ORIGINAL ARTICLE

Genetic algorithm with logistic regression feature selection for Alzheimer's disease classification

R. Divya[1] · R. Shantha Selva Kumari[1] on behalf of the Alzheimer's Disease Neuroimaging Initiative

Abstract

Cognitive impairment must be diagnosed in Alzheimer's disease as early as possible. Early diagnosis allows the person to receive effective treatment benefits apart from helping him or her to remain independent longer. In this paper, different feature selection techniques are utilized with different classifiers in the classification of this chronic disease as normal control (NC), mild cognitive impairment (MCI) and Alzheimer's disease (AD) based on the MRI images of ADNI dataset. Dimensionality reduction plays a major role in improving classification performance when there are fewer records with high dimensions. After different trials to select the ample features, support vector machine (SVM) with radial basis function kernel is found to produce better results with 96.82%, 89.39% and 90.40% accuracy for binary classification of NC/AD, NC/MCI and MCI/AD, respectively, with repeated tenfold stratified cross-validation. Combining mini-mental state examination (MMSE) score to the MRI data, there has been an improvement of 2.7% in the MCI/AD classification, but it does not have much influence in the NC/AD and NC/MCI classification.

Keywords Alzheimer's disease (AD) · Feature selection · Genetic algorithm (GA) · Mini-mental state examination (MMSE) score

1 Introduction

Most elderly people are affected by dementia mainly due to Alzheimer's disease. Alzheimer's disease slowly progresses from mild, moderate to severe stage of dementia. It occurs mainly due to the abnormal build-up of proteins like amyloid plaques and tau tangles in the brain. In the preclinical stage, changes take place in the brain decades before the actual diagnosis of Alzheimer's disease (AD). The next stage is mild cognitive impairment (MCI) where slight but noticeable changes occur in memory and cognitive functionalities. The final stage is dementia due to memory loss and impaired daily activities. None of the treatments available is a complete cure for the disease. There is a strong need for identifying the disease at an earlier stage to effectively implement preventive measures.

Significant studies were carried out to understand the pathological conditions in the brain. Imaging modalities like structural and functional magnetic resonance imaging (sMRI, fMRI) [1–3], positron emission tomography (PET) [4] and cerebrospinal fluids (CSF) [5] were used as biomarkers to classify the disease stage either separately or combined [5–7]. These imaging biomarkers quantify the structural and functional information of the brain. Apart from these biomarkers, mini-mental state examination (MMSE) scores were considered for the diagnosis of AD [1, 2]. MMSE is a 30-point questionnaire for measuring cognitive impairment.

Different approaches were used for identifying AD. Traditional machine learning approaches based on the

✉ R. Divya
  rdivya@mepcoeng.ac.in

  R. Shantha Selva Kumari
  rshantha@mepcoeng.ac.in

1 Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi 626 005, India

features obtained manually were used in [4, 8]. Quantitative studies were developed to analyze the volume, thickness, surface, shape and texture of the brain [9–11]. In [12], region of interest (ROI)-based analysis was used for obtaining the features. Voxel-based morphometry in statistical parametric mapping [13] and volumes generated from FreeSurfer [14] were used to extract the brain features. In [15], a multi-feature kernel discriminant dictionary learning technique that combines the sMRI, fluorodeoxyglucose (FDG) PET and florbetapir-PET imaging features was used. In [16], three different two-dimensional convolutional neural networks were used for Alzheimer's disease classification based on the slice importance of sMRI images. Instead of using the entire brain, the hippocampus region was segmented and used for the disease classification [17]. Networks were constructed from cortical gray matter volume, cortical thickness, cortical surface area, cortical curvature, cortical folding index and subcortical volume. Node and edge features of these networks were selected using F-score and were used for classification [6]. In [18], support vector machine (SVM) used recursive feature elimination (RFE) to reduce model complexity.

Most of the previous studies used only baseline data. Despite many efforts in identifying the biomarkers for early diagnosis, classification of the disease state is still quite hard. In this study, sMRI images are used for the classification of AD disease stages. Volumetric segmentation is used where the images are normalized and registered using Desikan–Killiany atlas. The predominant challenge is the limited records with large dimensions of imaging features. Cortical parcellation volume (CV), subcortical parcellation volume (SV), surface area (SA), cortical thickness average (TA), cortical thickness standard deviation (TS), hippocampal subfields parcellation volume (HS) biomarkers are considered for the classification. Though the RFE and Genetic Algorithm (GA) were used in other studies for the feature selection, there was no study on the classification based on the above-proposed biomarkers. For these biomarkers, different wrapper-based feature selection techniques using RFE and GA are performed to find the optimal feature set. MMSE score is used along with the optimal volumetric segmentation features to check whether it affects the classification performance. Classifier with better predicting accuracy is determined.

2 Methods

Data used in the preparation of this article is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Longitudinal sMRI images are used which are processed for gradient inhomogeneity,

B1 non-uniformity correction, and scaling by the Mayo Clinic. Volumetric segmentation is carried out using FreeSurfer version 5.1 based on the image processing framework mentioned in [19, 20]. The quality control process is used on the segmented images and is provided for the usage in the ADNI website.

For the classification of sMRI imaging data into normal control (NC), mild cognitive impairment (MCI) and Alzheimer's disease (AD), the imaging data which pass the overall quality control process are considered. This reduces the data size considerably due to the partial or total segmentation failure based on quality control. Apart from this neuropsychological measure, mini-mental state examination (MMSE) scores are also obtained from the LONI Image Data Archive website. Missing feature readings are not calculated to avoid bias being introduced in the results. Hence, records with missing values are discarded.

After the preprocessing of data, total records include 347 normal controls (NC), 558 MCI, and 171 AD. Among these records, 27 MCI subjects converted to AD in three years time period from the baseline. The demographic information of the data samples used in this study is shown in Table 1.

3 Features selection

Totally 69 CV, 50 SV, 70 SA, 68 TA, 68 TS, 16 HS features are present for every record. The records are scaled to their inter-quartile range so that they are robust to the outliers. Training time is high if all the features are used for classification purposes. When the feature dimension is large, it leads to problems like overfitting, models with higher complexity, and lesser accuracy. To overcome the above-mentioned problems, feature selection strategies are carried out.

Univariate methods such as t-test and Fisher's criteria were used for feature selection in [21, 22]. Though the features selected were the best as an individual, they might not be the best as a whole. Multivariate feature selection methods do not individually rank the features but they rank sets of features.

Table 1 Demographic information of the data samples

Disease state	Number of records	Gender	MMSE score
NC	347	137M/210F	29.04 ± 1.27
MCI	558	274M/284F	27.68 ± 1.98
AD	171	81M/90F	22.05 ± 4.31

MMSE score is out of 30

RFE is a backward elimination technique that starts with the total feature set, and then the most irrelevant features are removed one after another. This RFE technique is used to select the most prominent and discriminative features from the entire feature set. GA is an evolutionary computing algorithm suitable for searching an efficient subset of features that are optimal from the high-dimensional feature set. In this work, RFE and GA methods are combined with the logistic regression and linear support vector machine classifiers in the wrapper technique to select the highly relevant features from the large feature set. The classifiers are then trained to verify the usefulness of the different sets of the selected features.

3.1 RFE features selection

Recursive feature elimination is a simple heuristic approach to select features that are most relevant for predicting the target. RFE looks for a subset of features by starting with all the features and removes features iteratively until the desired feature set is obtained. It is achieved by fitting LR or SVM model with all the features. Then, features are ranked by importance score and the least important features are discarded. The model is re-fitted. This process is repeated until a specified number of features remains.

3.2 GA features selection

Genetic algorithm is a simple meta-heuristic algorithm that imitates the biological evolution is used for feature selection. In this feature selection, volumetric measures are encoded as genomes using binary strings. The steps involved in GA feature selection process are listed below.

Step 1. Random population of 50 chromosomes are generated which are the solutions.
Step 2. The fitness function is evaluated for each chromosome in the random population. Here, the fitness function is the accuracy measured from the classifier (LR or SVM). This accuracy is calculated using fivefold cross-validation technique.
Step 3. Tournament selection is performed for selecting the parent chromosomes based on their fitness value. Tournament size of 3 is used for this tournament selection.
Step 4. Uniform cross-over operator is used to produce offspring with cross-over probability of 0.5
Step 5. With a mutation probability of 0.1 new offspring is mutated by flipping the binary bits. This offspring is added into the population. For the further iterations, newly generated population is used.

Step 6. If the chromosomes generated are same for the last 10 generations, then stop the process and return the best solution (feature set) from the current population. Otherwise go to step 2.

4 Proposed workflow

ADNI longitudinal sMRI features that are preprocessed and normalized are given as input to the four wrapper-based feature selection processes. The feature subsets obtained are then available to be trained by the classifiers. Three different binary classifications are performed in AD diagnosis: NC/AD, NC/MCI and MCI/AD classification. For this purpose, logistic regression, SVM, random forest (RF) and extreme gradient boosting classifiers (XGB) are used. Based on these classifications, the best feature set is selected. To this feature set, a neuropsychological measure MMSE score is added. Once again, classification is performed on these combined measures. The workflow of this classification process is shown in Fig. 1.

4.1 Classifiers

The following classifiers are considered for the diagnosis of Alzheimer's disease stage using sMRI features.

4.1.1 LR classifier

Consider the training set of n points $\{(x_i, y_i) \mid x_i \in R^m, y_i \in \{0,1\}\}$. Logistic regression generalizes from the linear regression, where sigmoid function is applied to the linear regression function as given by Eq. (1):

$$y = g(\theta^T x)$$
$$\text{where } g(z) = \frac{1}{1 + e^{-z}} \tag{1}$$

This enables to classify the inputs into binary valued labels, $y_i \in \{0,1\}$.

4.1.2 SVM classifier

Consider the training set of n points $\{(x_i, y_i) \mid x_i \in R^m, y_i \in \{\pm 1\}\}$ where $i = 1$ to n, x_i denotes the feature vectors and y_i denotes the class label. SVM maximizes the margin around the hyperplane. This is achieved by solving the following optimization problem in Eq. (2):

$$\min_{w,\xi,b}\left\{\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i\right\} \tag{2}$$
$$\text{s.t. } y_i(w.x_i - b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1,\ldots,n$$
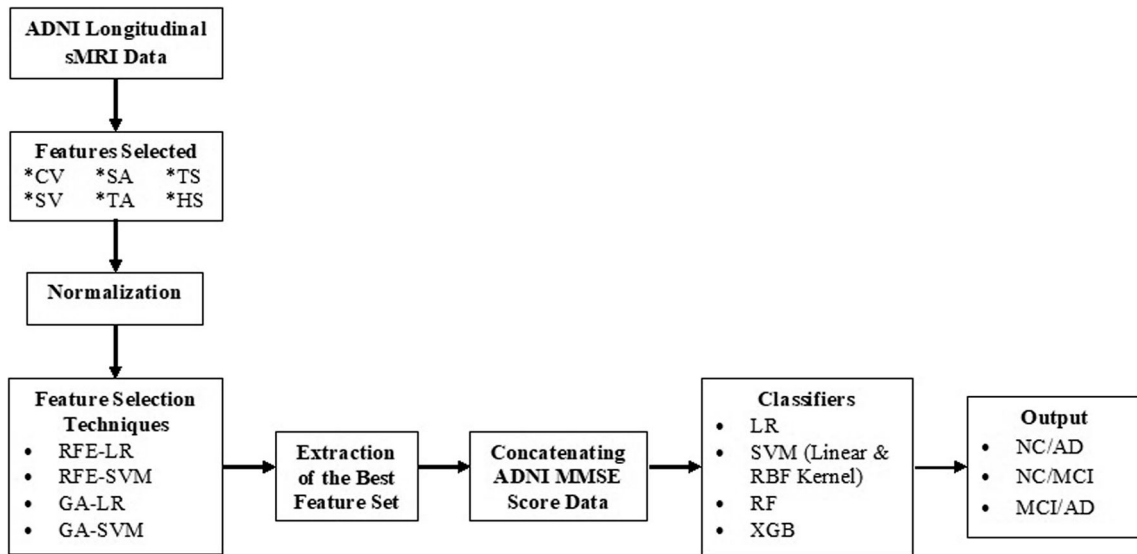
Fig. 1 Workflow of the classification process

When radial basis function kernel is used, dot product is replaced by the Gaussian kernel function which is given in Eq. (3).

$$K\left(x_i, x_j\right) = \exp\left(-\gamma||x_i - x_j||^2\right), \quad \gamma > 0, \tag{3}$$

4.1.3 RF classifier

Random forest is an ensemble tool that builds many decision trees and combines them to produce better results. Consider the m set of points for the training $\{(x_i, y_i)| x_{i-} \in R^m, y_i \in \{0,1\}\}$.

for $b = 1$ to B:

1. n samples are taken with replacement from m set of points.
2. A classifier tree f_b is constructed on these n samples.

After training, predictions for the unseen samples $\hat{x}$ are made by averaging the predictions from all these individual trees as per Eq. (4):

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(\hat{x}) \tag{4}$$

When data are provided to the decision tree, splitting of data is based on Gini index and entropy.

4.1.4 XGBoost classifier

XGBoost (XGB), extreme gradient boosting classifier, is an optimized gradient boosting algorithm. Trees are grown one by one and attempt to reduce misclassification rate in subsequent iterations. The next tree is grown by giving

higher weights to the misclassified points of the previous tree. Consider a dataset $\{(x_i, y_i)| x_i \in R^m, y_i \in \{0,1\}\}$. The objective function for XGBoost is given in Eq. (5).

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y_i}^{(t-1)} + f_t(x_i)\right) + \Omega(f_t)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2}\lambda||w||^2 \tag{5}$$

f_t represents a tree with weights w. T is the number of leaves in a tree. l denotes differentiable convex loss function that measures the difference between the prediction $\widehat{y_i}$ and the target y_i. The second term Ω penalizes the complexity of the model.

4.2 Performance metrics

For analyzing the performance of the models, accuracy, sensitivity and specificity are used. Accuracy is calculated by the formula given in Eq. (6):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{6}$$

Sensitivity is calculated by the formula given in Eq. (7):

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

Specificity is calculated by the formula given in Eq. (8):

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{8}$$

where TP, TN, FP and FN stand for true positive, true negative, false positive and false negative, respectively. The performance measures are the means of these measures computed in the cross-validation runs.

5 Results and discussion

In RFE, the volumetric features corresponding to the lowest rankings are discarded and the remaining metrics are used for training the classifier at each level. The feature ranking for all the three binary classifications using RFE method is shown in Fig. 2.

One hundred and seventy features are chosen among the 341 features for LR and linear SVM wrapper method. Between the two wrapper methods, there are 140, 132 and 118 identical features in NC/AD, NC/MCI and MCI/AD classification, respectively.

In GA, genomes are represented using binary strings which are the set of volumetric measures encoded. The bit whose value is zero indicates the feature is not selected; otherwise, the feature is selected. The initial population is chosen as 50, and tournament size is selected as 3. Forty generations are performed with fivefold cross-validation, where four folds are used for training and one fold is used for testing in all the iterations. A total of 82 to 104 features are chosen among the 341 features for LR and SVM wrapper methods as shown in Fig. 3. There are 19, 44 and 25 identical features in NC/AD, NC/MCI and MCI/AD classification, respectively, when the two wrapper methods are compared.

The features whose ranking scores are 1 based on RFE and GA will be selected for further classification purposes. The remaining features are discarded. After the four feature subsets are identified, hyper-parameters C and γ are tuned using grid search on SVM which is implemented using the LIBSVM library [23]. Fivefold cross-validation is performed for grid search. Similarly, models are built using LR, RF and XGB classifiers.

All the models are developed using tenfold stratified cross-validation, and they are repeated fifteen times. The records are randomly partitioned into ten subsets with roughly the same proportions of the different class labels. Each class is almost equally represented across the training and test fold. Stratified cross-validation is performed due to the imbalance in the data size. For each of the 10 folds, the model is trained with nine folds and the remaining fold is used for testing. This process is repeated fifteen times with different randomization of the samples. The results of the binary classification experiments are presented in Tables 2, 3 and 4.

From Tables 2, 3 and 4, the selected features from the wrapper-based GA-LR algorithm differentiate the disease stages better when compared to the other feature selecting algorithms. From the 341 features of the structural MRI image, only 82, 89 and 84 features are selected for NC/AD, NC/MCI and MCI/AD classification, respectively. This helps in attaining better classifiers with reduced number of features which makes the models have lesser complexity. In the GA-LR feature selection technique, SVM with RBF kernel gives better accuracy when compared with the other classifiers. LR provides the least accuracy in all the classification models. RF and XGB produce approximately equal accuracy results.

The performance metrics of the models developed with MMSE score along with the best feature subset obtained in NC/AD, NC/MCI and MCI/AD classification are shown in Fig. 4. Receiver operating characteristic (ROC) curves for the three binary classifiers are shown in Fig. 5

For each classifier, the ROC curve represents the mean ROC from the 150 cross-validation runs and area under curve (AUC) corresponds to the mean AUC of 150 cross-

Fig. 2 Feature ranking using RFE **a** NC vs AD, **b** NC vs MCI, **c** MCI vs AD
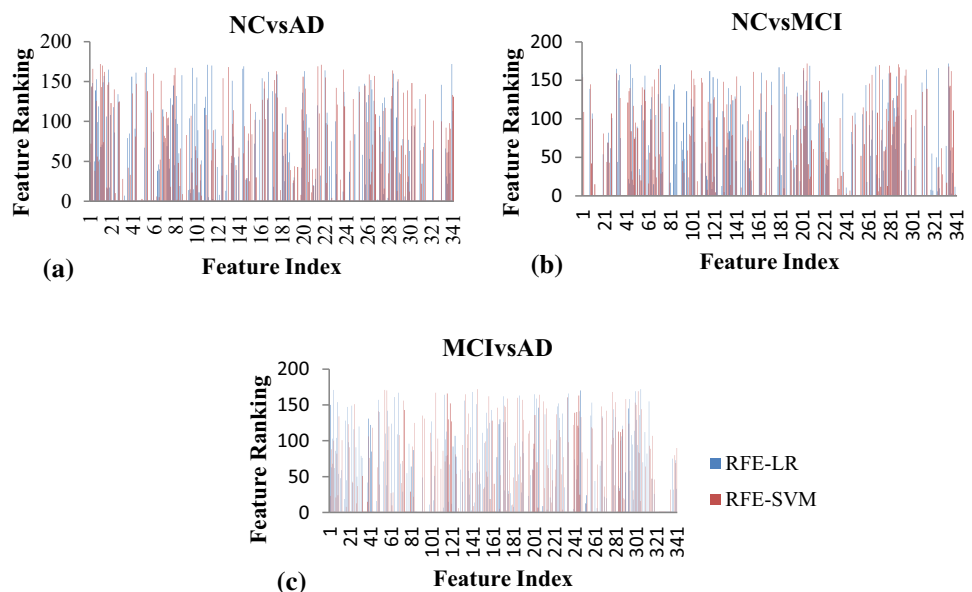
Fig. 3 Features selected using GA **a** NC vs AD, **b** NC vs MCI, **c** MCI vs AD
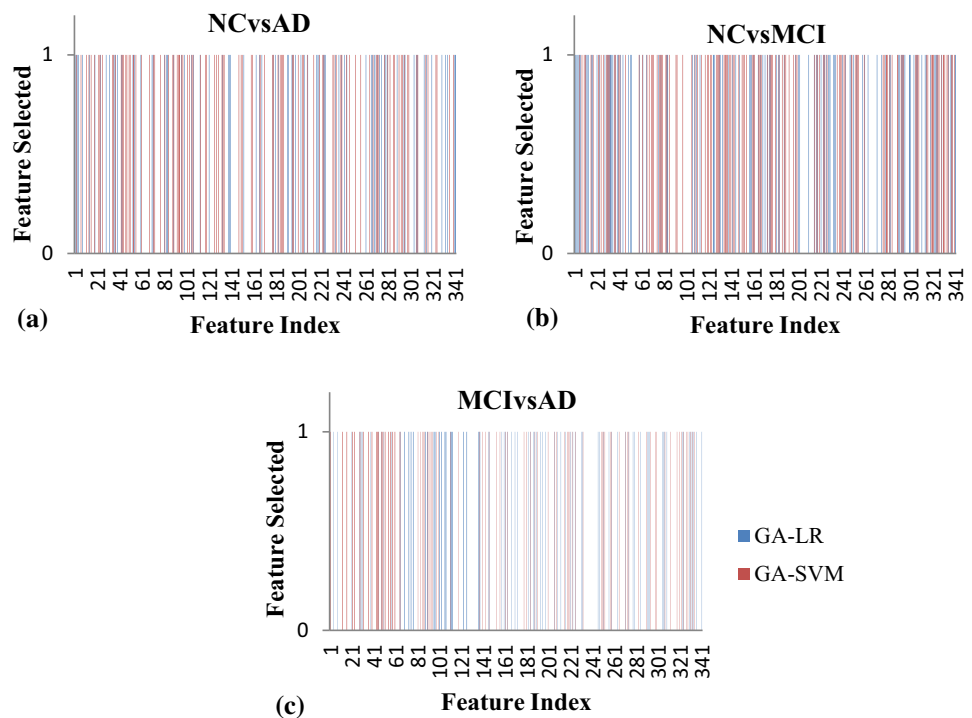

Table 2 NC/AD classification using different feature selection techniques

Feature selection	No. of features	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
RFE-LR	170	LR	92.85	84.90	96.77
		RF	93.61	87.56	96.60
		SVM (linear)	94.39	89.05	97.02
		SVM (RBF) $c = 10$ $\gamma = 0.01$	93.07	84.17	97.47
		XGB	93.82	87.96	96.72
RFE-SVM	170	LR	92.89	85.14	96.72
		RF	93.60	87.48	96.62
		SVM (linear)	94.25	89.12	96.79
		SVM (RBF) $c = 1$ $\gamma = 1$	95.66	91.62	97.65
		XGB	94.02	88.42	96.78
GA-LR	82	LR	91.16	82.37	95.48
		RF	93.28	87.20	96.27
		SVM (linear)	92.11	84.01	96.10
		SVM (RBF) $c = 1$ $\gamma = 1$	96.41	92.17	98.50
		XGB	92.66	86.36	95.77
GA-SVM	104	LR	91.11	81.83	95.69
		RF	93.82	87.48	96.95
		SVM (linear)	92.42	84.33	96.42
		SVM (RBF) $c = 1$ $\gamma = 1$	95.79	89.44	98.92
		XGB	95.13	89.58	97.87

validation runs. SVM with RBF kernel has achieved higher AUC, i.e., 0.99, 0.95 and 0.94, respectively, for NC vs AD, NC vs MCI and MCI vs AD classification, respectively.

Upon adding MMSE score to the feature subset, in the NC/AD classification and NC/MCI classification, there is not much difference in the performance of the model. But,

Table 3 NC/MCI classification using different feature selection techniques

Feature selection	No. of features	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
RFE-LR	170	LR	73.96	86.42	53.90
		RF	86.61	95.19	72.80
		SVM (linear)	74.61	84.26	59.07
		SVM (RBF) $c = 1\ \gamma = 1$	88.37	96.56	75.20
		XGB	84.37	92.44	71.37
RFE-SVM	170	LR	72.25	84.61	52.36
		RF	86.98	94.42	74.20
		SVM (linear)	74.91	84.38	59.65
		SVM (RBF) $c = 1\ \gamma = 1$	88.22	96.56	74.80
		XGB	84.25	92.08	71.64
GA-LR	89	LR	71.06	83.95	50.33
		RF	86.66	94.79	73.57
		SVM (linear)	72.11	83.11	54.41
		SVM (RBF) $c = 1\ \gamma = 1$	89.39	95.21	80.05
		XGB	83.04	90.78	70.59
GA-SVM	93	LR	68.33	82.42	45.66
		RF	86.15	94.58	72.58
		SVM (linear)	69.32	80.10	51.98
		SVM (RBF) $c = 1\ \gamma = 1$	82.76	88.18	74.05
		XGB	80.79	89.59	66.63

Table 4 MCI/AD classification using different feature selection techniques

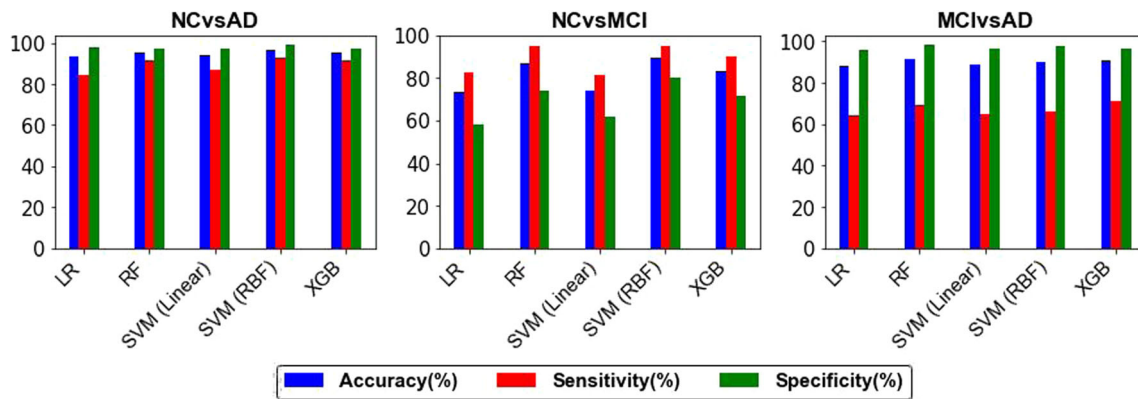| Feature selection	No. of features	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
RFE-LR	170	LR	87.32	60.48	95.56
		RF	88.99	62.86	96.99
		SVM (linear)	87.70	65.05	94.64
		SVM (RBF) $c = 1\ \gamma = 1$	85.19	40.49	98.90
		XGB	88.72	66.68	95.48
RFE-SVM	170	LR	86.91	60.79	94.92
		RF	89.15	62.83	97.23
		SVM (linear)	87.32	64.80	94.23
		SVM (RBF) $c = 10\ \gamma = 0.01$	87.31	57.43	96.47
		XGB	89.06	67.82	95.58
GA-LR	84	LR	85.78	54.28	95.44
		RF	89.04	63.94	96.74
		SVM (linear)	85.72	56.19	94.78
		SVM (RBF) $c = 1\ \gamma = 1$	90.13	66.53	97.37
		XGB	89.09	67.84	95.60
GA-SVM	84	LR	85.61	56.20	94.64
		RF	88.76	62.83	96.71
		SVM (linear)	86.34	59.90	94.46
		SVM (RBF) $c = 1\ \gamma = 1$	89.85	64.32	97.68
		XGB	88.79	67.10	95.45

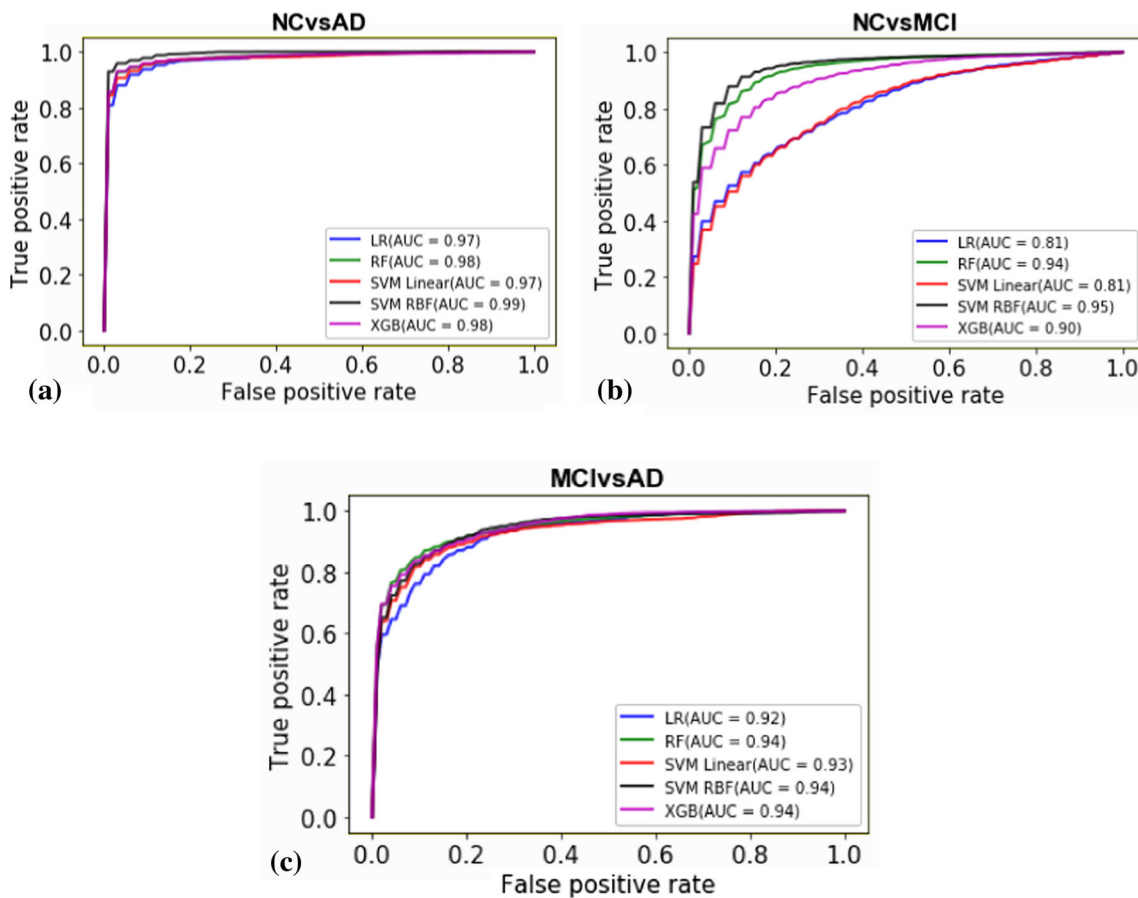**Fig. 4** Classification upon combining MMSE feature to MRI features

Fig. 5 ROC curves for disease classification **a** NC vs AD, **b** NC vs MCI, **c** MCI vs AD

for the MCI/AD classification, accuracy approximately increases by 2.7%, sensitivity improves by 8.7%, and specificity improves by 1.4% for RF classifier when MMSE score is combined with the sMRI features subset. But adding the MMSE score to the features had not improved SVM with RBF kernel performance. On the other hand, the XGB classifier has a better performance over SVM but underperforms when compared to RF. Thus, RF classifier is

ideal for classifying the MCI and AD subjects with improved accuracy.

From the above-mentioned results, the MMSE score does not play a major role in classifying NC/AD or NC/ MCI classification. However, the MMSE score influences RF and XGB classifiers in MCI/AD classification. These results indicate that features derived from sMRI images

Table 5 Performance comparison of binary classifications

Model	MCI vs AD			NC vs AD			NC vs MCI		
	Acc (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)
Ahmed et al. [17]	72.23	75.00	50.00	87	75.50	100	78.22	70.73	83.34
Ren et al. [16]	85.32	78.79	87.30	93.75	94.23	93.40	88.50	82.16	85.24
Liu et al. [6]	90.85	91.77	89.56	95.24	94.26	95.74	86.35	89.49	85.68
Li et al. [15]	74.47	72.44	78.99	98.18	99.81	96.49	78.50	76.00	81.06
Proposed	91.41	69.55	98.11	96.82	92.83	98.79	89.39	95.18	80.07

Acc accuracy, *Sen* sensitivity, *Spe* specificity

play a vital role in the classification rather than the neuropsychological measure MMSE score.

The proposed model outperforms the hippocampus-based method [17] in terms of the performance measures. The model developed achieves better results when compared to the other whole-brain methods [6, 15, 16] mainly because of their smaller and effective feature set as shown in Table 5. The proposed model performs much better in MCI/AD and NC/MCI classification with a single imaging modality. With the minimal feature set, consistent performance over Alzheimer's disease classification is attained.

6 Conclusion

The influence of different feature selection algorithms on CV, SV, SA, TA, TS, HS features of the longitudinal structural MRI images has been investigated. GA-LR feature selection performs better than the other algorithms. Appending MMSE score to the structural MRI reduced feature set helps in improving the accuracy and specificity of the classifier that distinguishes MCI and AD. The model developed from the features selected using GA-LR combined with the MMSE score has a 2.7% increase in accuracy, an 8.7% increase in sensitivity and a 1.4% increase in specificity for RF classifier in MCI/AD classification. SVM with RBF kernel produces better results with 96.82% and 89.39% accuracy for binary classification of NC/AD and NC/MCI, respectively. The proposed models have been developed with lesser features when compared to the existing works and they exhibit high accuracy. As the future work, other imaging measures will be reviewed for improving the classification accuracy.

Compliance with ethical standards

Conflict of interest Authors declare no conflict of interest.

References

1. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104:398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002
2. Minhas S, Khanum A, Riaz F, Khan SA, Alvi A (2018) Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers. IEEE J Biomed Health Inform 22:818–825. https://doi.org/10.1109/JBHI.2017.2703918
3. Ju R, Hu C, Zhou P, Li Q (2019) Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. IEEE/ACM Trans Comput Biol Bioinform 16:244–257. https://doi.org/10.1109/TCBB.2017.2776910
4. Pan X, Adel M, Fossati C, Gaidon T, Guedj E (2019) Multilevel feature representation of FDG-PET brain images for diagnosing Alzheimer's disease. IEEE J Biomed Health Inform 23:1499–1506. https://doi.org/10.1109/JBHI.2018.2857217
5. Zhang D, Wang Y, Zhou L, Yuan H, Shen D (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55:856–867. https://doi.org/10.1016/j.neuroimage.2011.01.008
6. Liu J, Wang J, Tang Z, Hu B, Wu FX, Pan Y (2018) Improving Alzheimer's disease classification by combining multiple Measures. IEEE/ACM Trans Comput Biol Bioinform 15:1649–1659. https://doi.org/10.1109/TCBB.2017.2731849
7. Shi J, Zheng X, Li Y, Zhang Q, Ying S (2018) Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J Biomed Health Inform 22:173–183. https://doi.org/10.1109/JBHI.2017.2655720
8. Minhas S, Khanum A, Riaz F, Alvi A, Khan SA (2017) A Nonparametric approach for mild cognitive impairment to AD conversion prediction: results on longitudinal data. IEEE J Biomed Health Inform 21:1403–1410. https://doi.org/10.1109/JBHI.2016.2608998
9. Xu L, Yao Z, Li J, Lv C, Zhang H, Hu B (2019) Sparse feature learning with label information for Alzheimer's disease classification based on magnetic resonance imaging. IEEE Access 7:26157–26167. https://doi.org/10.1109/ACCESS.2019.2894530
10. Cui R, Liu M (2019) Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis.

IEEE J Biomed Health Inform 23:2099–2107. https://doi.org/10.1109/JBHI.2018.2882392

11. Yue L, Gong X, Li J, Ji H, Li M, Nandi AK (2019) Hierarchical feature extraction for early Alzheimer's disease diagnosis. IEEE Access 7:93752–93760. https://doi.org/10.1109/ACCESS.2019.2926288

12. Liu J, Li M, Lan W, Wu FX, Pan Y, Wang J (2018) Classification of Alzheimer's disease using whole brain hierarchical network. IEEE/ACM Trans Comput Biol Bioinform 15:624–632. https://doi.org/10.1109/TCBB.2016.2635144

13. Li W, Zhao Y, Chen X, Xiao Y, Qin Y (2019) Detecting Alzheimer's disease on small dataset: a knowledge transfer perspective. IEEE J Biomed Health Inform 23:1234–1242. https://doi.org/10.1109/JBHI.2018.2839771

14. Westman E, Muehlboeck JS, Simmons A (2012) Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. NeuroImage 62:229–238. https://doi.org/10.1016/j.neuroimage.2012.04.056

15. Li Q, Wu X, Xu L, Chen K, Yao L (2018) Classification of Alzheimer's disease, mild cognitive impairment, and cognitively unimpaired individuals using multi-feature kernel discriminant dictionary learning. Front Comput Neurosci 11:1–14. https://doi.org/10.3389/fncom.2017.00117

16. Ren F, Yang C, Qiu Q, Zeng N, Cai C, Hou C, Zou Q (2019) Exploiting discriminative regions of brain slices based on 2D CNNs for Alzheimer's disease classification. IEEE Access 7:181423–181433. https://doi.org/10.1109/ACCESS.2019.2920241

17. Ben Ahmed O, Benois-Pineau J, Allard M, Ben Amar C, Catheline G (2014) Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. Multimed Tools Appl 74:1249–1266. https://doi.org/10.1007/s11042-014-2123-y

18. Richhariya B, Tanveer M, Rashid AH (2020) Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). Biomed Signal Process Control. https://doi.org/10.1016/j.bspc.2020.101903

19. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31:968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

20. Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. NeuroImage 61:1402–1418. https://doi.org/10.1016/j.neuroimage.2012.02.084

21. Chu C, Hsu AL, Chou KH, Bandettini P, Lin CP (2012) Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. NeuroImage 60:59–70. https://doi.org/10.1016/j.neuroimage.2011.11.066

22. Kaur T, Saini BS, Gupta S (2018) A novel feature selection method for brain tumor MR image classification based on the Fisher criterion and parameter-free Bat optimization. Neural Comput Appl 29:193–206. https://doi.org/10.1007/s00521-017-2869-z

23. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:1–39. https://doi.org/10.1145/1961189.1961199