Semi-Supervised Adaptive Neuro Fuzzy Inference System (SSANFIS) for Effective Classification of Alzheimer's Disease

C. Geetha, D. Ramyachitra

Abstract--- Alzheimer's disease (AD) is the most popular dementia in elderly people worldwide. It affects memory of the patient. The early detection of Alzheimer's disease still a challenge because of the estimation of the scans depends on manual directing and visual reading. To overcome this issue, Semi-Supervised Adaptive Neuro Fuzzy Inference System (SSANFIS) is introduced for effective classification of alzheimer's disease. In this work, an Improved Artificial Bee Colony (IABC) algorithm is used for preprocessing step which provides higher classification performance. It removes noises from the images based on the employed bees, onlooker bees and scout bees calculation. Then using the Adaptive Median Filter (AMF) enhances the quality of the images by comparing each pixel in the image to its surrounding neighbor pixels. The Hybrid Wavelet Transform (HWT) is enhanced using Haar with Walsh wavelet transform; it is used to extract the features from the MRI images. Unsupervised Fuzzy C Means (USFCM) is applied for selecting the important features from the extracted features. The model is learned by using Semi-Supervised Adaptive Neuro Fuzzy Inference System (SSANFIS) approach. It selects optimal fuzzy rules based on the higher frequency rules which are used to increase the accurate classification results which provide higher results than other methods.

Keywords--- Alzheimer's Disease (AD), Adaptive Median Filter, Fuzzy rules and Improved Artificial Bee Colony (IABC).

I. INTRODUCTION

Alzheimer's Disease (AD) is one among the most prevalent neurodegenerative diseases that affects elder adults during their mid-60s [1]. Degradation of the nervous system is chiefly caused because of the accumulation of amyloid plaques and neuro fibrillary tangles [2]-[5]. Symptoms usually consist of memory loss, thinking impairment, poor judgement and trouble in identifying people. AD is actually an untreatable dementia. But, several drugs have been discovered for slowing down the impact of the disease. No specialized treatments are available for AD [6]. Hence, the detection and diagnosis of AD at a primitive stage is required for the treatment and is quite helpful in managing the disease [3]. The transition happening to AD from normal aging is gentle; therefore, several researchers have focused their attention on the prior detection of AD. AD's early stage is indicated by the term, Mild Cognitive Impairment (MCI) [7]. Two diverse subtypes of MCI include Amnestic Mild Cognitive Impairment (aMCI) and Non amnestic Mild Cognitive Impairment (naMCI). aMCI patients are affected with considerably degenerated memory, but their cognitive skills are strong. On the other side, in naMCI, one or more number of cognitive functions is reduced but the memory problem stays as it is. Patients affected with MCI have potential chances of getting affected with AD eventually. Therefore, detection of AD during the early stage is quite a challenge and needs investigations further through well-known neuroimaging methodologies [8]. Medical imaging strategies have a significant role to play in the AD's early detection. Few of the generally employed imaging modalities include Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Single-Photon Emission Computed Tomography (SPECT), Functional MRI (fMRI), Positron Emission Tomography (PET) and Electroencephalogram (EEG).

In the recent times, different classification techniques have also been studied to find the people affected with AD from Normal Control (NC) employing MRI data [9]-[10]. In many of these available classification techniques, two important steps are usually included :) extraction and selection of differentiating features from the actual neuroimaging data, and 2) learning of an optimal separating hyperplane in a high dimensional feature space for carrying out AD classification. The high-dimensional pattern classification techniques, e.g., Support Vector Machines (SVM), have been extensively explored for analyzing the structural and functional brain images to help in diagnosing AD inclusive of its prodromal stage, i.e., MCI. Several available classification techniques perform the extraction of the features from neuroimaging data and then build a single classifier to carry out the classification. But, owing to noise and smaller sample size of neuroimaging data, it is a challenge to train just a global classifier, which can be sufficiently robust to accomplish a good classification performance. Rather than having one single global classifier built, Liu et al developed a local patch-based subspace ensemble technique that develops several individual classifiers depending on various subsets of local patches and thereafter integrates them for more accurate and reliable classification [11]. But, many of the available pattern classification techniques just make use of one individual modality of biomarkers for diagnosing AD or MCI that might have an effect over the entire classification performance.

Zhang et al [12] demonstrated the combination of three modalities of biomarkers, i.e., MRI, FDG-PET, and CSF biomarkers, to differentiate between AD and healthy controls, employing a kernel combination technique.

Manuscript received September 16, 2019.

C. Geetha, Department of Computer Science, Research Scholar, Bharathiar University, Sri Kayaka Parameswari Arts and Science College for Women,Chennai, T.N, India. E-mail:geethacvphd@gmail.com

D. Ramyachitra, Assistant Professor Department of Computer Science, Bharathiar University, Coimbatore, T.N, India.

SEMI-SUPERVISED ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (SSANFIS) FOR EFFECTIVE CLASSIFICATION OF ALZHEIMER'S DISEASE

Particularly, for every MR or FDG-PET image, 93 volumetric features get extracted from the 93 Regions of Interest (RoI), which are automatically marked by means of an atlas warping algorithm. In the case of CSF biomarkers, their actual values are directly utilized in the form of features. Afterwards, a linear Support Vector Machine (SVM) is used for evaluating the classification accuracy. Consequently, for the classification of AD from healthy controls, the system attains a classification accuracy of 93.2% while having all the three modalities of biomarkers combined, and just 86.5% when making use of even the best ones of the individual modality of biomarkers.

Kong et al [13] developed an Independent Component Analysis (ICA) [4] to identify the vital genes for the diagnosis of Alzheimer's disease. ICA finds the gene expression profiles to be linear combinations of elementary expression patterns, which may be considered to be significant regulation pathways. Classifying Alzheimer's disease and the control samples becomes easier and efficient with the help of fewer amounts of gene data employing the significant genes extracted. This technical is the first one helping to validate the ICA paradigm over the analysis of microarray gene expression Alzheimer's disease data. Suggested a Support vector machine to be used for classification of Alzheimer's disease from whole-brain anatomical MRI. The system concentrated on the features of the distribution of the Gray Matter (GM), White Matter (WM), and Cerebro Spinal Fluid (CSF) that intuitively is meaningful while tackling with neurodegenerative diseases generally and AD in specific. Especially, the technique does the parcellation of the subjects' brain MRI into RoIs. As seen from this analysis, one parameter gets extracted, which provides a characterization of the subject and implies a physiological meaning, as it denotes the relative weight of GM in comparison with WM and CSF. At last, depending upon this parameter that estimated in each of the RoIs of the entire brain, the technique performs the classification of the subjects with the assistance of an SVM algorithm.

Lopez et al [14] suggested an automated tool to help the interpretation of Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) images for diagnosing AD. The major issue that has to be dealt with is the so-known small size sample that comprises of about a small number of available images in comparison with to the huge extent of features. This issue is encountered when aggressively decreasing the dimension of the feature space through the Principal Component Analysis (PCA). The system designed is dependent on Bayesian classifiers that make use of a posteriori information to decide, which class the subject belongs to, rendering 88.6 and 98.3% accuracy values for SPECT and PET images, correspondingly.

II. PROPOSED METHODOLOGY

Image enhancement is performed by using Adaptive Median Filter (AMF) and feature extraction utilizing Hybrid Wavelet Transform (HWT). Improved Artificial Bee Colony (IABC) algorithm is utilized for preprocessing step that gives greater classification performance. It eliminates noises from the images dependent upon the employed bees, onlooker bees and scout bees calculation. After that utilize AMF is introduced for improving the quality of the images by matching up with every pixel in the image to its nearby pixels. The HWT is improved by means of Haar with Walsh wavelet transform; it is utilized to take out the features from the MRI images. In order to choose the necessary features from the taken out features Unsupervised Fuzzy C Means (USFCM) clustering method is presented. By means of SSANFIS technique, this method is learned. It chooses optimal fuzzy rules dependent upon the greater frequency rules that are utilized for improving the exact classification outcomes. It gives greater outcomes while matched up with other algorithms.

2.1. Image Denoising - Improved Artificial Bee Colony (IABC)

Pre-processing methods is utilized for enhancing the identification of the distrustful area from Magnetic Resonance Image (MRI). The pre-processing as well as enhancement technique comprises two steps; primary step is the elimination of film artifacts for instance labels and X-ray marks are eliminated from the MRI by means of tracking algorithm. Next the elimination of high frequency components is performed by using Improved Artificial Bee Colony (IABC) algorithm. The performance of the presented method is assessed by Peak Single-To Noise-Ratio (PSNR), Average Signal-to-Noise Ratio (ASNR).

The primary step of pre-processing by means of tracking functions include those operations, which are vital beforehand the main data inspection and taking out of details, and are clustered as radiometric or geometric developments. MRI comprises film artifacts or label on it, such as, patient name, age and marks. Film artifacts are removed by utilizing tracking algorithm. Currently, start from the first row and first column, the intensity value of the pixels is examined, so the threshold value of the film artifacts is identified. The threshold value is beyond that of the threshold value is removed from MRI. The greater intensity value of film artifacts are removed from MRI. Sometime in the removal of film artifacts, the image comprises salt and pepper noise.

In this research, ABC is a population based optimization algorithm [15] and attempts to attain worldwide minimum or maximum repeatedly for getting greater noise eliminated image samples. The termination conditions for ABC must be Maximum Cycle Number (MCN) or satisfactory error value. The amount of images in AD prediction in ABC hive comprises three main categories of bees; employed bees, onlooker bees and scout bee. The employed bees and onlooker bees exploit nectar sources identified all over the hive and the scout bee discovers the solution space. In case of ABC algorithm, amount of nectar sources everywhere the hive matches to amount of employed bees. Also the amount of employed bees matches to amount of onlooker bees. In a cycle, amount of scout bee is capable of be accomplished. This algorithm contains four stages successively realized; initialization, employed bee, onlooker bee and scout bee phases

Initialization: this algorithm starts to perform by arbitrarily choosing noises in the image samples. The primary solutions are generated for employed bees by means of Equation (1).

$$p_{i,j} = p_j^{min} + \lambda(p_j^{max} - p_j^{min}), i = 1, \dots N, j = 1, \dots D \quad (1)$$

Here, $p_{i,j}$ is j^{th} image of ith pixel in employee bee, p_j^{min} and p_j^{max} is lower and upper bounds of jth parameters, correspondingly, N is known as the amount of employed bee and 'D' is the dimensionality of the optimization problem, λ is known as a random number in range of [0,1],. Furthermore, in this stage, the Abandonment Counter (AC) of every employed bee is rearranged.

Employed Bee Phase: In this stage a novel quality of the pixels solutions are generated from every employed bee. Initially, the solution of employed bee is copied to novel noise eliminated image solution ($v_i = p_i$). After that, merely one parameter of solution is brought up to date by means of Equation (2).

$$v_{i,j} = p_{i,j} + \phi(p_{i,j} - p_{k,j}), i, k \in \{1, \dots N\}, j \in \{1, \dots D\} \quad (2)$$

Here $v_{i,j}$ j^{th} image of ith pixel solution, $p_{i,j}$ is j^{th} image of ith pixel in employee bee, $p_{k,j}$ is j^{th} image of kth pixel in employee bee, N is called the amount of employed bee and 'D' is the dimensionality of the optimization problem. φ is known as a random number in range of [−1,+1], Furthermore, the neighbor of quality finalized pixel solution (k) and dimension of the image (j) are arbitrarily chosen amongst the employed bee images (population) and amid qualities of the images. Subsequently the novel quality outcomes are generated and computed the objective function (PSNR) values particular for the noise elimination, fitness values of noise eliminated solutions of employed bees are computed in this manner:

$$fit_i = \begin{cases} \dfrac{1}{1 + PSNR_i} if(PSNR_i \geq 0) \\ 1 + abs(PSNR_i) elsewhere \end{cases} \quad (3)$$

Here fit_i is known as the fitness value of ith pixel (employee bee), $PSNR_i$ is called the objective function value of ith the pixel (employee bee). When the fitness value of candidate solution is superior to fitness value of employed bee, the employed bee is substituted with candidate solution ($v_i = p_i$) and the AC of the employed bee is retuned, if not AC is incremented by 1 (AC = AC+1).

Onlooker Bees Phase: Every onlooker bee selects an employed bee with the purpose of enhancing its solution. This choice is accomplished in keeping with PSNR values of employed bees by roulette wheel provided in Equation (4).

$$pr_i = \dfrac{Fit_i}{\sum_{j=1}^{N} fit_j} \quad (4)$$

here, pr_i is known as the likelihood of being chosen i^{th} pixel (employee bee). The onlooker bees attempt to enhance solution of the chosen image quality by using Equation (5.18). When the PSNR value of the novel solution identified by the onlooker bee is superior PSNR value of the solution of the employed bee, the onlooker bee is altered with the employed bee and the abandonment counter of the

employed bee is retuned, if not the AC is incremented by 1 (AC = AC+1). Concisely, onlooker and employed bees utilizes identical equation for exploitation PSNR value. Though onlooker bees observe employed bees' dance in dance area of the hive, neighborhood selection is completely arbitrary (p_i Equation (5)).

Scout Bee Phase: The ACs of the entire employed bees is assessed with the amount that is chosen by designer. The employed bee that could not increase self-solution till the AC attains to the limitation; turn out to be scout bee. Henceforth, a quality is made for the scout bee by means of Equation (1) and the AC is retuned. The scout bee that a noise eliminated image solution was created for itself; turn out to be the employed bee. So, scout bees in ABC algorithm prevent stagnation of employed bee population.

The employed bees exploit nectar sources and transfer location info regarding the nectar sources to the hive and share the location info with onlooker bees by dancing in the dance area of the hive. The onlooker bees select quality improved solutions of employed bees by roulette wheel selection dependent upon PSNR values of employed bees and attempt to enhance the noise elimination outcomes. Though onlooker bees observe the dance, neighbor selection of onlooker bees is completely arbitrary (p_k in Equation (2)). Concisely, the onlooker's bee's doesn't taken info shared for own neighbor selection. The presented a board to be utilized neighborhood selection by stimulating ABC model [16]. Subsequent to employed bee phase, average PSNR values got from employed bees is identified in this manner:

$$avg_t^{pop}(PSNR) = \frac{1}{N} \sum_{i=1}^{N} fit_i \quad (5)$$

Here, $avg_t^{pop}(PSNR)$ is an average PSNR value of employed bees image at iteration t and N is amount of employed bees. The PSNR values of employed bees are assessed $avg_t^{pop}(PSNR)$ and the solutions of employed bees that are superior to $avg_t^{pop}(PSNR)$ bestated in the board. Time period on the board of solutions is computed by Equation (6).

$$D_i = K. fit_i \quad (6)$$

Here, D_i is known as waiting time on the board of the solution and waiting time of the solutions is proportional to PSNR values of employed bees. K is known as a positive constant number, fit_i called is fitness value of ith employed bees. Consequently, neighbors for onlooker bees (p_k in Equation (2)) are no longer chosen from the board.

2.2. Adaptive Median Filter (AMF)

Consequently, for eliminating the noises that improve the quality of the image in image processing applications, the Adaptive Median Filter (AMF) algorithms are used extensively. The AMF algorithm carries out spatial processing for identifying which pixels in an image are affected by impulse noise. The AMF classifies pixels as noise by matching up with every pixel in the image to its neighbor pixels. The neighborhood size is flexible, and the

threshold for the comparison. A pixel, which is diverse from most of its neighbors, and being not physically brought into line with those pixels to which it is alike, is tagged as impulse noise. These noise pixels are substituted by the median pixel value of the pixels in the neighborhood, which passed the noise labeling test.

Algorithm 1: AMF Algorithm

Input : AMF changes size of S_{xy} (the size of the neighborhood) for the period of operation

Notation: Z_{min} = minimum gray level value in S_{xy}, Z_{max} = maximum gray level value in S_{xy}, Z_{med} = median of gray levels in S_{xy}, Z_{xy} = gray level at coordinates (x, y), S_{max} = maximum allowed size of S_{xy}

Level A

$A_1 = Z_{med} - Z_{min}$, $A_2 = Z_{med} - Z_{max}$

    if $A_1 > 0$ and $A_2 < 0$, go to level B

        Else increase the window size

            if window size $< S_{max}$, repeat level A

                else output Z_{xy}

Level B

      $B_1 = Z_{xy} - Z_{min}$

      $B_2 = Z_{xy} - Z_{max}$

      if $B_1 > 0$ and $B_2 < 0$, output Z_{xy}

        Else output Z_{med}

Explanation

Level A

    If $Z_{min} < Z_{med} < Z_{max}$, then Z_{med} is not an noise

        (1) go to level B to test if Z_{xy} is an noise

Else

    Z_{med} is an noise

    (1) the size of the window is increased and

    (2) level A is repeated until ...

        (a) Z_{med} is not an impulse and go to level B or

        (b) S_{max} reached: output is Z_{xy}

Level B

      If $Z_{min} < Z_{xy} < Z_{max}$, then Z_{xy} is not an noise

        (1) Output is Z_{xy} (distortion reduced)

      Else

      Either $Z_{xy} = Z_{min}$ or $Z_{xy} = Z_{max}$

        (2) Output is Z_{med} (standard median filter)

    Z_{med} is not an noise (from level A)

The standard median filter does not carry out well while noise is greater.

2.3. Hybrid Wavelet Transform (HWT)

The HWT is improved by means of Haar with Walsh wavelet transform; it is utilized to take out the features from the MRI images. It is utilized for calculating the HWT with coefficients that is enthused by certain relations amid Walsh-Hadamard Transform (WHT) and Haar wavelet transforms, on the other hand that is basically diverse from traditional wavelet transform approaches. Its marked change is obvious from the reality that it comprises completely of Haar wavelet transforms. Demonstrate display that the algorithm that calls the Cascading Haar Wavelet (CHW) algorithm equals the serial complexity of the standard algorithms for the natural or dyadic signals, needing accurately $n \log_2 n$ addition operations for its calculation. Specified $m \geq 0$, the $2^m \times 2^m$ Hadamard matrix with columns in dyadic (Paley) order, H_m, is described by the recursion

$$H_0 := 1, H_{m+1} = \frac{1}{\sqrt{2}}\begin{bmatrix} H_m \otimes (1\ 1) \\ H_m \otimes (1\ -1) \end{bmatrix} form \geq 0 \quad (7)$$

Here $\otimes$ specifies the Kronecker product. Specified $m \geq 0$, the $2^m \times 2^m$ Haar matrix, Ψ_m, is described by the recursion

$$\Psi_0 := 1, \Psi_{m+1} = \frac{1}{\sqrt{2}}\begin{bmatrix} \Psi_m \otimes (1\ 1) \\ I_m \otimes (1\ -1) \end{bmatrix} form \geq 0 \quad (8)$$

Here write I_m for the $2^m \times 2^m$ identity matrix. The CHW algorithm is dependent upon a specific decomposition of a Hadamard matrix in regard to Haar wavelet transform matrices

$$H_m = \left\{ \prod_{r=1}^{m-1} I_{r-1} \otimes \begin{bmatrix} I_{m-r} & 0 \\ 0 & \Psi_{m-r} \end{bmatrix} \right\} \Psi_m, form \geq 1 \quad (9)$$

WHT algorithm is calculated by primarily calculating the Haar wavelet transform, and after that using a divide and-conquer method as well comprising Haar wavelet transforms, as depicted in Figure 1. That is possibly deceptive, because the identity transforms don't need to be carried out.
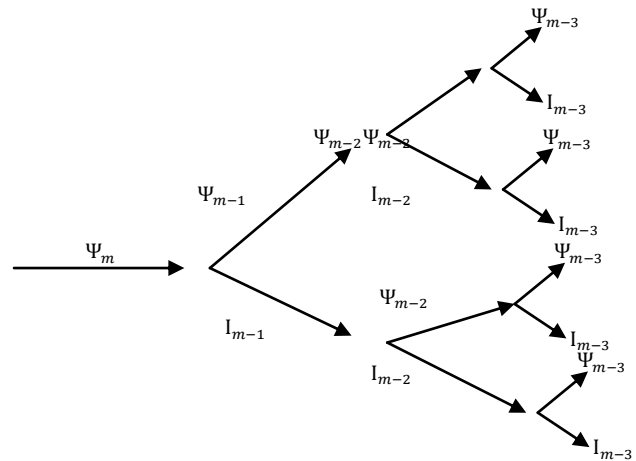

Figure 1: An Illustration of the Cascading Haar Wavelet Algorithm

In this segment, we create the observation that there is a normal means to parallelize the CHW algorithm that owns an amount of attractive features.
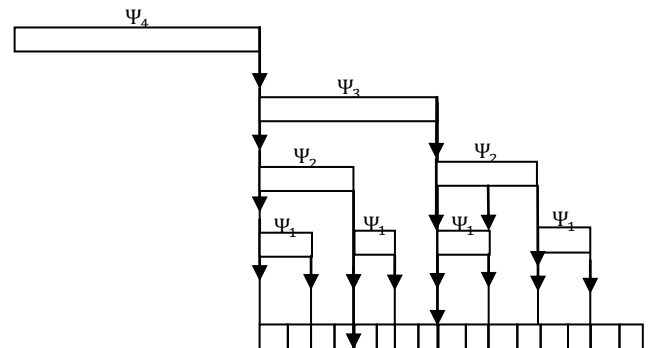

Figure 2: An Illustration of the Proposed Parallel Implementation of the CHW Algorithm

According to CHW algorithm, a signal of length 2m is cascaded via a succession of Haar wavelet transforms. It is likely to take a parallel architecture where in every $m - 1$ node is dedicated to do Haar wavelet transforms of some size.

A scheduling chart exemplifying this process for m = 4 is depicted in Figure 2.

Contain three nodes; every one id dedicated to carry out the Haar wavelet transforms Ψ_1, Ψ_2 and Ψ_3. A full Haar wavelet transform Ψ_4 must be carried out (by one among the three nodes, or by an additional one), and afterward every node is engaged for almost half of the whole running time.

The yield is the WHT coefficients in dyadic order. The gorgeous features of this method: every node require merely be programmed to carry out a solitary task, and communication of the output from any specified node obeys fixed as well as straightforward rule.

2.4. Feature Selection- Unsupervised Fuzzy C Means (USFCM)

As soon as the features are taken out from the samples, after that feature selection is carried out by means of USFCM in which targets are emerging feature selection that could be utilized in labeled as well as unlabeled samples.

In this feature selection, a known pre-identified set of groups with unlabeled training feature vectors of AD are used for prompting a classification function [17].

The unsupervised clustering is capable of performing the unlabeled training feature vectors of AD grouping exploiting the types in the primary unlabeled training feature vectors of AD, along with expanding as well as changing the existing set of class with the intention of showing the residual abnormalities in the dataset.

The fuzzy partition matrix U determination using USFCM is an iterative optimization process. The foundation of USFCM is utilizing the unlabeled training feature vectors of AD to manage the iterative optimization process. The objective function of the USFCM optimization problem is stated as:

$$J_m(U, v) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^m d_{ik}^2 \qquad (10)$$
$$+ a \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik} - f_{ik})^m d_{ik}^2$$

Here, U Fuzzy Partition Matrix, v cluster center, u_{ik} Membership degree of kth data point be in the ith cluster value amid 0 and 1, d_{ik} Distance measure of kth data point from i^{th} cluster center, f_{ik} Membership degree of kth labeled sample be in the ith cluster value is 0 or 1. The coefficient 'a' denotes the scaling factor and 'm' denotes the fuzzy coefficient.

The function of 'a' is keeping a part of unsupervised component in the optimization technique and parameter 'm' controls the amount of fuzziness in the feature selection.

The conventional value of m is 2 and a=L/n, L denoting the size of labeled samples.

The function J_m contain a large amount of values, with the least one getting associated with finest clustering.

Algorithm for USFCM

An effective method for fuzzy feature selection demoted as iterative optimization process is studied in this manner. The following are steps of this algorithm 2.

Algorithm 2: USFCM Algorithm for Feature Selection in AD Prediction

1. Input : Set the number of clusters c. Initialize the membership values of feature matrix F of size $c \times n$ with 0 or 1 based upon the class labels
2. Begin
3. Initialize the fuzzy partition matrix $U^{(0)}$ with randomized values amid 0 and 1, here $U \in M_{fc}$

$$M_{fc} = \left\{ U|_{u_{ik}} \in [0,1]; \sum_{i=1}^{c} u_{ik} = 1; 0 \qquad (11) \right.$$
$$\left. < \sum_{k=1}^{n} u_{ik} < n \right\}$$

4. Start the iterative procedure and fix the iteration count, t = 1.
5. Calculate the centers of the clusters applying the equation (11) observed below is the ith cluster center described by 'm' features and organized in the feature vector form represented as,

$$v_{ij}^{(t)} = \frac{\sum_{k=1}^{n} \left(U_{ik}^{(t-1)} \right)^m Z_{kj}^{(train)}}{\sum_{k=1}^{n} \left(U_{ik}^{(t-1)} \right)^m}, j = 1,..m \qquad (12)$$

$v_i^{(t)}$ denotes the ith cluster center described by m features of AD

database and structured in the vector form represented as $v_i^{(t)} = \left\{ v_{i1}^{(t)}, ..., v_{im}^{(t)} \right\}$ denotes the kth AD instance regarding the mth chosen feature variable. The input feature matrix, Z, denotes the input train pixel vector set derived for object identification process.

6. Calculate the distance, $d_{ik}^{(t)}$, amid the ith cluster center and k^{th} feature vector. The distance measure used is Euclidean Distance as given by equation (13).

$$d_{ik}^{(t)} = \left\| Z_k - v_i^{(t)} \right\| = \sqrt{\sum_{j=1}^{m} \left(Z_{kj}^{(train)} - v_{ij}^{(t)} \right)^2} \qquad (13)$$

7. Bring up-to-date the fuzzy partition matrix, $U^{(t+1)}$, for the subsequent iteration in this manner:

$$u_{ik}^{(t+1)} = (1-a) \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^{(t)}}{d_{jk}^{(t)}} \right)^{\frac{2}{m-1}} \right]^{-1} + a f_{ik} \forall I_k \qquad (14)$$
$$= \emptyset$$

8. If $U^{(t+1)} - U^{(t)} \leq \varepsilon$ (ε being iterative accuracy), terminate the iteration
9. Output v (cluster center), U (fuzzy matrix); or else increment the iteration count, t = t+1 and return to Step 3.

$$u_{ik} = \sum_{j=1}^{c} \left(\frac{\left\| Z_k^{(test)} - v_i \right\|}{\left\| Z_k^{(test)} - v_j \right\|} \right)^{\frac{2}{m-1}} \qquad (15)$$

10. End begin

Here v_i and v_j denote the values derived from last cluster center v of the USFCM iterative training algorithm. The feature selection performance designed using USFCM algorithm is gauged by the assessment of the subsequent measures for train set, test set of features in AD detection. Unsupervised learning is a form of the machine learning task of obtaining a function utilized for defining the unseen structure from the unlabeled training feature vectors of AD.

Since the examples given to the learner are unlabeled, for assessing a good solution, there is no error or reward signal.

The unsupervised classification is used when there exists no former info concerning the classes, which are existent in the training feature vectors of AD used for feature selection.

So, it is an automated process of classification and does not need any labelled data set in training feature vectors of AD. According to the unsupervised classification, primarily, select a needed amount of cluster centers in consort with threshold values.

Subsequent to each iteration, the clusters are refined. The iteration stops when the greatest amount of iteration specified by the user is reached.

If the clusters are formed, each one of the clusters is allotted to one among the appropriate classes. This type of classification would cluster the training feature vectors of AD to a class that are identical and the class derived may not be a class of interest.

2.5. *Semi-Supervised Adaptive Neuro Fuzzy Inference System (SSANFIS) classifier*

Semi-Supervised ANFIS (SSANFIS) Learning leverages labeled as well as unlabeled training feature vectors of AD. It merges unsupervised learning that uses no training feature vectors of AD labels, and supervised learning that uses data for which labels are there.

The learning algorithm utilizes a smoothness supposition that explains that when two training feature vectors of AD are comparatively close in feature space, after that their equivalent class outputs must be close in class space. According to semi-supervised learning a classifier is repeatedly constructed on its own predictions. Primarily, a classifier is constructed on the labeled training feature vectors of AD and utilized for categorizing unlabeled training feature vectors of AD.

Characteristically the most positively anticipated samples are after that repeatedly interleaved into the training feature vectors of AD and a novel classifier is produced [18]. A classifier is developed by means of ANFIS model. In Figure 3, the block diagram of ANFIS with SSL is shown.
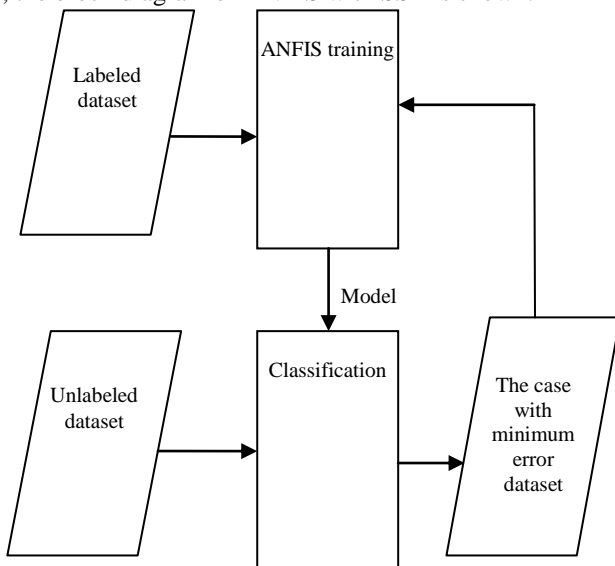

Figure 3: The diagram of SSANFIS classifier

This research presented a novel SSANFIS-based classifier. A common self-training algorithm 3 for the imbalanced training feature vectors of AD follows.

Algorithm 3: SSANFIS Classifier in AD Prediction

1. **Input:** L_{fv} – the original labeled training feature vectors of AD, U - unlabeled training feature vectors of AD ,
2. **Output:** L_M – final labeled training feature vectors of AD with the entire cases from U categorized as good or bad
3. **Begin**
4. Combine the entire good cases from L_{fv} and the similar amount of bad cases arbitrarily chosen from L_{fv} to generate novel training feature vectors of AD 'D' with balanced class labels
5. Utilize training feature vectors of AD 'D' to build an ANFIS-based predictive model (M)
6. Categorize unlabeled training feature vectors of AD U_{fv} with M
7. Calculate the error
8. Choose case u with the minimum error and direct the case from U_{fv} to the L_M.
   - 8.1. $L_M = L_{fv}+u$
   - 8.2.   U=U-u
   - 8.3.   Until U is null
9. End process

In section 2.5., ANFIS-based predictive model process is conversed that is utilized for AD prediction. On the other hand the presented SSANFIS method is diverse from the classification model, as the presented SSANFIS model takes the label as well as unlabeled samples.

III. RESULTS AND DISCUSSION

In this section, performance of proposed SSANFIS model is evaluated. The experiments were carried out on the MATLAB platform. The datasets includes T2-weighted MRI in axial plane and 256×256 resolution, which were obtained from the website of Harvard Medical School (URL: http: //med. harvard.edu/AANLIB/), OASIS dataset (URL: http:// www.oasis-brains.org/), and ADNI dataset (URL: http://adni. loni. uc- la.edu/). The proposed SSANFIS model is compared with the existing Discrete Cosine Transform (DCT) with Artificial Neural Network (ANN) and Support Vector Machine (SVM) classifier. In the classification result is based on four classes such as Alzheimer's disease, Mild Alzheimer's disease, Huntington's disease and normal. The confusion matrix results of the four classifiers such as the SVM, DCT-ANN , DWT-FNN , SHFNN and SSANFIS classifiers is discussed in table 1.

Table 1: Confusion Matrix for Alzheimer's Disease Classification Methods

Parameters	Alzheimer's disease classification methods				
	SVM	DCT-ANN	DWT-FNN	SHFNN	SSANFIS
TP	51	51	53	55	57
TN	30	32	34	36	38
FP	10	8	6	4	2
FN	9	9	7	5	3

Some preliminary results, in terms of sensitivity, specificity, and accuracy are reported in Table 2 for the MR

Classification Methods	Accuracy (%)	Error Rate (%)	Specificity (%)	Sensitivity (%)
SVM	81	19	75	85
DCT - ANN	83	17	80	85
DWT - FNN	87	13	85	88.33
SHFNN	91	9	90	91.67
SSANFIS	95	05	95.009	95

brain images.

Table 2: Overall Comparison Results (Proposed SSANFIS Model)

Figure 4 shows the step by step results of the proposed work , at initially two input images are taken , then those images are given as to preprocessing stage, then noises are removed by using IABC algorithm, finally features are selected using the MFCM clustering.
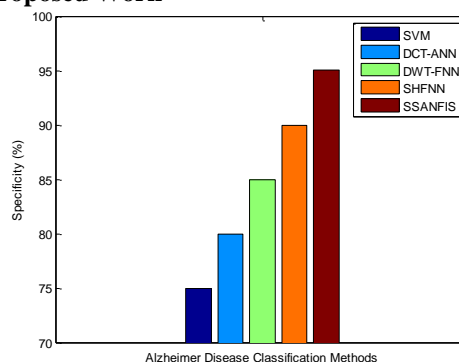

(A) Input Images

(B) Preprocessing Stage

(C) IABC based preprocessing results

(D) Feature selection results

Figure 4: Image Results of the Proposed Work

Figure 5: Accuracy Vs Methods (Proposed SSANFIS Model)

Figure 6: Specificity Vs Methods (Proposed SSANFIS Model)

Figure 5 shows the evaluation comparison of the accuracy in terms of accuracy (%). The figure 4 shows that the proposed SSANFIS has provides higher average accuracy of 95% which is 4%, 8%, 12% and 14% higher when compared to other SHFNN, DWT-FNN, DCT-ANN and SVM classifiers. Figure 6 shows the comparison results in terms of specificity metric. SSANFIS has provides higher average specificity of 95.009% which is 5.009%, 10.009%, 15.009% and 20.009% higher when compared to other SHFNN, DWT-FNN, DCT-ANN and SVM algorithms.
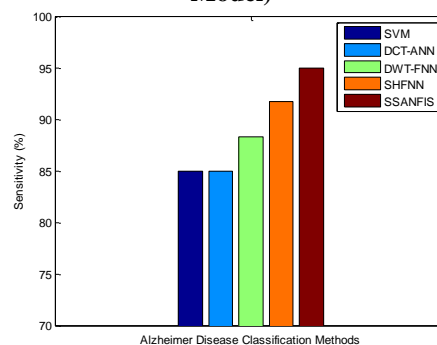

Figure7: Sensitivity Vs Methods (Proposed SSANFIS Model)

Figure 7 shows the comparison results in terms of sensitivity metric. SSANFIS has provides higher average sensitivity of 95% which is 3.33%, 6.67%, 10 %, and 10% higher when compared to other SHFNN, DWT-FNN, DCT-ANN and SVM algorithms.
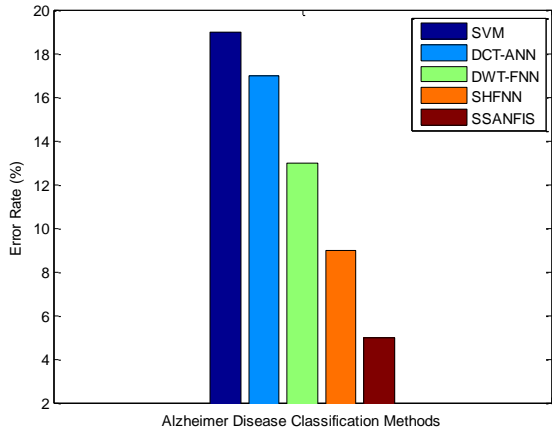

Figure 8: Error Rate Vs Methods (Proposed SSANFIS Model)

Figure 8 shows the comparison results in terms of error rate metric. SSANFIS has provides higher average error rate of 5% which is 4%, 8%, 12%, and 14% lesser when compared to other SHFNN, DWT-FNN, DCT-ANN and SVM algorithms.
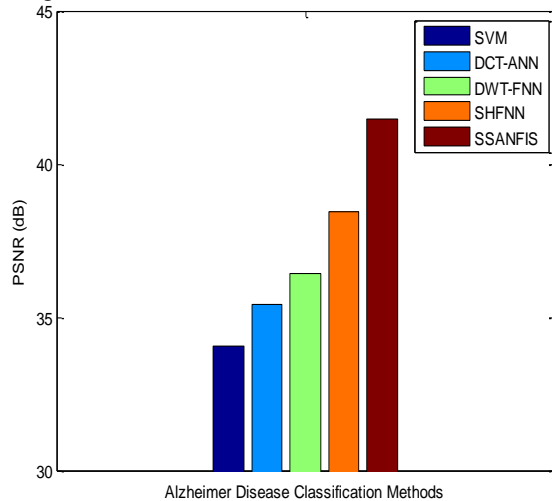

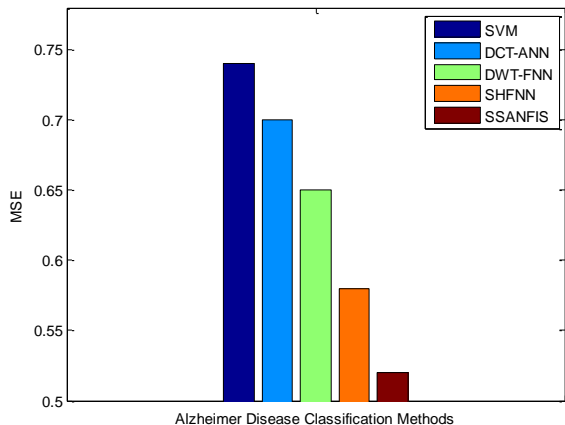**Figure 9: PSNR Vs Methods (Proposed SSANFIS Model)**

Figure 10: MSE Vs Methods (Proposed SSANFIS Model)

From Figure 9, it can be proved that the proposed SSANFIS classifier has provides higher PSNR (dB) of

41.48 dB, higher than 3.04 dB, 5.066 dB, 6.076 dB and 7.426 dB when compared to other SHFNN, DWT-FNN, DCT-ANN and SVM algorithms. From Figure 10, it can be proved that the proposed SSANFIS classifier has provides lesser MSE of 0.52 which is 0.06, 0.13, 0.18 and 0.22 when compared to other SHFNN, DWT-FNN, DCT-ANN and SVM algorithms. The table values of MSE are discussed in table3.
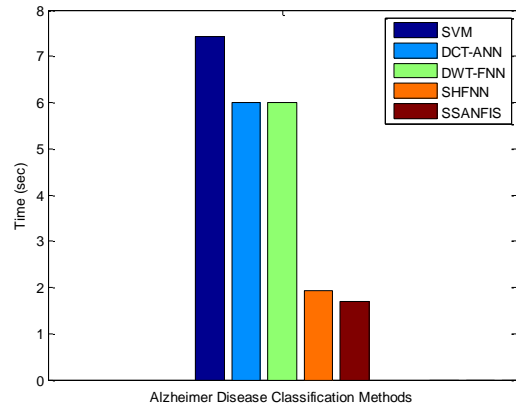

Figure 11: Classification Time Vs Methods

From Figure 11, it can be proved that the proposed SSANFIS classifier has takes lesser classification time of 1.6898 seconds which is 0.2517 seconds, 4.2991 seconds, 4.301 seconds and 5.7322 seconds lesser when compared to other SHFNN, DWT-FNN, DCT-ANN and SVM algorithms are discussed in table 3.

Table 3: Image Quality and Statistical Measures Vs Methods (Proposed SSANFIS Model)

Methods	PSNR(dB)	MSE	Time (seconds)
SVM	34.06	0.74	7.422
DCT - ANN	35.41	0.7	5.9908
DWT - FNN	36.42	0.65	5.9889
SHFNN	38.44	0.58	1.9415
SSANFIS	41.486	0.52	1.6898

IV.   CONCLUSION AND FUTURE WORK

The proposed system designed a Semi-Supervised Adaptive Neuro Fuzzy Inference System (SSANFIS) for effective classification of alzheimer's disease. In this work, Improved Artificial Bee Colony (IABC) algorithm is used for preprocessing step which provides higher classification performance. Feature extraction by HWT is used to extract the features from the MRI images.

Unsupervised Fuzzy C Means (USFCM) clustering algorithm is proposed for selecting the necessary features from the extracted features. The model is learned by using SSANFIS approach. It selects optimal fuzzy rules based on the higher frequency rules which are used to increase the accurate classification results. It provides higher results when compared to other algorithms. In the future, the various Scale Invariant Feature Transform (SIFT) images can be improved by using hybrid optimization based FS and transformation with advanced classification strategies.

REFERENCES

1. Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D. and Alzheimer's Disease Neuroimaging Initiative. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage* **65** (2013) 167-175.

2. Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C. and Snyder, P.J. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia: The journal of the Alzheimer's Association* **7** (3) (2011) 270-279.

3. Glabe, C.C. Amyloid accumulation and pathogensis of Alzheimer's disease: significance of monomeric, oligomeric and fibrillar Aβ. *In Alzheimer's Disease*, 2005, 167-177.

4. Ray, S., Britschgi, M., Herbert, C., Takeda-Uchimura, Y., Boxer, A., Blennow, K. and Kaye, J.A. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nature medicine* **13** (11) (2007) 1359-1362.

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

6. Varnum, M.M. and Ikezu, T. The classification of microglia activation phenotypes on neurodegeneration and regeneration in Alzheimer's disease brain. *Archive immunological therapies experimentalism* **60** (4) (2012) 251-266.

7. Zhou, J., Greicius, M.D., Gennatas, E.D., Growdon, M.E., Jang, J.Y., Rabinovici, G.D. and Seeley, W.W. Divergent network connectivity changes in behavioral variant front temporal dementia and Alzheimer's disease. *Brain* **133** (5) (2010) 1352-1367.

8. Liu, F., Zhou, L., Shen, C. and Yin, J. Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. *IEEE journal of biomedical and health informatics* **18** (3) (2014) 984-990.

9. Farhan, S., Fahiem, M.A. and Tauseef, H. An ensemble-of-classifiers based approach for early diagnosis of Alzheimer's disease: classification using structural features of brain images. *Computational and mathematical methods in medicine* (2014), 1-12.

10. Liu, X., Tosun, D., Weiner, M.W., Schuff, N. and Alzheimer's Disease Neuroimaging Initiative. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. *Neuro image* **83** (2013), 148-157.

11. Liu, M., Zhang, D., Shen, D. and Alzheimer's Disease Neuroimaging Initiative. Ensemble sparse classification of Alzheimer's disease. *NeuroImage* **60** (2) (2012) 1106-1116.

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

13. Kong, W., Mou, X. and Yang, B. Study DNA microarray gene expression data of Alzheimer's disease by independent component analysis. IEEE *International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing,* 2009, 44-47.

14. López, M., Ramírez, J., Górriz, J.M., Salas-Gonzalez, D., Alvarez, I., Segovia, F. and Puntonet, C.G. Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules. *Electronics Letters* **45** (8) (2009) 389-391.

15. Karaboga, D. and Basturk, B. On the performance of artificial bee colony (ABC) algorithm. *Applied soft computing* **8** (1) (2008) 687-697.

16. Wong, L.P., Loq, M.Y.H. and Chong, C.S. A Bee Colony Optimization Algorithm for Travelling Salesman Problem. *Second Asia International Conference on Modelling & Simulation,* Kuala, Lumpur, 2008, 818- 823.

17. Li, C., Liu, L. and Jiang, W. Objective Function of Semi-Supervised Fuzzy C-Means Clustering Algorithm. *IEEE International Conference on Industrial Informatics*, Daejeon, Korea, 2008, 737-742.

18. Chapelle, O., Scholkopf, B. and Zien, A. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, (2006).