Alzheimer's Disease Detection using Correlation based Ensemble Feature Selection and Multi Support Vector Machine

Jhansi Rani Kaka[1*] & K. Satya Prasad[2]

[1] Department of Electronics and Communication Engineering,
University college of Engineering Kakinada, JNTUK, Kainada, India
[2]Department of Electronics and Communication Engineering,
Vignan's Foundation for Science, Technology & Research (VFSTR), Guntur, India

E-mail address:[1]jhansikaka@gmail.com, [2]prasad_kodati@yahoo.co.in

Abstract: In recent decades, machine learning techniques have been playing a crucial role in the field of computer aided diagnosis. This paper address the issue of automated Alzheimer's disease detection on the basis of magnetic resonance imagining, and proposed a new supervised machine learning technique for Alzheimer's disease diagnosis. Initially, an adaptive histogram equalization and region growing are employed on the collected brain scans for contrast improvement and skull removal. Next, Fuzzy C Means (FCM) clustering algorithm is applied in the enhanced brain scans to segment tissues like White Matter (WM), Cerebro Spinal Fluid (CSF), and Grey Matter (GM). Ina addition, feature extraction is accomplished in the segmented brain tissues using Gabor and local directional pattern variance features. In order to decrease the dimension of the extracted feature vectors, the correlation based on ensemble feature selection algorithm was proposed. Finally, the obtained optimal feature vectors are fed to Multi Support Vector Machine (MSVM) to classify Mild Cognitive Impairment (MCI), Alzheimer's disease, and healthy controls classes. From the simulation outcome, the proposed ensemble feature selection with multi support vector machine model shows 9.58% and 5.09% improvement in classification accuracy on Open Access Series of Imaging Studies (OASIS) and Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets compared to the existing models.

Keywords: Adaptive Histogram Equalization; Alzheimer's Disease; Ensemble Feature Selection; Fuzzy C Means Clustering; Multi Support Vector Machine.

1. INTRODUCTION

Alzheimer's disease (AD) is a neuro-degenerative disorder, which is generally characterized as dementia [1-2]. In AD, the brain cells are destroyed, which causes thinking and memory losses, and finally leads to death. AD approximately affects over 22 million people worldwide, so the recognition of AD at an early phase is essential to reduce mortality rate [3-4]. In recent times, many imaging techniques are available for AD detection such as X ray, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), functional MRI (FMRI), Electro-Encephalography (EEG) and histopathology [5-7]. In MRI imaging, various multi-slices images are obtained that assist the clinicians in accurate detection of AD [8-10]. In recent decades, numerous automated systems are developed by the researchers for AD detection on the basis of machine learning techniques, which comprises of both supervised and unsupervised techniques [11-13]. By surveying the existing research works, it was found that researchers faced problems like more time consumption to recognize patterns from the brain images and lack of human intervention to interpret the data, especially in case of an enormous dataset size. In order to resolve the above stated problems, a new supervised model is proposed in this research paper to enhance the performance of AD detection. The major contributions of the proposed model are listed below;

- Several imaging techniques are available for AD detection, in which MRI is the standardized imaging modality used in clinical practice.
  - In this research, MRI brain scans are collected from ADNI and OASIS datasets.
  - Image pre-processing is accomplished by using adaptive histogram equalization and region growing techniques for contrast improvement and skull removal.
  - Brain tissues like WM, CSF, and GM are segmented by FCM clustering algorithm. It is robust to noise and outliers, while retaining computational simplicity.
  - The Feature extraction is performed by Local Directional Pattern Variance (LDPV) and Gabor feature descriptors to extract the feature vectors from the segmented tissues.
  - Ensemble feature selection is developed to optimize or diminish the dimension of extracted features for accurate classification.

E-mail:author's email

- By utilizing the obtained optimal feature vectors, MSVM classifies AD, healthy controls patients, and MCI.
- Various performance measures; Positive Predictive Value (PPV), Negative Predictive Value (NPV), Fowlkes-Mallows (FM) index, accuracy, sensitivity and specificity are utilized to validate the proposed model's performance.

This paper is prepared as follows; In Section 2, a few recent research papers on the topic "AD detection" are surveyed. The detailed explanation about the proposed model is given in the Section 3. The experimental analysis of the proposed model is represented in the Section 4. The conclusion of this work is indicated in the Section 5.

2. RELATED WORKS

M. Khajehnejad et al. [14] developed a semi-supervised manifold learning system for classifying MRI brain scans into two classes named as normal condition and MCI. Initially, a voxel morphometry analysis was carried-out to extract important AD features from the brain scans. Then, the dimensional reduction was carried out on the extracted features by using Principal Component Analysis (PCA) to achieve precise and faster classification. Finally, the optimized AD features were fed into label propagation methodology for classifying the testing brain scans into two classes; normal condition and MCI. The simulation result showed that the developed system obtained effective performance on OASIS dataset by means of specificity, error rate, accuracy, and sensitivity. Unlike other machine and deep learning methods, the label propagation methodology consumes more memory space for experimental analysis. U.R. Acharya et al. [15] implemented a new system to investigate the severity of brain abnormalities caused by AD. In this literature, the brain scans were collected from T2 weighted brain MRI dataset. Initially, median filtering technique was used to improve the quality of collected brain scans by removing noises. Then, the feature extraction was accomplished by complex wavelet transform, contourlet transform, dual tree complex wavelet transform, empirical wavelet transform, curvelet transform, discrete wavelet transforms and shearlet transform to extract AD related features from the pre-processed images. Further, optimal subset of feature vectors was selected from the extracted features using students test. Finally, K Nearest Neighbor (KNN) was applied to categorize the brain scans into normal and AD categories. The classification phase might be slow with larger datasets, and the classification accuracy completely depends on the data quality. These were the two major concerns related to KNN classifier.

X. Bi et al. [16] developed a two phase model for early detection of AD from MRI scans. In the initial phase, Convolutional Neural Network (CNN) was developed on the basis of PCA Net for extracting the features from the collected brain scans. In the second phase, k-means clustering was employed for classifying the brain scans as AD, normal condition and MCI. In this experimental section, the developed dual phase model performance was validated on ADNI dataset by means of classification accuracy. Additionally, S. Basaia et al. [17] developed a deep learning algorithm to classify the MRI brain scans as normal, AD and MCI patients. In this literature, CNN model was applied on T1 weighted images, which were collected from ADNI dataset. In the practical applications, CNN model was computationally expensive, since it needs a good Graphic Processing Unit (GPU) based system in order to achieve precise performance in AD detection. J. Samper-Gonzalez et al. [18] presented a novel system for early detection of AD using MRI and PET scans. In this literature study, the brain scans were collected from OASIS, and ADNI datasets. After image collection, the PET-partial volume correction software's and statistical parametric mapping were used for image denoising, and then the voxel and region feature vectors were extracted from the denoised images. Finally, image classification was accomplished by using RF, logistic regression and SVM for classifying the images as normal condition, AD, and MCI. The extensive experiment shows that the developed system obtained effective performance in AD detection in light of accuracy. The experimental segment validated that the developed system was suitable only for single modality classification problem, which proved to be a major issue in this literature study.

I. Beheshti et al. [19] developed a new automated system to detect AD on the basis of feature ranking approaches and classification errors. Initially, the brain scans were acquired from ADNI dataset, and then feature extraction was accomplished by using Voxel Based Morphometry (VBM) to extract raw feature vectors and voxel values from the collected brain images. Then, feature ranking was performed on the extracted feature vectors by using seven methods that are information gain, mutual information, Gini index, statistical dependency, fisher's criterion, t-test score and Pearson's correlation co-efficient. Hence, feature vectors with higher scores were fed to the SVM classifier for classifying the patients as MCI, AD, and normal. From the experimental investigation, the developed system obtained effective performance in AD detection in light of accuracy, specificity, sensitivity and Area under Curve (AUC). R.S. Kamathe and K.R. Joshi, [20] used band expansion processes, thresholding, independent component analysis, and skull removal to segment WM, CSF, and GM from the collected brain scans. Then, Grey Level Co-occurrence Matrix (GLCM) and SVM were applied for extracting the feature vectors and classifying the images as AD, normal conditions and MCI. The extensive experiment showed that the developed system achieved significant performance in AD detection in light of recall, similarity index, precision, accuracy and tanimoto index. However, SVM was a binary classification technique, which is suitable for two-class classification and is inappropriate for multi-class classification.

K. Shankar et al. [21] presented a new model for AD detection by using MRI scans. In this study, the brain scans were acquired from ADNI datasets. Initially, histogram, texture and scale invariant transform features were extracted from the collected brain scans. After extracting the feature vectors, a Group Grey Wolf Optimization (GGWO) algorithm was introduced to select the optimal feature sub-sets. Finally, classification was carried-out by

using CNN, KNN and decision tree to classify AD, MCI and healthy controls. Among CNN, KNN and decision tree, CNN with GGWO model obtained a maximum accuracy of 96.2%, which was better when compared to the traditional methods. As a future enhancement, it was important to choose the relevant feature vectors using an improved algorithm to tackle the concern of overfitting. I. Beheshti et al. [22] introduced a new technique on the basis of VBM and Probability Distribution Function (PDF) to detect ADs. After collecting the images from ADNI dataset, statistical feature extraction was accomplished to extract the feature vectors. Then, PDF methodology was used to compress the statistical information from higher dimensional feature vectors into lower dimensional feature vectors. Further, the optimized feature vectors were fed into SVM classifier to classify the images as AD and healthy controls. The Simulation outcome showed that the developed technique obtained effective performance in AD detection in light of sensitivity, accuracy, AUC and specificity. The developed PDF-SVM technique includes many outliers which results in misclassification. A. Savio et al. [23] developed VBM approach to determine the differences in brain tissues by voxel wise comparison of brain scans. The standard deviation and mean values were extracted from each voxel location cluster. The extracted feature values were fed to Radial Basis Function (RBF) networks, SVM, and multi-layer perceptron trained with back propagation to classify AD and MCI patients. In this literature study, the combined RBF-diverse Adaboost-SVM classifier obtained a maximum classification accuracy of 86% on OASIS dataset. From the experimental validation, it was essential to resolve the multi-variate interpolation problem in RBF networks in order to achieve better accuracy in AD detection. To highlight the above stated concerns, a new model; ensemble feature selection-MSVM is proposed to enhance the segmentation and classification performance of AD.

3. METHODOLOGY

The proposed AD detection model comprises of six phases; data collection: ADNI and OASIS datasets, pre-processing: adaptive histogram equalization and region growing, image segmentation: FCM clustering algorithm, feature extraction: Gabor and LDPV feature descriptors, feature selection: ensemble feature selection, and classification: MSVM classifier. The Workflow of the proposed ensemble feature selection-MSVM model is represented in figure 1.

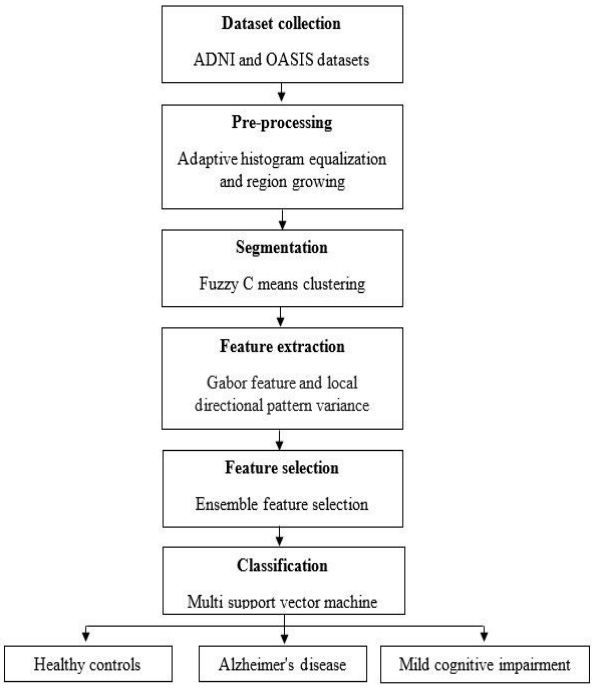
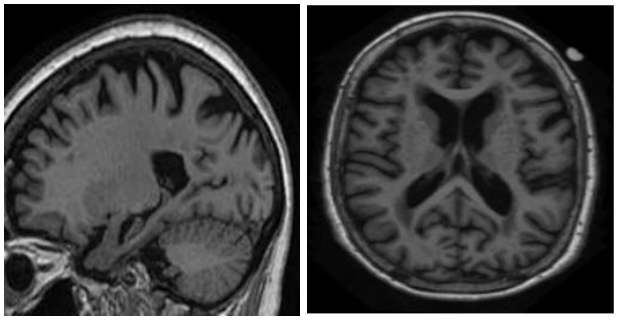


Figure 1. Work flow of proposed ensemble feature selection-MSVM model

A. Image collection and pre-processing

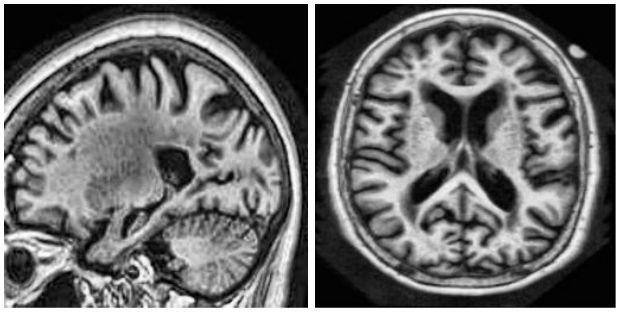
In order to perform experimental analysis, the input brain images are collected from two datasets namely: ADNI and OASIS. The ADNI dataset comprises of 3T MRI and 1.5T MRI brain images of 819 subjects which comprises of 398 subjects with MCI, 192 subjects with AD and the remaining 229 subjects are healthy controls. http://adni.loni.usc.edu/. By using standard and functional cognitive measures, the brain images are recorded for one year in ADNI dataset [24]. Additionally, OASIS-1 dataset contains 1.5T MRI brain images which are recorded from 416 subjects, where the individual's age ranges from 18 to 96. https://www.oasis-brains.org/. The OASIS dataset contains information like subject's age, clinical dementia rating, subject's education, gender, number of patients, socio-economic status, and Mini-Mental State Exam (MMSE) score [25]. In OASIS dataset, the image acquisition details are given as follows: sequence: MP-RAGE, echo time: 4 msec, slice number: 128, repetition time: 9.7 msec, orientation: axial, coronal and sagittal, flip angle: 10, resolution pixel: 256×256 $(1 \times 1\ mm)$ and Thickness, gap (mm): 1.25, 0. Sample collected images of ADNI and OASIS datasets are represented in figure 2.



(a) (b)

Figure 2. Sample collected images, (a) ADNI dataset, and b) OASIS dataset

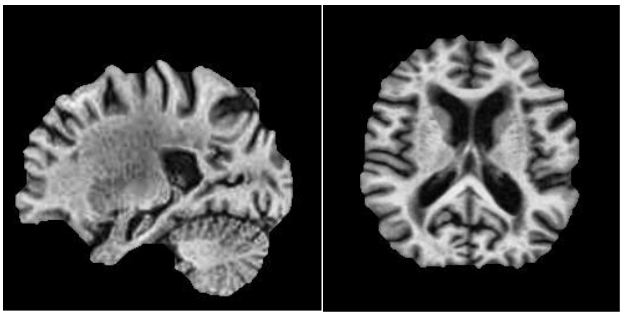
After collecting the brain images, adaptive histogram equalization approach is used to improve the contrast of the images by redistributing the lightness value [26-27].



(a) (b)

Figure 3. Output of adaptive histogram equalization, a) ADNI dataset, and b) OASIS dataset

In existing method, adaptive histogram equalization approach effectively enhances the edges and local contrast in each region of a brain image. Further, the enhanced brain image is given as input to region growing for skull removal [28-29]. The Region growing approach identifies the neighbor pixels of seed points and determines whether the pixels need to be added at the selected regions or not. The skull regions are precisely eliminated from the enhanced brain scans after the 200th iteration. The enhanced brain images are represented in figures 3 and 4.



(a) (b)

Figure 4. Output of region growing, a) ADNI dataset, and b) OASIS dataset

B. Image segmentation

After image pre-processing, segmentation is accomplished by using FCM clustering algorithm in order to segment WM, GM and CSF regions of the brain scan images. Generally, FCM clustering algorithm is used to localize the objects like WM, GM and CSF in the complex templates. Fuzzy set theory is adopted in FCM clustering algorithm to assign data objects to the clusters. In this algorithm, each object is considered as a member of each cluster with a variable degree of membership. The Euclidean distance is used in FCM clustering algorithm to estimate the similarity between objects that helps to find the correct clusters [30]. In FCM clustering algorithm, the objective function J needs to be reduced in each iteration, which is mathematically defined in equation (1).

$$J = \sum_{i=1}^{n} \sum_{j=1}^{C} \delta_{ij} \left\| x_i - c_j \right\|^2 \quad (1)$$

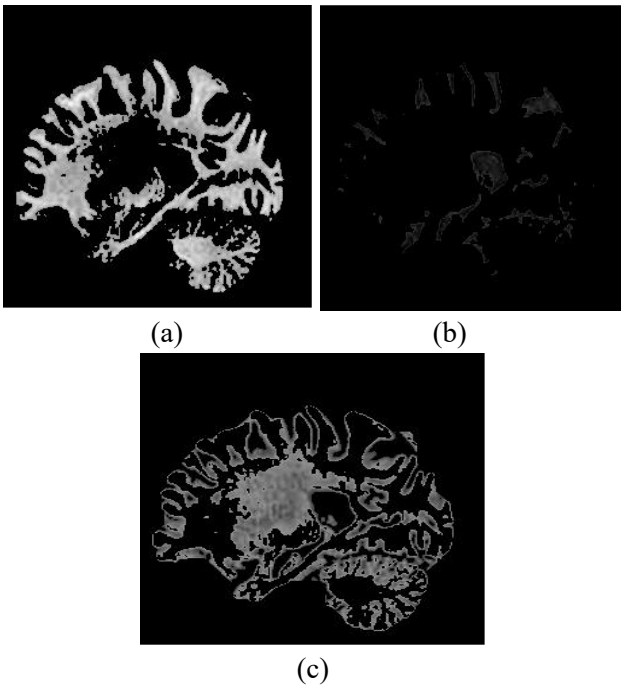
Where, n is indicated as data points, C is denoted as clusters, c_j is indicated as center vector of the cluster j and δ_{ij} is represented as degree of membership for i^{th} data point x_i in cluster j. The term $\left\| x_i - c_j \right\|$ determines the similarities between the data points x_i in cluster j. For each data point x_i, δ_{ij} is estimated by equation (2).

$$\delta_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

Where, fuzziness coefficient is represented as m. Meanwhile, center vector c_j is estimated by using equation (3).

$$c_j = \frac{\sum_{i=1}^{n} \delta_{ij}^{m} \cdot x_i}{\sum_{i=1}^{n} \delta_{ij}^{m}} \quad (3)$$

The fuzziness coefficient m is used to calculate the clustering tolerance, where the maximum value of fuzziness coefficient m states higher overlap between the clusters j. The maximum value of m uses more data points x_i in cluster j, so δ_{ij} is either zero or one.



(a) (b)

(c)

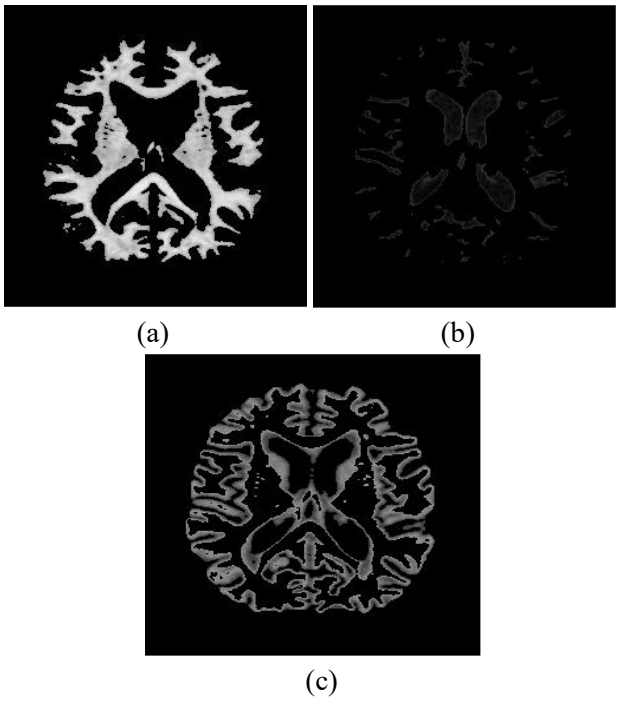
Figure 5. Segmented regions in ADNI dataset; a) WM, b) CSF, and c) GM

The degree of membership δ_{ij} finds the number of iterations which is accomplished by FCM clustering algorithm. In this scenario, the accuracy a is determined by utilizing δ_{ij} from one iteration k to the succeeding iteration $k+1$, which is mathematically represented in equation (4).

$$a = \Delta_i^n \Delta_i^c \left| \delta_{ij}^{k+1} - \delta_{ij}^k \right| \quad (4)$$

Where, δ_{ij}^{k+1} and δ_{ij}^k are indicated as δ_{ij} of iterations $k+1$ and k respectively. The segmented GM, WM and CSF regions of the enhanced images are represented in the figures 5 and 6.

Then, feature extraction is carried-out by Gabor and LDPV feature descriptors to extract the feature vectors from the segmented regions.



(a) (b)

(c)

Figure 6. Segmented regions in OASIS dataset; a) WM, b) CSF, and c) GM

C. *Feature extraction and selection*

In feature extraction phase, Gabor and LDPV feature descriptors are applied for extracting the features from the regions like WM, GM, and CSF. The Gabor features are calculated by the convolution process of segmented image I with Gabor filter bank ψ [31], as denoted in equation (5).

$$G_{u,v}(X,Y) = I(X,Y) \times \psi(X,Y) \quad (5)$$

Further, $G_{u,v}(X,Y)$ are convolved on the basis of filtering operation with the orientation v and size u, where the convolutional procedure is accomplished for both imaginary and real part of images. The general Gabor feature representation is mathematically defined in the equations (6-8).

$$O(X,Y)_{M,N} = ((\left(Re\left(O(X,Y)\right)_{M,N}\right)^2 + (Im\left(O(X,Y)\right)_{M,N})^2)^{1/2} \quad (6)$$

Where,

$$Re\left(O(X,Y)\right)_{M,N} = I(X,Y) \times Re(\psi(X,Y,\lambda_{M,}\theta_N)) \quad (7)$$

$$Im\left(O(X,Y)\right)_{M,N} = I(X,Y) \times Im(\psi(X,Y,\lambda_{M,}\theta_N)) \quad (8)$$

The extracted Gabor $O(X,Y)_{M,N}$ features from the segmented images have high redundant features and are multi-dimensional in nature. Additionally, LDPV is used for extracting the features from the segmented brain regions, where LDPV encodes contrast information and direction pattern of the brain images based on local derivative variation. By utilizing the concept of local directional pattern, LDPV provides information about nature of textures in the brain images [32-33]. The General representation of LDPV is given in equation (9).

$$LDPV = \sum_{r=1}^{M} \sum_{e=1}^{N} w(LDP(r,e),\tau) \quad (9)$$

Where, local directional pattern is denoted as LDP, M and N are indicated as size of the segmented images, and τ is stated as LDP code value. Finally, the extracted features $F = O(X,Y)_{M,N} + LDPV$ are given as input to feature selection algorithm.

After extracting the features, the correlation based on ensemble feature selection is accomplished to reduce the dimension of the extracted feature vectors. In this scenario, ensemble feature selector finds the correlation values of every feature vector selected from grasshopper algorithm [34], Particle Swarm Optimization (PSO) algorithm [35], and Genetic algorithm [36]. In each feature subset, higher similarity feature vectors are eliminated and the residual feature vectors are given as the input to ensemble feature selector to select the optimal feature values. Based on majority voting, the ensemble feature selector selects the optimal features. Step by step process of correlation based feature selector is given below,

Step 1: Arithmetic mode of the features selected by three bio-inspired optimization algorithms is calculated using equation (10).

$$Out_{ensemble\ feature\ selection} = mode \{Out_{grasshopper}, Out_{Particle\ swarm,} Out_{Genetic}\} \quad (10)$$

Step 2: Then, the correlation coefficient matrix is determined for the feature vectors in the output of $Out_{ensemble\ feature\ selection}$ using equation (11). Where, F is represented as total features, p and q are denoted as feature vectors under consideration.

$$Correlation\ coefficient = \frac{F\sum pq - (\sum p)(\sum q)}{\sqrt{[F\sum p^2 - (\sum p)^2][F\sum q^2 - (\sum q)^2]}} \quad (11)$$

Step 3: If the correlation value is higher than 0.95, p and q are highly correlated and eliminated. Or else, p and q are selected by the correlation based on ensemble feature selector, which is given as the input to MSVM classifier. On both datasets, the extracted features' size F is 450×34267, and the size of the selected features p and q is 450×22497.

D. Classification

The obtained feature vectors are fed to MSVM in order to classify AD, MCI and healthy controls. Usually, SVM is a binary classifier, which is appropriate for two class classification: either AD vs MCI or MCI vs healthy controls or AD vs healthy controls. To deal with multi-class classification; AD vs MCI vs healthy control it is necessary to create a multi-SVM classifier with hierarchical structure. One against All (OAA) and One against One (OAO) are the most frequently used approaches that decomposes the i^{th} class problems into a set of binary classification problems, and then combines all i^{th} binary classifiers. The OAO approach creates $i_2 \times (i_2 - 1)/2$ classifiers that discriminate class I and class II. Further, OAA approach creates i_1 binary SVM classifiers with distinct one class from all the residual classes [37]. In OAA approach, i^{th} SVM classifier trains the training image sets of i^{th} class with both positive and negative labels. In MSVM classification technique, OAO and OAA approaches are integrated to create $i = i_1 + i_2 \times (i_2 - 1)/2$ class problems for classifying three classes on both ADNI and OASIS datasets. Mathematical expressions of MSVM classifier are indicated in the equations (12), (13) and (14).

$$min\Phi(w_z, \xi) = 1/2 \sum_{z=1}^{class}(w_z) + constant \sum_{i=1}^{l} \sum_{z \neq yi} \xi_i^z \quad (12)$$

Subjected to,

$$\begin{pmatrix} w_{yi} \times \\ (p,q)_i \end{pmatrix} + B_{yi} \geq \left(w_{yi} \times (p,q)_i\right) + B_z + 2 - \xi_i^z, \quad (13)$$

$$\xi_i^z \geq 0, i = 1,2,3 \dots l, z, yi \in \{1,2,3 \dots i\}, z \neq yi \quad (14)$$

Where, yi is denoted as class of training data vectors $(p,q)_i$, l is represented as training data point, ξ_i^z is stated as slack variable, B is represented as thresholds in the new space, and w_z is denoted as a sum of norms, which increases the margin among three classes; AD, MCI and healthy controls.

4. Experimental Results

In this research, the proposed ensemble feature selection-MSVM model is simulated by using MATLAB (2018a) software tool with the system requirements; operating system: Windows 10, RAM: 16 GB, and processer: Intel Core i7. In this study, the proposed ensemble feature selection-MSVM model is compared with some benchmark existing models like Semi-supervised manifold learning system [14], Feature ranking and Classification errors [19], CNN-GGWO [21] and RBF-diverse Adaboost-SVM [23] on OASIS and ADNI datasets to investigate the proposed model's effectiveness. In this research, the proposed model performance is analyzed in light of accuracy, PPV, sensitivity, NPV, specificity and FM. The mathematical expressions of the performance measures are represented in the equations (15-20).

Accuracy is one of the important performance measures in AD detection, which represents how close the obtained result to the true value. Sensitivity is a test that accurately identifies the features with the AD disease, and specificity is a test, which accurately identifies the features without the AD disease. The mathematically expressions of accuracy, sensitivity and specificity are represented in the equations (15-17).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (15)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (16)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (17)$$

In the diagnostic and statistics tests, PPV and NPV are the proportions of positive and negative results which are called as true positive and true negative results.

Fowlkes-mallows index is utilized to identify the similarity between two clusters, and also it's a metric to estimate confusion matrices. The mathematical expressions of PPV, NPV, and Fowlkes-mallows are represented in the equations (18-20).

$$PPV = \frac{TP}{TP+FP} \times 100 \quad (18)$$

$$NPV = \frac{TN}{TN+FN} \times 100 \quad (19)$$

$$FM = \sqrt{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}} \times 100 \quad (20)$$

Where, TP denotes true positive, TN states true negative, FP indicates false positive, and FN denotes false negative.

A. Quantitative investigation on ADNI dataset

In this section, the proposed ensemble feature selection-MSVM model performance is validated on ADNI dataset. In this research, k-fold cross validation is applied to train and test the collected brain scans. In this scenario, the quantitative analysis is done for 450 brain scans (150 AD class, 150 MCI class and 150 healthy control class) with 20% testing and 80% training of brain scans. As denoted in table 1, the proposed ensemble feature selection-MSVM model performance is analyzed in terms of FM index, PPV, sensitivity, NPV, specificity and accuracy.

By inspecting table 1, the performance analysis is carried out with different feature selection and classification techniques like grasshopper algorithm, PSO algorithm, genetic algorithm, Random Forest (RF), Artificial Neural Network (ANN), KNN, Decision Tree (DT) and MSVM. These techniques are investigated with different combinations in that ensemble feature selection with MSVM classifier achieved effective performance in AD detection. The combination of ensemble feature selection with MSVM classifier obtained maximum classification accuracy of 97.57%, sensitivity of 95.6%, specificity of 97.76%, PPV of 92.99%, NPV of 95.90% and FM index value of 98.45%. The graphical depiction of the proposed ensemble feature selection-MSVM model on ADNI dataset in terms of PPV, sensitivity, FM index, NPV, specificity and accuracy is denoted in the figures 7 and 8.

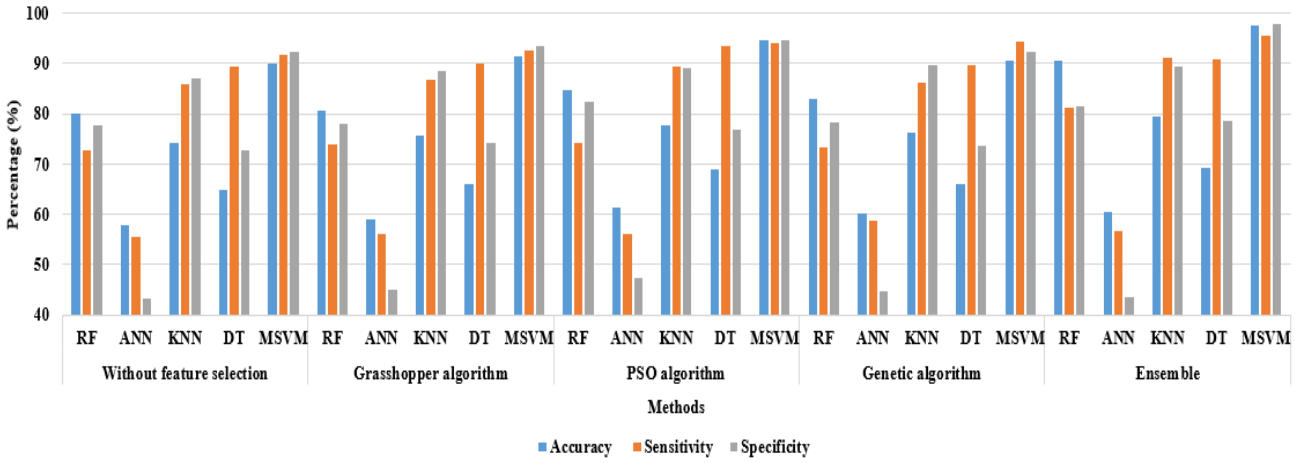


Figure 7. Graphical depiction of proposed ensemble feature selection-MSVM model on ADNI dataset by means of specificity, sensitivity, and accuracy

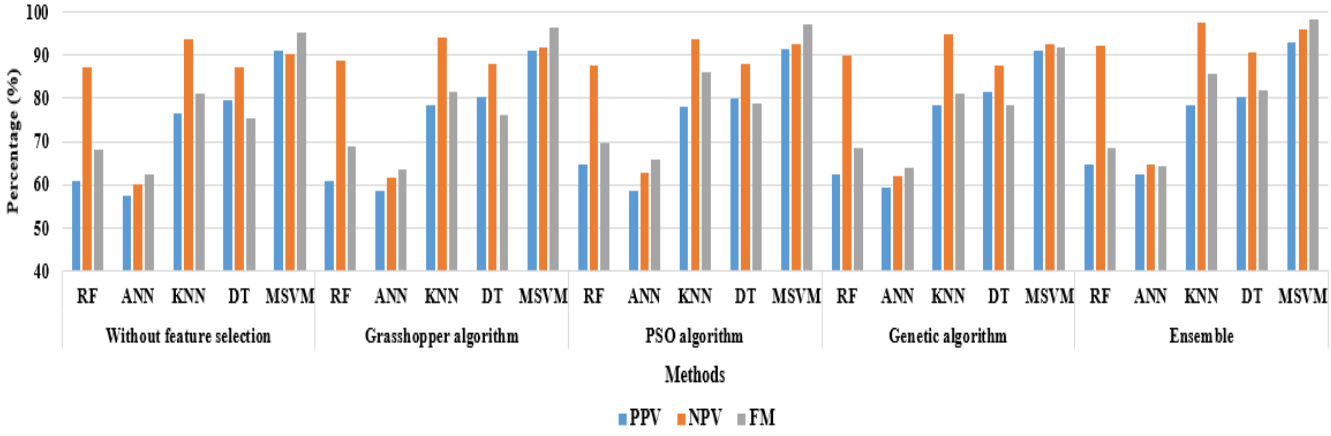


Figure 8. Graphical depiction of proposed ensemble feature selection-MSVM model on ADNI dataset by means of PPV, NPV, and FM index

TABLE I. PERFORMANCE ANALYSIS OF THE PROPOSED MODEL ON ADNI DATASET

ADNI dataset							
Feature selection	***Classifier***	***Accuracy (%)***	***Sensitivity (%)***	***Specificity (%)***	***PPV (%)***	***NPV (%)***	***FM (%)***
Without feature selection	RF	80.10	72.71	77.71	60.78	87.39	68.08
	ANN	57.81	55.52	43.12	57.24	60.24	62.56
	KNN	74.19	86	87.14	76.57	93.62	81.27
	DT	64.95	89.29	72.86	79.51	87.35	75.43
	MSVM	90.10	91.71	92.23	91.08	90.48	95.35
Grasshopper algorithm	RF	80.75	74.01	77.87	60.88	88.67	69
	ANN	58.97	56.03	44.87	58.44	61.77	63.46
	KNN	75.63	86.86	88.51	78.48	94.01	81.33
	DT	65.99	90.06	74.06	80.20	88.10	76.22
	MSVM	91.41	92.68	93.43	91.22	91.81	96.61
PSO algorithm	RF	84.67	74.07	82.27	64.85	87.76	69.84
	ANN	61.24	56.18	47.38	58.49	62.58	65.67
	KNN	77.64	89.48	89.07	78.25	93.84	86
	DT	68.98	93.43	76.88	79.91	88.04	79.01
	MSVM	94.52	93.97	94.68	91.28	92.64	97.25
Genetic algorithm	RF	82.90	73.33	78.25	62.51	90.07	68.55
	ANN	60.07	58.67	44.82	59.46	62.18	63.88
	KNN	76.31	86.24	89.79	78.37	95.02	81.28
	DT	66	89.73	73.48	81.68	87.68	78.36
	MSVM	90.68	94.25	92.37	91.16	92.73	91.73
Ensemble	RF	90.69	81.27	81.46	64.71	92.31	68.53
	ANN	60.45	56.61	43.53	62.22	64.84	64.18

E-mail:author's email

	KNN	79.49	91.27	89.42	78.47	97.51	85.75
	DT	69.17	90.76	78.47	80.38	90.68	81.92
	MSVM	97.57	95.65	97.76	92.99	95.90	98.45

B. Quantitative investigation on OASIS dataset

In this section, OASIS dataset is utilized to investigate the performance of the proposed model by means of accuracy, PPV, sensitivity, NPV, specificity and FM index. In this scenario, the proposed ensemble feature selection-MSVM model is validated for all three slices in OASIS datasets that are Axial, Coronal and Sagittal, which are denoted in the tables 2, 3, and 4. By inspecting the tables 2, 3 and 4, the combination ensemble feature selection with MSVM classifier shows better performance in AD detection compared to other feature selection and classification techniques like grasshopper algorithm, PSO algorithm, genetic algorithm, ANN, RF, KNN and DT. MSVM model averagely achieved 94.58% of classification accuracy, 94.7% of sensitivity, 96.26% of specificity, 96.69% of PPV, 95.35% of NPV, and 95.43% of FM index, which are better compared to other combinations. Feature selection is an important phase in this research, where the proposed model significantly reduces the "curse of dimensionality" problem and improves the AD detection performance by decreasing the dimensions of the extracted feature vectors. MSVM classifier effectively reduces the size of resulting dual problem by creating a classification error bound, and speeds up the training procedure by maintaining a competitive classification accuracy.

TABLE II. PERFORMANCE ANALYSIS OF THE PROPOSED MODEL ON OASIS DATASET (AXIAL SLICE)

OASIS dataset (Axial slice)							
Feature selection	***Classifier***	***Accuracy (%)***	***Sensitivity (%)***	***Specificity (%)***	***PPV (%)***	***NPV (%)***	***FM (%)***
Without feature selection	RF	86.40	84.69	84.53	80.51	85.54	70.36
	ANN	56.25	55.04	41.86	56.62	59.36	60.83
	KNN	82.71	84.99	86.32	75.97	92.76	80.52
	DT	73.25	88.26	72.51	78.45	86.69	74.28
	MSVM	86.70	91.51	90.36	90.93	89.28	94.17
Grasshopper algorithm	RF	86.35	85.35	84.56	80.37	86.41	72.06
	ANN	57.37	54.73	42.58	56.55	60.04	62.50
	KNN	84.15	85.30	86.82	76.04	93.43	80.94
	DT	74.31	89.07	72.11	79.42	86.86	74.63
	MSVM	87.72	90.98	91.80	90.28	89.90	94.63
PSO algorithm	RF	88.53	87.44	88.09	82.23	88.86	74.46
	ANN	58.36	58.83	45.63	60.30	63.30	63.02
	KNN	86.33	87.54	88.66	78.58	93.85	81.37
	DT	75.98	91.33	75.05	82.26	89.76	75.99
	MSVM	89	91.85	93.32	91.13	90.59	95.21
Genetic algorithm	RF	87.89	89.75	87.43	84.23	94.34	76.30
	ANN	64.66	57.32	49.35	62.43	67.13	62.72
	KNN	87.48	86.80	87.47	81.82	98.98	86.78
	DT	79.55	95.26	75.23	82.25	89.17	80.58
	MSVM	92.23	92.76	94.26	97.48	90.60	93.02
Ensemble	RF	91.09	94.59	88.93	89.67	91.83	72.35
	ANN	68.34	61.66	44.86	65.60	62.02	73.25
	KNN	85.28	91.93	88.32	88.12	94.36	91.87
	DT	80.37	90.14	80.16	80.81	87.62	85.61
	MSVM	94.03	95.15	95.58	96.89	97.68	96.06

TABLE III. PERFORMANCE ANALYSIS OF THE PROPOSED MODEL ON OASIS DATASET (SAGITTAL SLICE)

OASIS dataset (Sagittal slice)							
Feature selection	***Classifier***	***Accuracy (%)***	***Sensitivity (%)***	***Specificity (%)***	***PPV (%)***	***NPV (%)***	***FM (%)***
Without feature selection	RF	85.47	84.38	83.90	79.89	85.23	69.74
	ANN	55.67	54.55	41.08	56.57	59.35	59.87
	KNN	82.60	84.96	85.80	75.33	92.19	79.55
	DT	72.70	87.65	72.31	77.87	86.05	73.32
	MSVM	85.73	91.34	89.91	90.14	89.05	84.04
Grasshopper algorithm	RF	85.57	85.14	83.94	80.17	85.58	71.44
	ANN	57.24	54.62	42.55	55.73	59.90	61.79
	KNN	83.82	84.45	86.43	75.15	93.19	80.36
	DT	74.16	88.94	71.49	79.40	86.03	74.09
	MSVM	87.31	90.19	90.93	89.29	89.45	93.87
PSO algorithm	RF	88.52	86.72	87.36	81.86	88.74	74.46
	ANN	58.02	58.54	45.60	59.66	63.07	62.79
	KNN	85.41	87.03	88.23	78.21	93.24	80.80
	DT	75.03	90.57	74.60	81.44	88.90	75.60
	MSVM	88.09	91.60	92.80	90.74	90.27	94.84
Genetic algorithm	RF	87.24	89.71	86.98	83.99	94.02	75.50

	ANN	64.22	56.75	48.39	62.35	66.72	61.93
	KNN	87.26	86.19	86.84	81.38	98.87	86.58
	DT	79.10	95.20	74.39	81.46	88.49	80.54
	MSVM	92.18	92.21	93.27	96.76	89.72	92.19
Ensemble	RF	90.74	94.46	88.44	88.89	90.90	72.33
	ANN	68.30	61.64	44.49	64.78	61.42	72.41
	KNN	85.12	91.01	88.13	87.44	93.49	91.42
	DT	79.50	89.34	79.95	79.85	86.78	85.04
	MSVM	93.67	94.32	94.74	96.10	97.34	95.81

TABLE IV. PERFORMANCE ANALYSIS OF THE PROPOSED MODEL ON OASIS DATASET (CORONAL SLICE)

OASIS dataset (Coronal slice)							
Feature selection	***Classifier***	***Accuracy (%)***	***Sensitivity (%)***	***Specificity (%)***	***PPV (%)***	***NPV (%)***	***FM (%)***
Without feature selection	RF	77.49	82.79	83.09	71.60	84.42	61.97
	ANN	50.89	49.14	39.15	53.55	58.57	51.08
	KNN	75.11	84.58	82.41	73.18	84.70	73.39
	DT	65.72	81.07	65.98	73.24	82.65	66.07
	MSVM	85.31	90.77	85.30	86.31	86.20	79.87
Grasshopper algorithm	RF	80.07	78.99	77.06	78.33	83.82	68.33
	ANN	50.43	52.27	40.19	53.52	55.81	55.36
	KNN	77.03	83.98	84.80	70.14	91.35	78.52
	DT	68.75	83.50	69.96	76.71	85.44	70.39
	MSVM	84.10	90.39	88.07	89.80	83.87	79.03
PSO algorithm	RF	87.09	86.66	93.23	86.98	91.35	71.05
	ANN	65.60	57.62	44.68	57.40	68.45	67.58
	KNN	86.77	87.32	87.51	79.45	96.66	87.35
	DT	81.29	90.43	81.25	86.96	86.90	78.41
	MSVM	92.07	98.28	90.39	92.28	97.08	86.33
Genetic algorithm	RF	89.62	92.99	91.46	81.59	91.01	71.10
	ANN	58.57	62.54	43.52	65.01	63.89	61.18
	KNN	90.31	85.03	87.76	77.06	94.82	86.25
	DT	81.41	92.85	78.14	79.68	94.50	77.41
	MSVM	92.21	99.68	98.28	94.71	94.26	88.34
Ensemble	RF	93.54	87.80	89.38	85.33	91.58	92.70
	ANN	64.43	62.01	48.73	67.50	69.17	60.18
	KNN	82.82	88.40	89.94	79.57	99.93	81.26
	DT	74.72	91.51	83.92	85.28	91.37	76.59
	MSVM	96.05	94.85	98.46	97.09	91.04	94.42

C. Comparative investigation

The comparative investigation of the proposed and the existing models are indicated in table 5. M. Khajehnejad et al. [14] introduced an effective semi-supervised manifold learning framework to classify healthy controls and MCI patients. After collecting the brain scans from OASIS dataset, voxel morphometry analysis was performed to extract the feature vectors from the collected brain scans. Then, the PCA was used to optimize the dimension of extracted feature values in order to obtain better classification. In the classification section, the label propagation methodology was implemented to classify the images of healthy controls and MCI patients. This Extensive experiment showed that the developed semi-supervised manifold learning framework achieved 93.86% of classification accuracy, 94.65% of sensitivity and 93.22% of specificity on OASIS dataset for early diagnosis of AD. In addition, I. Beheshti et al. [19] introduced a new system based on ranking approaches and classification errors for early diagnosis of AD. Initially, the feature vectors and voxel values were extracted from the brain scans using VBM methodology. Then, the extracted features were ranked by using seven methods namely statistical dependency, information gain, mutual information, Gini index, fisher's criterion, t-test score and Pearson's correlation coefficient. The feature vectors with high scores were fed into SVM classifier to categorize the images as AD, healthy controls and MCI. From the simulation results, the developed framework obtained 92.48% of classification accuracy, 91.07% of sensitivity, and 93.89% of specificity on ADNI dataset for AD detection.

K. Shankar et al. [21] introduced a three phase model for AD detection by using MRI brain scans. Initially, histogram, scale invariant transform and texture features were extracted from the brain scans, which were collected from ADNI dataset. The GGWO algorithm was developed in the second phase to reduce the dimension of extracted features. In the final phase, the classification was accomplished by using various machine and deep learning classifiers like decision tree, CNN and KNN. In that, CNN with GGWO model obtained maximum accuracy of 96.23%, sensitivity of 94.55% and specificity of 96.23% in AD detection. Additionally, A. Savio et al. [23] utilized VBM method to segment GM, WM and CSF from the brain scans, and then mean and standard deviation values were extracted from the segmented regions. The extracted feature values were fed to the RBF-diverse Adaboost-SVM classifier to classify the brain images as AD and MCI classes. On OASIS dataset, the developed model achieved

maximum classification accuracy of 85%, sensitivity of 78% and specificity of 92% in AD detection. Compared to the existing methods, the proposed ensemble feature selection-MSVM model obtained better performance in AD detection in light of classification accuracy, sensitivity and specificity on both ADNI and OASIS datasets.

TABLE V. COMPARATIVE ANALYSIS OF PROPOSED AND EXISTING MODELS

Method	*Dataset*	*Accuracy (%)*	*Sensitivity (%)*	*Specificity (%)*
Semi-supervised manifold learning system [14]	OASIS	93.86	94.65	93.22
Feature ranking and classification errors [19]	ADNI	92.48	91.07	93.89
CNN-GGWO [21]	ADNI	96.23	94.55	96.23
RBF-diverse Adaboost-SVM [23]	OASIS	85	78	92
Proposed ensemble feature selection-MSVM	ADNI	97.57	95.65	97.76
	OASIS	94.58	94.77	96.26

D. Discussion

As stated previously, feature selection is an important phase of AD detection in this research article. Two feature descriptors namely Gabor and LDPV are applied for extracting the features from the collected brain scans, in which the extracted feature vectors are high dimensional in nature that leads to "Curse of dimensionality" problem. So, ensemble feature selection algorithm is proposed to diminish the dimension of extracted feature vectors. The effectiveness of the proposed ensemble feature selection algorithm is represented in the tables 1, 2, 3, and 4. Related to individual feature selection algorithms like grasshopper algorithm, PSO and genetic algorithm, the proposed ensemble feature selection algorithm achieved better performance in AD detection in light of PPV, classification accuracy, sensitivity, NPV, specificity and FM index. In comparative analysis phase, ensemble feature selection-MSVM model showed 9.58% improvement on OASIS dataset and 5.09% improvement on ADNI dataset in AD detection by means of classification accuracy.

5. CONCLUSION

In this research paper, ensemble feature selection-MSVM model is proposed for automatic detection of AD. The proposed model includes two major phases for AD detection that are: segmentation and feature selection. In this article, FCM clustering algorithm is used for segmenting the brain tissues like GM, WM, and CSF, where it provides better results for overlapped datasets and comparatively effective related to other segmentation algorithms. Then, ensemble feature selection algorithm is proposed for selecting the relevant feature vectors from the extracted features to achieve better classification results. In the experimental section, the proposed ensemble feature selection-MSVM model achieved significant performance on both datasets in terms of FM index, PPV, sensitivity, accuracy, NPV, and specificity. Related to the existing methods, the proposed model showed 9.58% and 5.09% improvement in AD detection on OASIS and ADNI datasets, respectively in terms of accuracy. In future work, hybrid segmentation method will be included in the proposed model to further enhance its performance in AD detection.

REFERENCES

[1] Y. Li, L. Zhang, A. Bozoki, D.C. Zhu, J. Choi, and T. Maiti, "Early prediction of Alzheimer's disease using longitudinal volumetric MRI data from ADNI," Health Services and Outcomes Research Methodology, vol. 20, no. 1, pp. 13-39, March 2020.

[2] M. Lopez-Martin, A. Nevado, and B. Carro, "Detection of early stages of Alzheimer's disease based on MEG activity with a randomized convolutional neural network," Artif. Intell. Med., vol. 107, pp. 101924, July 2020.

[3] P. Kishore, C.U. Kumari, M.N.V.S.S. Kumar, and T. Pavani, "Detection and analysis of Alzheimer's disease using various machine learning algorithms," Materials Today: Proceedings, September 2020.

[4] B. Ghoraani, L.N. Boettcher, M.D. Hssayeni, A. Rosenfeld, M.I. Tolea, and J.E. Galvin, "Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning," Biomed. Signal Process. Control, vol. 64, pp. 102249, Feburary 2020.

[5] S. Soundarya, M.S. Sruthi, S.S. Bama, S. Kiruthika, and J. Dhiyaneswaran, "Early detection of Alzheimer disease using Gadolinium material," Materials Today: Proceedings, April 2020.

[6] C. Ge, Q. Qu, I.Y.H. Gu, and A.S. Jakola, "Multi-stream multi-scale deep convolutional networks for Alzheimer's disease detection using MR images," Neurocomputing, vol. 350, pp. 60-69, July 2019.

[7] D. Chyzhyk, M. Graña, A. Savio, and J. Maiora, "Hybrid dendritic computing with kernel-LICA applied to Alzheimer's disease detection in MRI," Neurocomputing, vol. 75, no. 1, pp. 72-77, January 2012.

[8] K.M. Poloni, I.A.D. de Oliveira, R. Tam, R.J. Ferrari, and Alzheimer's Disease Neuroimaging Initiative, "Brain MR image classification for Alzheimer's disease diagnosis using structural hippocampal asymmetrical attributes from directional 3-D log-Gabor filter responses," Neurocomputing, vol. 419, pp. 126-135, January 2020.

[9] Y. Zhu, M. Kim, X. Zhu, D. Kaufer, G. Wu, and Alzheimer's Disease Neuroimaging Initiative, "Long range early diagnosis of Alzheimer's disease using longitudinal MR imaging data," Med. Image Anal., vol. 67, pp. 101825, January 2020.

[10] R.R. Janghel, and Y.K. Rathore, "Deep Convolution Neural Network Based System for Early Diagnosis of Alzheimer's Disease," IRBM, July 2020.

[11] T. Altaf, S.M. Anwar, N. Gul, M.N. Majeed, and M. Majid, "Multi-class Alzheimer's disease classification using image and clinical features," Biomed. Signal Process. Control, vol. 43, pp. 64-74, May 2018.

[12] A. De, and A.S. Chowdhury, "DTI based Alzheimer disease classification with rank modulated fusion of CNNs and random forest," Expert Syst. Appl., pp. 114338, November 2020.

[13] Y. Gupta, K.H. Lee, K.Y. Choi, J.J. Lee, B.C. Kim, and G.R. Kwon, "Alzheimer's disease diagnosis based on cortical and subcortical features," J. Healthcare Eng., March 2019.

[14] M. Khajehnejad, F.H. Saatlou, and H. Mohammadzade, "Alzheimer's disease early diagnosis using manifold-based semi-supervised learning," Brain Sci., vol. 7, no. 8, pp. 109, August 2017.

[15] U.R. Acharya, S.L. Fernandes, J.E. WeiKoh, E.J. Ciaccio, M.K.M. Fabell, U.J. Tanik, V. Rajinikanth, and C.H. Yeong, "Automated detection of Alzheimer's disease using brain MRI images-a study

with various feature extraction techniques," Journal of Medical Systems, vol. 43, no. 9, pp. 302, September 2019.

[16] X. Bi, S. Li, B. Xiao, Y. Li, G. Wang, and X. Ma, "Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology," Neurocomputing, vol. 392, pp. 296-304, June 2020.

[17] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, and Alzheimer's Disease Neuroimaging Initiative, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," NeuroImage: Clinical, vol. 21, pp. 101645, January 2019.

[18] J. Samper-Gonzalez, N. Burgos, S. Bottani, S. Fontanella, P. Lu, A. Marcoux, A. Routier, J. Guillon, M. Bacci, J. Wen, and A. Bertrand, "Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data," NeuroImage, vol. 183, pp. 504-521, December 2018.

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

[20] R.S. Kamathe, and K.R. Joshi, "A novel method based on independent component analysis for brain MR image tissue classification into CSF, WM and GM for atrophy detection in Alzheimer's disease," Biomed. Signal Process. Control, vol. 40, pp. 41-48, Febuary 2018.

[21] K. Shankar, S.K. Lakshmanaprabu, A. Khanna, S. Tanwar, J.J. Rodrigues, and N.R. Roy, "Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier," Computers & Electrical Engineering, vol. 77, pp. 230-243, July 2019.

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

[23] A. Savio, M.T. García-Sebastián, D. Chyzyk, C. Hernández, M. Graña, A. Sistiaga, A.L. De Munain, and J. Villanúa, "Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI," Comput. Biol. Med, vol. 41, no. 8, pp. 600-610, August 2011.

[24] R.C. Petersen, P.S. Aisen, L.A. Beckett, M.C. Donohue, A.C. Gamst, D.J. Harvey, C.R. Jack, W.J. Jagust, L.M. Shaw, A.W. Toga, and J.Q. Trojanowski, "Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization," Neurology, vol. 74, no. 3, pp. 201-209, January 2010.

[25] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, and R.L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," J. Cognit. Neurosci., vol. 19, no. 9, pp. 1498-1507, September 2007.

[26] Y. Zhu, and C. Huang, "An adaptive histogram equalization algorithm on the image gray level mapping," Physics Procedia, vol. 25, pp. 601-608, January 2012.

[27] S. Anand, and S. Gayathri, "Mammogram image enhancement by two-stage adaptive histogram equalization," Optik, vol. 126, no. 21, pp. 3150-3152, November 2015.

[28] Y.A.N. Gao, J.F. Mas, N. Kerle, and J.A. Navarrete Pacheco, "Optimal region growing segmentation and its effect on classification accuracy," International journal of remote sensing, vol. 32, no. 13, pp. 3747-3763, July 2011.

[29] A.K. Qin, and D.A. Clausi, "Multivariate image segmentation using semantic region growing with adaptive edge penalty," IEEE Trans. Image Process., vol. 19, no. 8, pp. 2157-2170, March 2010.

[30] H. Huang, F. Meng, S. Zhou, F. Jiang, and G. Manogaran, "Brain image segmentation based on FCM clustering algorithm and rough set," IEEE Access, vol. 7, pp. 12386-12396, January 2019.

[31] A. Muthukumar, and A. Kavipriya, "A biometric system based on Gabor feature extraction with SVM classifier for Finger-Knuckle-Print," Pattern Recognit. Lett., vol. 125, pp. 150-156, July 2019.

[32] A.M. Shabat, and J.R. Tapamo, "Angled local directional pattern for texture analysis with an application to facial expression recognition," IET Comput. Vision, vol. 12, no. 5, pp. 603-608, Febuary 2018.

[33] M.H. Kabir, T. Jabid and O. Chae, "A Local Directional Pattern Variance (LDPv) Based Face Descriptor for Human Facial Expression Recognition," 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 526-532, August 2010.

[34] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: theory and application," Adv. Eng. Software, vol. 105, pp. 30-47, March 2017.

[35] J.C. Bansal, "Particle swarm optimization, In Evolutionary and swarm intelligence algorithms, Springer, Cham, pp. 11-23, Febuary 2019.

[36] G.T. Reddy, M.P.K. Reddy, K. Lakshmanna, D.S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," Evolutionary Intelligence, vol. 13, no. 2, pp. 185-196, June 2020.

[37] S. Choi, and Z. Jiang, "Cardiac sound murmurs classification with autoregressive spectral analysis and multi-support vector machine technique," Comput. Biol. Med., vol. 40, no. 1, pp. 8-20, January 2010.