colliot, O., Chételat, G., Chupin, M., Desgranges, B., Magnin, B., Benali, H., … & Lehéricy, S. (2008). Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology, 248*(1), 194–201.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., & Habert, M. O. … & Alzheimer’s Disease Neuroimaging Initiative. (2011). Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage, 56*(2), 766–781.

Dasarathy, B. V., & Holder, E. B. (1991). Image characterizations based on joint gray level—run length distributions. *Pattern Recognition Letters, 12*(8), 497–502.

Davatzikos, C., Fan, Y., Wu, X., Shen, D., & Resnick, S. M. (2008). Detection of prodromal Alzheimer’s disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging, 29*(4), 514–523.

Dyer, C. R., & Rosenfeld, A. (1976). Fourier texture features- Suppression of aperture effects(Landsat geological terrain image power spectra). *IEEE Transactions on Systems, Man, and Cybernetics, 6*, 703–705.

El-Dahshan, E. A., Salem, A. B. M., & Younis, T. H. (2009). A hybrid technique for automatic MRI brain images classification. *Studia Univ. Babes-Bolyai, Informatica, 54*(1), 55–67.

El-Dahshan, E. S. A., Hosny, T., & Salem, A. B. M. (2010). Hybrid intelligent techniques for MRI brain images classification. *Digital Signal Processing, 20*(2), 433–441.

Fan, Y., Batmanghelich, N., Clark, C. M., & Davatzikos, C. & Alzheimer’s Disease Neuroimaging Initiative. (2008). Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage, 39*(4), 1731–1743.

Frisoni, G. B., Testa, C., Sabattoli, F., Beltramello, A., Soininen, H., & Laakso, M. P. (2005). Structural correlates of early and late onset Alzheimer’s disease: voxel based morphometric study. *Journal of Neurology, Neurosurgery & Psychiatry, 76*(1), 112–114.

Galloway, M. M. (1975). Texture classification using gray level run length. *Comput. Graph. Image Process, 4*(2), 172–179.

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H. S., … & Eustache, F. (2009). Multidimensional classification of hippocampal shape features discriminates Alzheimer’s disease and mild cognitive impairment from normal aging. *Neuroimage, 47*(4), 1476–1486.

Gonzalez, R. C., & Woods, R. E. Digital image processing using MATLAB (2nd edn.). Pearson Prentice Hall (Chap. 11), 2010.

Gutman, B., Wang, Y., Morra, J., Toga, A. W., & Thompson, P. M. (2009). Disease classification with hippocampal shape invariants. *Hippocampus, 19*(6), 572–578.

Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6), 610–621.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory, 8*(2), 179–187.

Kim, Y., Street, W. N., & Menczer, F. (2003). Feature selection in data mining. *Data Mining: opportunities and Challenges, 3*(9), 80–105.

Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., & Frackowiak, R. S. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain, 131*(3), 681–689.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai, 14*(2) pp. 1137–1145).

Kusy, M., & Kowalski, P. A. (2018). Weighted probabilistic neural network. *Information Sciences, 430*, 65–76.

Laws, K. I. (1980, July). Rapid texture identification. In" image proc. for missile guid.". (Vol. 238, pp. 376–381).

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., … & Benali, H. (2009). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology, 51*(2), 73–83.

Munteanu, C. R., Fernandez-Lozano, C., Abad, V. M., Fernández, S. P., Álvarez-Linera, J., Hernández-Tamames, J. A., & Pazos, A. (2015). Classification of mild cognitive impairment and Alzheimer's disease with machine-learning techniques using 1 H magnetic resonance spectroscopy data. *Expert Systems with Applications, 42*(15), 6205–6214.

Nestor, P. J., Fryer, T. D., Ikeda, M., & Hodges, J. R. (2003). Retrosplenial cortex (BA 29/30) hypometabolism in mild cognitive impairment (prodromal Alzheimer's disease). *European Journal of Neuroscience, 18*(9), 2663–2667.

Pan, Y., Huang, W., Lin, Z., Zhu, W., Zhou, J., Wong, J., Ding, Z. (2015). Brain tumor grading based on neural networks and convolutional neural networks. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 699–702. https://doi.org/10.1109/EMBC.2015.7318458.

Ridha, B. H., Barnes, J., van de Pol, L. A., Schott, J. M., Boyes, R. G., Siddique, M. M., … & Fox, N. C. (2007). Application of automated medial temporal lobe atrophy scale to Alzheimer disease. *Archives of Neurology, 64*(6), 849–854.

Saad, N. M., Bakar, S. A. R. S. A., Muda, A. S., & Mokji, M. M. (2015). Review of brain lesion detection and classification using neuroimaging analysis techniques. *Jurnal Teknologi, 74*(6), 73–85.

Shen, K. K., Fripp, J., Mériaudeau, F., Chételat, G., Salvado, O., & Bourgeat, P. & Alzheimer's Disease Neuroimaging Initiative. (2012). Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models. *Neuroimage, 59*(3), 2155–2166.

Srinivasan, G. N., & Shobha, G. (2008). Statistical texture analysis. In *Proceedings of World Academy of Science, Engineering and Technology, 36*, 1264–1269.

Stoitsis, J., Golemati, S., & Nikita, K. S. (2006). A modular software system to assist interpretation of medical images—Application to vascular ultrasound images. *IEEE Transactions on Instrumentation and Measurement, 55*(6), 1944–1952.

Studholme, C., Drapaca, C., Iordanova, B., & Cardenas, V. (2006). Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Transactions on Medical Imaging, 25*(5), 626–639.

Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, p. 37.

Toews, M., Wells, W., Collins, D. L., & Arbel, T. (2010). Feature-based morphometry: discovering group-related anatomical patterns. *NeuroImage, 49*(3), 2318–2327.

Toga, A. W., Thompson, P. M., Mega, M. S., Narr, K. L., & Blanton, R. E. (2001). Probabilistic approaches for atlasing normal and disease-specific brain variability. *Anatomy and Embryology, 204*(4), 267–282.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273–289.

Weszka, J. S., Dyer, C. R., & Rosenfeld, A. (1976). A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (4), 269–285.

Wu, C. M., Chen, Y. C., & Hsieh, K. S. (1992). Texture features for classification of ultrasonic liver images. *IEEE Transactions on Medical Imaging, 11*(2), 141–152.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*, 55(3), 856–867.

Zhang, Y., Dong, Z., Wu, L., & Wang, S. (2011). A hybrid method for MRI brain image classification. *Expert Systems with Applications, 38*(8), 10049–10053.

Zöllner, F. G., Emblem, K. E., & Schad, L. R. (2012). SVM-based glioma grading: optimization by feature reduction analysis. *Zeitschrift Für Medizinische Physik, 22*(3), 205–214.