

A hybrid method for classifying cognitive states from fMRI data

S. Parida*

*Carrier Software and Core Network Department
Huawei Technologies India Pvt Ltd
Near EPIP Industrial Area, Whitefield
Bangalore-560 066, Karnataka, India
shantipriyap@huawei.com*

S. Dehuri

*Department of Systems Engineering, Ajou University
San 5, Woncheon-dong, Yeongtong-gu
Suwon 443-749, South Korea*

S.-B. Cho

*Department of Computer Science
Yonsei University, Seoul, South Korea*

L. A. Cacha and R. R. Poznanski

*Department of Clinical Sciences
Faculty of Biosciences and Medical Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru, Johor, Malaysia*

[Received 7 April 2014; Accepted 11 August 2015; Published 8 October 2015]

Functional magnetic resonance imaging (fMRI) makes it possible to detect brain activities in order to elucidate cognitive-states. The complex nature of fMRI data requires understanding of the analyses applied to produce possible avenues for developing models of cognitive state classification and improving brain activity prediction. While many models of classification task of fMRI data analysis have been developed, in this paper, we present a novel hybrid technique through combining the best attributes of genetic algorithms (GAs) and ensemble decision tree technique that consistently outperforms all other methods which are being used for cognitive-state classification. Specifically, this paper illustrates the combined effort of decision-trees ensemble and GAs for feature selection through an extensive simulation study and discusses the classification performance with respect to fMRI data. We have shown that our proposed method exhibits significant reduction of the number of features with clear edge classification accuracy over ensemble of decision-trees.

Keywords: Functional magnetic resonance imaging; genetic algorithms; decision tree ensemble; cognitive state classification.

*Corresponding author.

1. Introduction

Present advances in human neuroimaging techniques have documented the prospects of accurate decoding and detecting the contents of a person's conscious experience on the basis of only non-invasive measurements. Neurodynamic functional brain imaging methods allow for real-time monitoring of human brain functioning. Several experiments have been conceived and designed through functional brain imaging to identify brain regions that reflect their specific cognitive processes and activities caused by stimuli or motor action (Fan & Li, 2006; Cacha & Poznanski, 2011; Sun *et al.*, 2013; Bandettini, 2009).

Analyzing each functional magnetic resonance imaging (fMRI) image contains thousands of voxel results in overfitting the classification algorithm, i.e., the classifier can function properly on the training data, but not on new data (Jin *et al.*, 2009). In fMRI analysis, the voxels correspond to information about any kind of mental tasks or stimuli that are used as features, and the analysis finds the voxels in a particular region of interest (ROI). The feature selection option plays a crucial role in the classification process by searching for the best-features subset from the original features, based on the evaluation criteria (Kharrat *et al.*, 2010). Selecting appropriate methods of feature selection may improve the classification accuracy regardless of the sample size (Chu *et al.*, 2012). This is vital in cognitive state classification because it is computationally infeasible to use all of the available features (Kohavi & John, 1997).

Genetic algorithm (GA) with classifier fusion for designing high-performance classifier system (Kuncheva & Jain, 2000; Gabrys & Ruta, 2006; Parvin *et al.*, 2011) can also be successfully applied to the analysis of fMRI data. Kuncheva & Rodríguez (2010) applied the ensemble of decision tree (DT) on fMRI data for classification purpose. Richiardi *et al.* (2010) showed the efficiency of ensemble of DT in brain decoding, which results in high-end classification performance with inter-pretability of results. The effectiveness of GA ensemble method as compared to other algorithms yields slightly better performance (for single classifier with single objective GA) and effectually useful for classification and feature selection problem. Further, it is comparatively more stable in feature selection with high classification accuracy (Zhang & Yang, 2008).

GA-based feature selection techniques have shown consistent successful results in classifying normal or abnormal (Kharrat *et al.*, 2010) states of human brain. Boehm *et al.* (2011) demonstrated the effectiveness of GA as a feature-selection technique for classifying cognitive states of brain activities. Often the classification accuracy is improved by aggregating a group of classifiers, where the DTs are considered as base classifiers (Yang *et al.*, 2010). Plumpton *et al.* (2012) demonstrated that classifiers ensembles are a good approach for fMRI datasets with a large feature-to-instance ratio.

Among different feature selection techniques, GAs are well-established, population-based stochastic algorithms that solve optimization problem in a high-dimensional

feature space (Cordes *et al.*, 2012; Kuncheva & Jain, 2000). The classification accuracy is returned as a measure of the quality of transformation matrix, which is used by GA to search transformation that minimizes the dimensionality of the transformed patterns and maximizes the classification accuracy (Kharrat *et al.*, 2010).

The DT classification algorithm is most commonly used due to its easy implementation and understanding as compared to other classification algorithms (Anyanwu & Shiva, 2009). The node of DT contains either a single feature f (monolithic tree) or a function of multiple features $f(\cdot)$ (polythetic trees). When an entropy-based splitting criterion is used, the goal is to find cut points of f which minimizes the conditional entropy on class labels $c = \{1, \dots, C\}$ attached to the points in the corresponding sub domains of the discretized variable f' . The entropy of the dataset is partitioned by feature f' and is represented as follows:

$$H(C|f') \triangleq - \sum_{j=1}^2 p_j \sum_{c=1}^C P_{j,c} \log_2 P_{j,c}, \quad (1)$$

where P_j denotes the relative frequency of points in the subset that have value j for feature f' and $P_{j,c}$ is the relative frequency of points that belongs to class c and have value j for feature f' . The goal of DT growth is to minimize Eq. (1), which involves recursively selecting features (or discriminant functions) and computing the result obtained by applying different cut-off points. The edge weight (the features correspond to dimensions in the feature vector, which in turn correspond to edge weights in the functional connectivity graph) is put as zero by the feature selection and therefore are never included as a feature. The learning procedure is divided into two phases: growing and pruning. In the growing phase, at each decision node, a single feature or a discriminant function based on linear combination of features is used; in the pruning phase, the functional leaves containing the discriminant functions are replaced by a simple function that predicts the class value (Richiardi *et al.*, 2011).

Multiple component learners are being trained for the same task, and predictions of the component learners are being combined to deal with future instances (Zhou & Tang, 2003). The classifier ensemble has a significant role in classification as it has been observed that individually, classifiers are not very accurate in classification and tend to make mistake with different objects, which may form a very accurate ensemble (Kuncheva & Rodríguez, 2010). The ensemble methodology weighs the individual classifiers and combines them to obtain a classifier that outperforms any of the individual classifiers (Rokach, 2010).

The ensemble method is divided into two steps: first, to impart training to multiple component learners and then to combine their predictions (Zhou & Tang, 2003). An ensemble is constructed using popular techniques such as Bagging and Boosting (Banfield *et al.*, 2007). In Bagging, each training set is generated by forming a bootstrap replicate of the training set. Given a training set S of m examples, the new

training set S' is formed by uniformly drawing m examples (with replacement) from S (Dietterich, 2000). It is effective on unstable learning algorithms such as DT, where small changes in the training set results in large changes in the predictions (Maclin & Opitz, 2011). Boosting creates an ensemble of classifiers by re-sampling the training dataset, which is then combined by majority of voting. However, in boosting, re-sampling is directed to provide the most informative training data for each consecutive classifier. The AdaBoost (Adaptive Boosting) algorithm constitutes the best known member of boosting family. It generates a sequence of base classifiers C_1, C_2, \dots, C_M by using successive bootstrap samples T_1, T_2, \dots, T_M that are obtained by weighting the training instances in M iterations. It adjusts the initial equal weights assigned to all the training instance based on the misclassifications made by the resulting base classifier during successive iterations (Freund & Schapire, 1996).

Currently, hybrid approaches have achieved widespread adoption as a way of determining cognitive-state classification due to the complexities involved in fMRI data analysis and the limitations of an independent classifier. Only limited studies have considered DTs and GA as a hybrid technique for cognitive-state classification. The present study selected a set of relevant attributes via GAs and supplied them to an ensembler based on DT for classification purpose to achieve low computation time and high interpretability without compromising the accuracy of the classifier.

2. Method

The overall framework of the proposed method is shown in Fig. 1. In the proposed technique, the design process in fMRI data is supplied as an input to GA for selecting the most promising features from the high-dimensional dataset (2.1). The selected features are used to construct a classifier based on an ensemble of DTs for classification (c.f., 2.2). The classification accuracy of the constructed classifier was evaluated using the confusion matrix based on the classification result obtained by test data.

2.1. fMRI feature selection using GAs

The GA-based feature selection operation in fMRI context is illustrated in Fig. 2.

The design of population, fitness function and configuration parameters of the genetic-based feature selection is discussed as follows. The initial population is generated by populating a matrix with a dimension of population size (rows) and an independent variable known as the chromosome length (columns). The values in this matrix are integers, which are randomly selected from the processed input data based on ranking, as shown in Fig. 3.

The fitness of the population is estimated using the fitness function (e.g., classification accuracy). We have used the fitness function provided by the GA Toolbox, which maximizes the separation ability of two classes using a linear combination of the posterior probability and an empirical error rate of linear classifier.

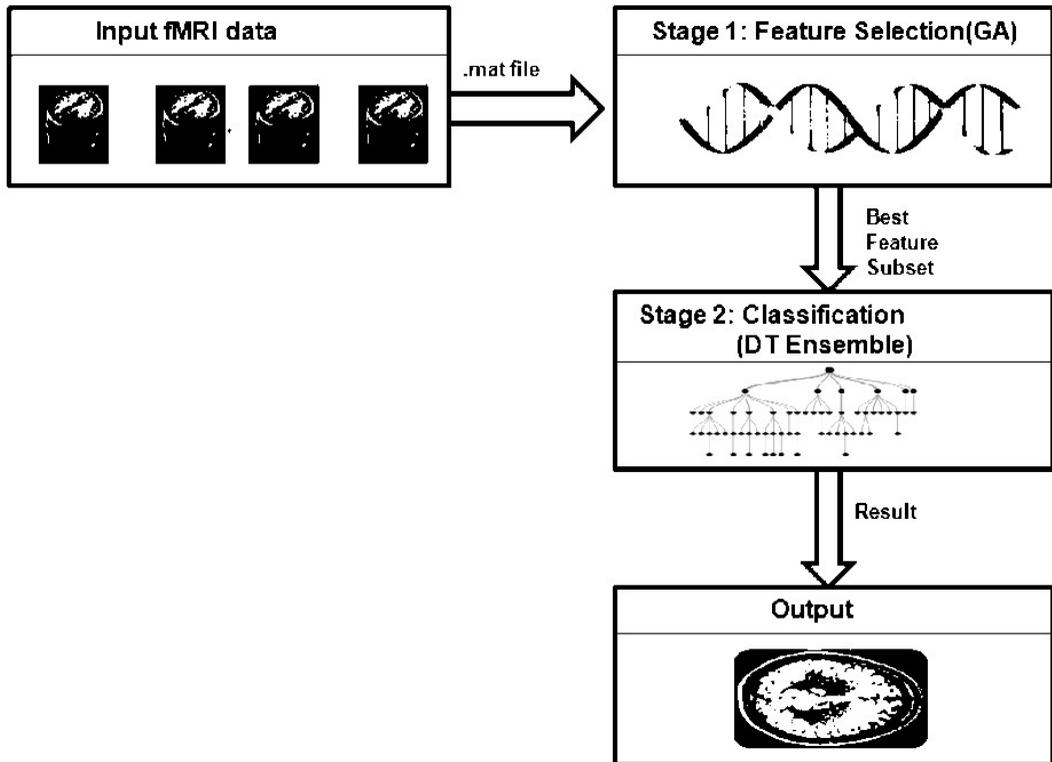


Fig. 1. A block diagram of the proposed technique. The processed StarPlus fMRI data is input to the experiment (7 ROIs considered for better classification accuracy). In Stage 1: The GA applied to select the best features from the high-dimensional processed fMRI data. In Stage 2: The DT ensemble method is used for the cognitive state classification.

We selected the Roulette wheel selection method, in which a circle is divided into N sectors and the width of each sector is proportional to individual's fitness value. The random selection is made similar to how the roulette wheel is rotated (Sharma *et al.*, 2012). A crossover point is randomly set, where portions of the two chromosomes beyond this cut-off point to the right exchanged to form the offspring. An operation rate P_c is used as a probability of crossover (Man *et al.*, 2012). The two point crossover is selected, where two points to be selected on the parent strings and all between the two points is swapped between the parents to generate the child strings. The mutation genetic operator randomly alters the values of genes in the parent string. At each step of evolution, crossover and mutation were applied stochastically to set their probabilities of occurrence (Scrucca, 2012). The uniform mutation was selected as the mutation function. The parameters configured for GA are: population size=25; number of generations=100; selection function=Roulette wheel selection; crossover function= Two point crossover; mutation function=0.02.

Based on the above-listed initial population and the parameter configuration, the input passed to GA Toolbox provided GA function, which returns the best

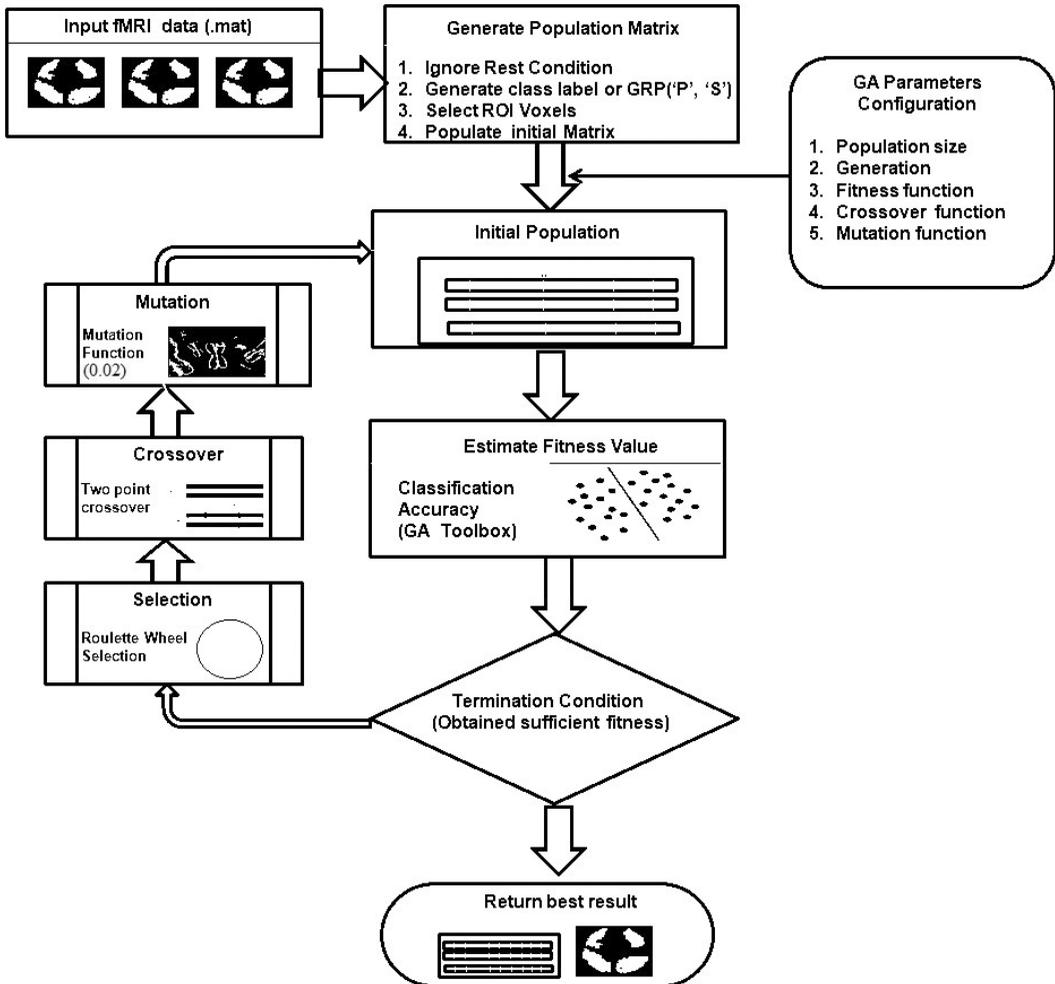


Fig. 2. Stage 1. GA-based Feature Selection. In the first step, the initial population matrix generated from the fMRI data (the rest conditions are ignored and only selected ROIs are considered). In the second step, the GA parameters (fitness/crossover/mutation functions) are configured. In the third step, the fitness value was estimated and checked for the termination conditioned till the sufficient fitness (this process continues in a loop as shown in the figure) is obtained. In the final step, the best features are selected which are used for classification subsequently.

features. The GA function runs multiple times to obtain the best set of features that can contribute significantly less number of features for subsequent DT-ensemble classification.

2.2. Decision tree ensemble

To overcome the issue with overfitting of learning to use single DT, the ensemble of DTs is popularly adopted (Paul *et al.*, 2012). The DT ensemble is an extremely popular ensemble because DTs are unstable classifiers, wherein the output undergoes significant changes in response to small changes in the training data. It is often

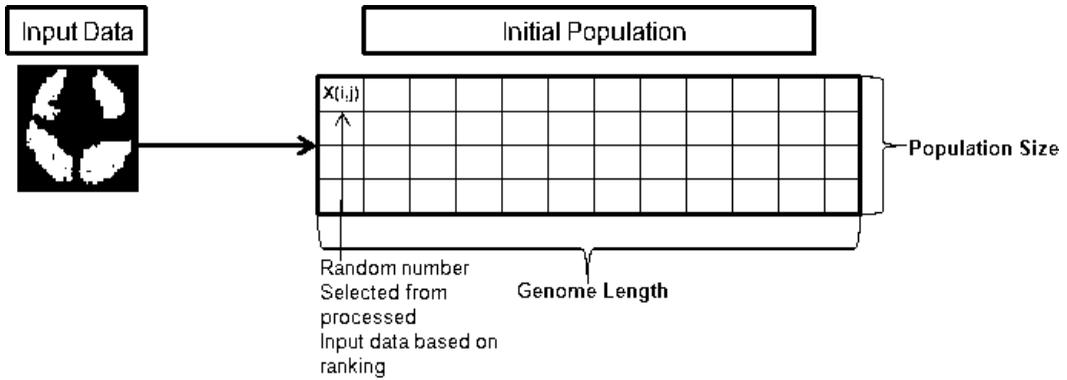


Fig. 3. Structure of the Initial Population Matrix. The initial population matrix as shown in the figure is populated from the processed fMRI data (StarPlus). To create the matrix, the elements are selected randomly from the input data based on the ranking. The GA is applied on the population matrix to obtain the best features.

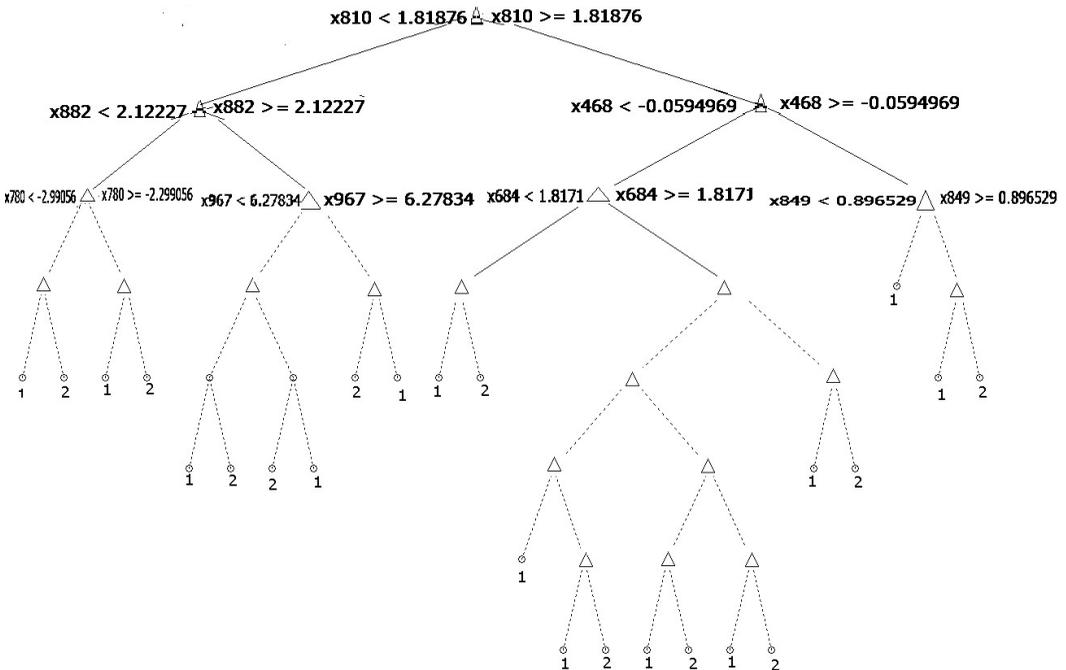


Fig. 4. Decision Trees (GA + Ensemble DT). The figure shows a DT grown by Matlab for the GA with Ensemble DT technique for subject 1. The leaf node shows the class label i.e., 1 for class label ‘Picture’ and 2 for ‘Sentence’. The branching node is split based on the feature value. The figure shows the branching node with values up to level 3 in the DT. At the training phase, the training data is utilized to split the nodes and build the DT. At the testing phase, test data pass through the test to obtain the class label. The tree predicts classification based on the features (ex: x810, x882, etc.) traversing from the root represented by a triangle (Δ) (beginning) down to a leaf node. At each decision, check the values of the features to decide which branch to follow (ex: as shown in the figure, if x810 less than 1.81876 follow left, else if x810 greater than or equal to 1.81876 then follow right).

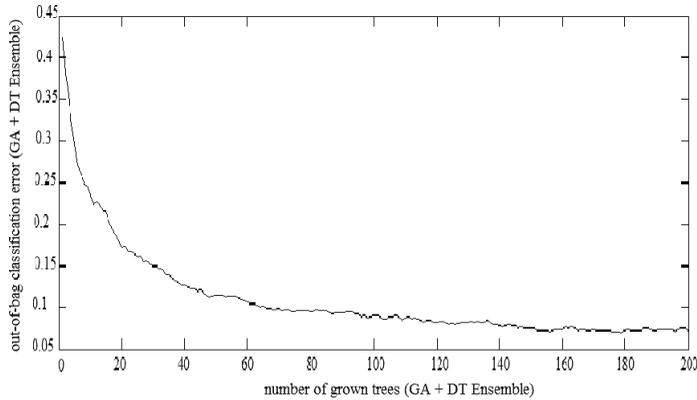


Fig. 5. Out-of-bag classification error. The number of trees used for GA with DT ensemble is 200. The figure shows the growth of trees with the out-of-bag classification error (subject 1). The horizontal axis shows the DT growth and vertical axis represents the classification error. The selected features using GA leading to a smallest OOB error.

possible to create synergy within DT ensembles. The ensemble of DT generally performs better in comparison to a single DT due to its representational power. However, it shows poor results on data that contains several irrelevant attributes (Gashler *et al.*, 2008).

We used the ensemble of DTs for classifying the true-class labels. The classification result was obtained by the voting of the trees. The best features obtained by the GA are input to the DT ensemble. We have used 200 trees for both DT ensembles with and without GA. Figure 4 shows the DTs for the ensemble DT with GA for subject 1. The dataset is divided into (80–20) ratio where 80% samples formed the training dataset and the 20% formed the test dataset. The number of training, test data are 1757 and 439 for both the subjects (04847, and 04799).

The training data was used for learning DTs and the test data was used for predicting the classification accuracy. This method estimates the out-of-bag classification error and then compares it with the number of trees grown, as shown in Fig. 5. Based on the prediction scores, the confusion matrix is created. The classification accuracy was calculated by using the confusion matrix.

3. Data Collection

The experiment was performed on a system running on the platform of 32 bit, Intel 2.70 GHz processor, 4.00 GB RAM, and Windows 7 operating system. The programs for the experiment were coded by using the MATLAB R2010a (The Mathworks c). The MATLAB provided GA Toolbox functions, which was used for feature selection. The fMRI dataset was collected from the Carnegie Mellon University (CMU)'s public StarPlus fMRI data repository (Just & Mitchell, 2001). The data were obtained for two subjects ('04847' and '04799') and were partitioned into trials. The experiment consisted of a set of trials. For some of these intervals, the subject is simply rested, or

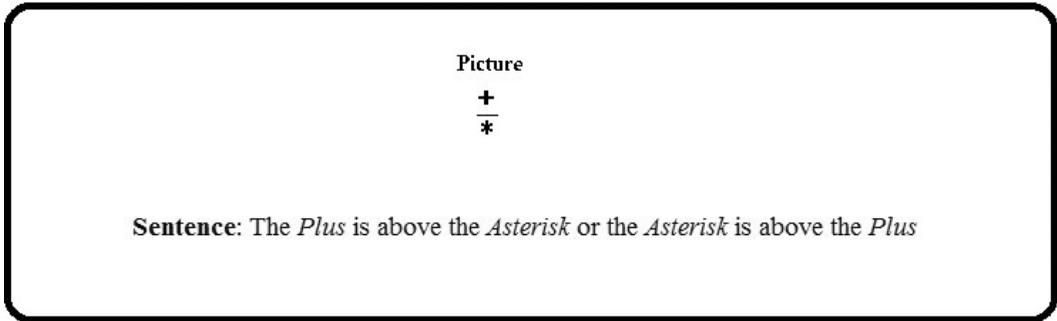


Fig. 6. Pictures contained geometric arrangement of the two symbols +, and *. In half of the trials, the picture was presented first, followed by the sentence. In the remaining trials, the sentence was presented first, followed by the picture. The first stimulus (sentence or picture) was presented at the beginning of the trail (image=1). After 4 s (image=9), the stimulus was removed and replaced by a blank screen. After 4 s (image=17), the second stimulus was presented. This remained on the screen for 4 s or until the subject pressed the mouse button, whichever came first. A rest period of 15 s (30 images) was added after the second stimulus was removed from the screen. Thus, each trial lasted a total of approximately 27 s (approximately 54 images).

gazed at a fixation point on the screen. In other trials, the subjects were shown an image and a sentence, and were instructed to press a button to indicate whether the sentence correctly described the picture, as shown in Fig. 6. For these trials, the sentence and picture were presented in sequence, half of the trails were shown the picture first, while the other half were shown the sentence first.

The images were collected every 500 ms. A total of 54 trials and 2800 snapshots were collected. The data were stored in a $[54 \times 1]$ cell array with one cell per/trial in

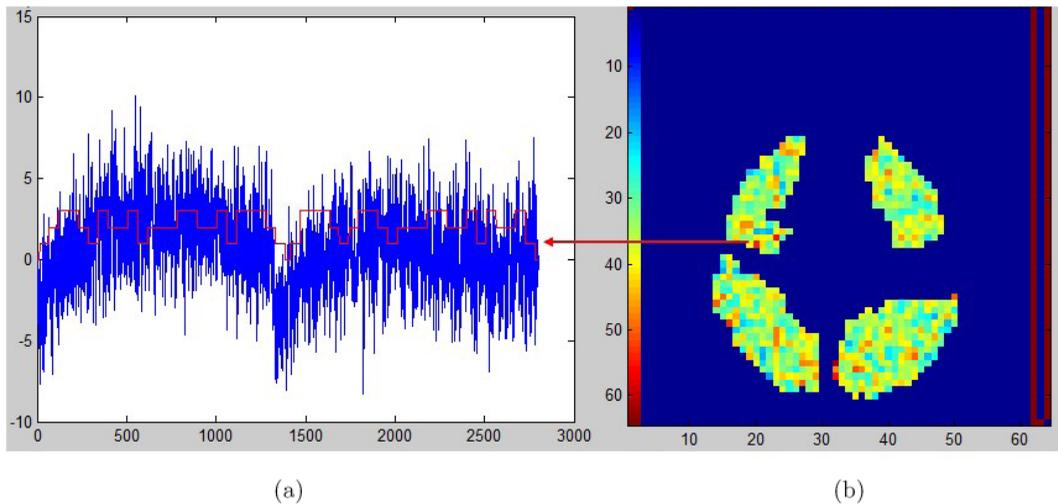


Fig. 7. (Color online) Voxel activity on a particular course of time. The image (a) shows time series data of a single voxel (arrow marked). The X-axis shows the time, and Y-axis shows the activation. The histogram (marked red) indicate the signal intensity. The image (b) shows the 2D slice image for a particular trial (trial=4) and the scale interval $[1, 64]$ (default matlab color map) is shown in the X, and Y axis.

the experiment. Each element in the cell array is an $N \times V$ array of the observed fMRI activations, where N is the number of snapshots and V is the number of voxels (features). The element data $x(t, v)$ denotes the fMRI observation at voxel v , at time t within trial x .

The sample of voxel activity at a particular time-course is shown in Fig. 7. The GA was applied to reduce the number of features (V) of each of the $N \times V$ array. During the initial population generation, we ignored the Cond=0, which indicates data to ignore, and Cond=1, which indicates, a segment is a rest or fixation interval. To train the classifier to distinguish whether the subject is viewing a picture or sentence and to obtain the best accuracy, we restricted the classifier input to 7 ROIs ('CALC' 'LIPL' 'LT' 'LTRIA' 'LOPER' 'LIPS' 'LDLPFC') determined by a domain expert to be most relevant. After selecting 7 ROIs, the number of voxels (features) reduced to 1715 for subject 1 (04847) and 1874 for subject 2 (04799). The dataset represented as a pattern and feature matrix of size 2196 and 1715 for subject 1 (04847), and 2196 and 1874 for subject 2 (04799) i.e., $D = (\cdot)_{2196 \times 1715}$ for subject 1 and $D = (\cdot)_{2196 \times 1874}$ for subject 2, where D is denoted as dataset.

4. Results

The important features for ensemble DT with GA for subject 1 are shown in Fig. 8. The performance measurement was estimated as percentages:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \quad (2)$$

$$\text{Sensitivity} = \frac{(\text{TP})}{(\text{TP} + \text{FN})}, \quad (3)$$

$$\text{Specificity} = \frac{(\text{TN})}{(\text{TN} + \text{FP})}, \quad (4)$$

where TP (True Positives) = correctly classifier positive cases, TN (True Negatives) = correctly classifier negative cases, FP (False Positives) = incorrectly classified negative cases, FN (False Negatives) = incorrectly classified positive cases.

The performance value (subject 1) for ensemble DT without GA are: sensitivity= 47; specificity= 86.88; and accuracy= 91.11 and for ensemble DT with GA are: sensitivity= 48.79; specificity= 92.6; and accuracy= 94.76. The performance value (subject 2) for ensemble DT without GA are: sensitivity= 51.99; specificity= 94.60; and accuracy= 91.57 and for ensemble DT with GA are: sensitivity= 50.12; specificity= 91.78; and accuracy= 91.79.

We have compared the result in terms of the number of features vs. the accuracy, and computation time for ensemble DT without GA vs. ensemble DT with GA for both the subjects. As comparing ensemble DT without GA with ensemble DT with GA, the number of features reduced from 1715 to 897, whereas the accuracy increased from 91.11 to 94.76 with reduced computation time (in s) from 5942.544 to

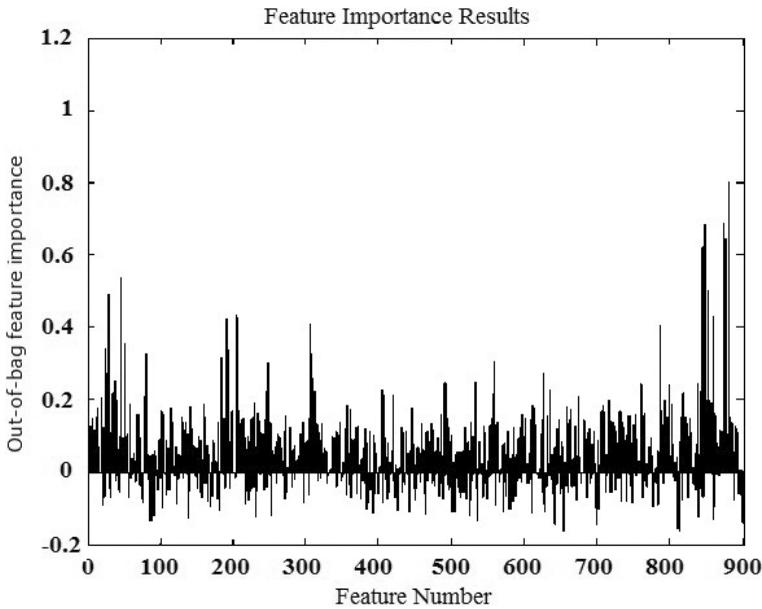
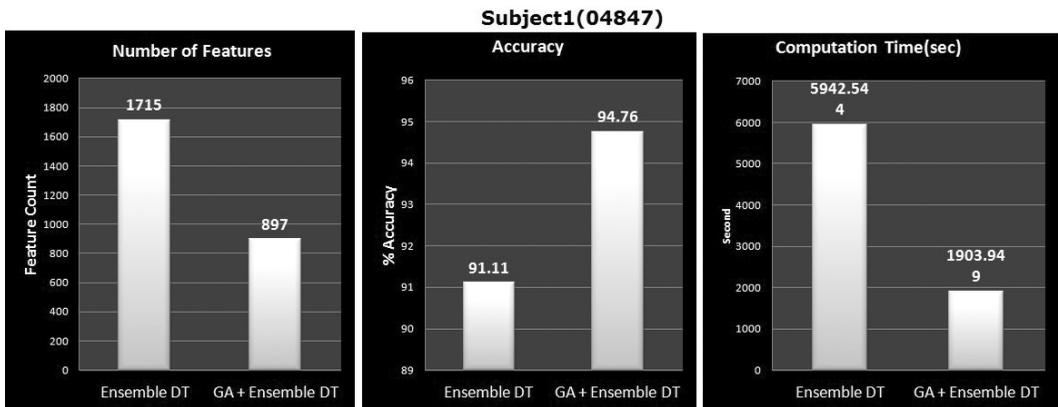


Fig. 8. Important Features. The figure shows the important features which have significant contribution in the GA-based ensemble DT technique for subject 1. The vertical bar represents the important feature (i.e., the important 897 features from the total 1715 number of features). The X-axis shows the OOB feature importance based on mean squared error (MSE).

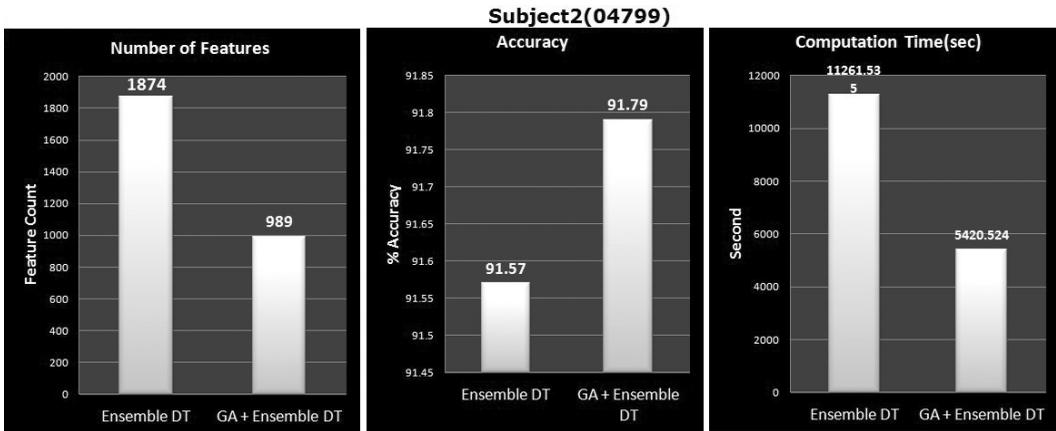
1903.949 for subject 1. Similarly, for subject 2, the number of features reduced from 1874 to 989, whereas the accuracy increased from 91.57 to 91.79 with reduced computation time (in s) from 11261.535 to 5420.524.

The comparison graphs in terms of number of features, accuracy and computation time are shown in Fig. 9. The computation graphs in Figs. 9(a) and 9(b) show the



(a)

Fig. 9. Comparison Graphs. (a) Represents the comparison of number of features, accuracy and computation time between the ensemble DT with and without GA-based feature selection for the subject 1 (04847). (b) represents the comparison of number of features, accuracy and computation time between the ensemble DT with and without GA-based feature selection for the subject 2 (04799).



(b)

Fig. 9. (Continued)

comparison between ensemble DT technique without and with GA-based feature selection for subject 1 and subject 2. The comparison in terms of number of features, classification accuracy and computation time is shown in separate graphs. The X-axis shows the comparison techniques and the Y-axis shows the number of features, accuracy percentage and the execution time in number of seconds. The comparison graphs shows the efficiency of the hybrid technique by increase in classification accuracy (from 91.11% to 94.76% for subject 1 and from 91.57% to 91.79% for subject 2) using reduced number of features (897 out of 1715 for subject 1 and 989 out of 1874) applying the GA-based feature selection technique. It reduced the number of features to 52.30% for subject 1 and 52.77% for subject 2. The reduction in number of features results in fast execution as shown in the comparison graph (computation time).

5. Discussion

Although many classification techniques have been developed for fMRI data analysis, none of the techniques could consistently outperform well for all datasets which raised the focus and potential of hybrid ensembled techniques. It is difficult to construct a single accurate DT during large feature dimensionality. The DT considered in the proposed approach as classifier ensembles using DT are very successful.

We have developed a hybrid method that combined the best attributes of GAs and an ensemble of DTs in classifying cognitive states for fMRI data analysis. The key differentiators of the proposed method include the significant reduction in the number of features when compared with the original dataset, hence slightly smoother matrix which is effective in fMRI-based data analysis, and the faster execution time emphasizes the importance of such methods to explore further. The directions for

future enhancement of the proposed method include continued improvement in the performance in terms of accuracy, analyses of the result for multiple subject and different cognitive tasks, exposing another potential area, creating discriminative graph for the selected features by GA and applying DT ensemble for classification purpose. The fMRI experiment involves different cognitive activities and various instantaneous states, which presents high-dimensional dataset; therefore, it demands techniques that can reduce the number of features and minimize the computation cost without compromising with prediction accuracy.

REFERENCES

- Anyanwu, M.N. & Shiva, S.G. (2009) Comparative analysis of serial decision tree classification algorithms. *Int. J. Computer Sci. Security*, **3**, 230–240.
- Bandettini, P.A. (2009) Seven topics in functional magnetic resonance imaging. *J. Integr. Neurosci.*, **8**, 371–403.
- Banfield, R.E., Hall, L.O., Bowyer, K.W. & Kegelmeyer, W.P. (2007) A comparison of decision tree ensemble creation techniques. *IEEE Trans. Pattern Anal.*, **29**, 173–180.
- Boehm, O., Hardoon, D.R. & Manevitz, L.M. (2011) Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms. *Int. J. Machine Learn. Cybern.*, **2**, 125–134.
- Cacha, L.A. & Poznanski, R.R. (2011) Associable representations as field of influence for dynamic cognitive processes. *J. Integr. Neurosci.*, **10**, 423–437.
- Chu, C., Hsu, A.L., Chou, K.H., Bandettini, P. & Lin, C. (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.
- Cordes, D., Herzmann, G., Nandy, R. & Curran, T. (2012) Optimization of contrast detection power with probabilistic behavioural information. *NeuroImage*, **60**, 1788–1799.
- Dietterich, T.G. (2000) An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learning* **40**, 139–157.
- Fan, J. & Li, R. (2006) Statistical challenges with high dimensionality: Feature selection in knowledge discovery. arXiv preprint math/0602133, 1096–1109.
- Freund, Y. & Schapire, R.E. (1996) Experiments with a new boosting algorithm, In: L. Saitta ed. *Proc. Thirteenth International Conference on Machine Learning (ICML '96)*, pp. 148–156.
- Gabrys, B. & Ruta, D. (2006) Genetic algorithms in classifier fusion. *Appl. Soft Comput.*, **6**, 337–347.
- Gashler, M., Giraud-Carrier, C. & Martinez, T. (2008) Decision tree ensemble: Small heterogeneous is better than large homogeneous. in: *Proc. Seventh International Conf. on Machine Learning and Applications, 2008. ICMLA'08*, pp. 900–905.
- Jin, B., Strasburger, A., Laken, S., Kozel, F.A., Johnson, K., George, M. & Lu, X. (2009) Feature selection for fMRI-based deception detection. *BMC Bioinf.*, **10**, S15.
- Just, M. & Mitchell, T. (2001) StarPlus fMRI data [Online], Available at <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>.

- Kharrat, A., Gasmi, K., Messaoud, M.B., Benamrane, N. & Abid, M. (2010) A hybrid approach for automatic classification of brain MRI using genetic algorithm and support vector machine. *Leonardo J. Sci.*, **17**, 71–82.
- Kohavi, R. & John, G.H. (1997) Wrappers for feature subset selection. *Artif. Intell.*, **97**, 273–324.
- Kuncheva, L.I. & Jain, L.C. (2000) Designing classifier fusion systems by genetic algorithms. *IEEE Trans. Evol. Comput.*, **4**, 327–336.
- Kuncheva, L.I. & Rodríguez, J.J. (2010) Classifier ensembles for fMRI data analysis: An experiment. *Magn. Reson. Imaging*, **28**, 583–593.
- Maclin, R. & Opitz, D. (2011) Popular ensemble methods: An empirical study. arXiv:1106.0257 40.
- Man, K., Tang, K. & Kwong, S. (2012) Genetic algorithms: Concepts and applications. *IEEE Trans. Ind. Electron.*, **43**, 519–534.
- Parvin, H., Minaei, B., Beigi, A. & Helmi, H. (2011) Classification ensemble by genetic algorithms. In: A. Dobnikar, U. Lotric and B. Ster, eds. *Adaptive and Natural Computing Algorithms*. Berlin, Heidelberg: Springer, pp. 391–399.
- Paul, J., Verleysen, M. & Dupont, P. (2012) The stability of feature selection and class prediction from ensemble tree classifiers, in *ESANN2012 Special Session on Machine Ensembles*, pp. 263–268.
- Plumpton, C.O., Kuncheva, L.I., Oosterhof, N.N. & Johnston, S.J. (2012) Naive random subspace ensemble with linear classifiers for real-time classification of fMRI data. *Pattern Recognit.*, **45**, 2101–2108.
- Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P. & Van De Ville, D. (2010) Brain decoding of fMRI connectivity graphs using decision tree ensembles. In: *Proc. 2010 IEEE International Symp. Biomedical Imaging: From Nano to Macro*, pp. 1137–1140.
- Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P. & Van De Ville, D. (2011) Decoding brain states from fMRI connectivity graphs. *Neuroimage*, **56**, 616–626.
- Rokach, L. (2005) Ensemble methods for classifiers, In: O. Maimon and L. Rokach, eds. *Data Mining and Knowledge Discovery Handbook*. US: Springer, pp. 957–980.
- Rokach, L. (2010) Ensemble-based classifiers. *Artif. Intell. Rev.*, **33**, 1–39.
- Scrucca, L. (2012) GA: A package for genetic algorithms in R. *J. Stat. Softw.*, **53**, 1–37.
- Sharma, D., Singh, V. & Sharma, C. (2012) GA based scheduling of FMS using roulette wheel selection process. In: K. Deep, A. Nagar, M. Pant and J.C. Bansal, eds. *Advances in Intelligent and Soft Computing*. India: Springer, pp. 931–940.
- Sun, H., Zheng, D., Wang, X., Lu, Z., Theysohn, N., Forsting, M. & Guo, Q. (2013) Functional segregation in the left premotor cortex in language processing: Evidence from fMRI. *J. Integr. Neurosci.*, **12**, 221–233.
- Yang, P., Hwa Yang, Y., Zhou, B. & Zomaya, A. (2010) A review of ensemble methods in bioinformatics. *Curr. Bioinf.*, **5**, 296–308.
- Zhang, Z. & Yang, P. (2008) An ensemble of classifiers with genetic algorithm based feature selection. *IEEE Intell. Inform. Bull.*, **9**, 18–24.
- Zhou, Z.H. & Tang, W. (2003) Selective ensemble of decision trees, In: G. Wang, Q. Liu, Y. Yao and A. Skowron, eds. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*. Berlin, Heidelberg: Springer, pp. 476–483.