11th International Conference on Vibration Problems
Z. Dimitrovová *et al.* (eds.)
Lisbon, Portugal, 9-12 September 2013

GENETIC ALGORITHMS BASED FEATURE SELECTION FOR COGNITIVE STATE CLASSIFICATION USING ENSEMBLE OF DECISION TREES

Shantipriya Parida*[1], Satchidananda Dehuri[2], Gi-Nam Wang[3]

[1]Carrier Software & Core Network Department, Huawei Technologies India Pvt Ltd,
The Leela Palace, 23 Airport Road, Bangalore - 560008, Karnataka, India.
shantipriyap@huawei.com

[2]Department of Systems Engineering, Ajou University,
San 5, Woncheon-dong, Yeongtong-gu, suwon 443-749, South Korea.
satchi@ajou.ac.kr

[3]Department of Industrial Engineering, Ajou University,
San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, South Korea.
gnwang@ajou.ac.kr

Keywords: functional Magnetic Resonance Imaging (fMRI), Genetic Algorithms, Decision Tree Ensemble, Cognitive State Classification.

Abstract.*The discovery of fMRI is a major breakthrough in the research of cognitive neuroscience and rise to an outstanding tool for research in this area for cognitive classification and brain activity prediction. Due to the nature of fMRI data, the cognitive classification poses several challenges to neuroscientists and researchers for advancing this arena of research. Although many classification techniques applied for fMRI data analysis none of the techniques could consistently outperform well for all datasets which raised the focus and potential of hybrid techniques. This paper proposes a novel hybrid method, which is a combination of decision tree ensemble and genetic algorithms for cognitive classification. The decision tree ensemble opted due to its fastness, accuracy, and interpretability in ensemble context. The genetic algorithms which significantly applying on fMRI analysis has shown substantial improvement over random and local method is chosen for selecting the best subset of feature. Moreover, this paper explained the decision tree ensemble and genetic algorithms, their usefulness with respect to fMRI analysis and discussed their performance. It also suggests the future perspective and other cognitive task where such technique can be used for feature selection and classification.*

1 INTRODUCTION

Recent advances in human neuroimaging have shown that it is possible to accurately decode a person"s conscious based only on non-invasive measurements of their activity [1]. Neuro-dynamic functional brain imaging methods offer unique insight into the human brain processes by capturing neuronal activity in real time [2]. The fMRI technique is based on the observation that activated brain regions show a change in blood supply as well as blood oxygenation i.e. increases in neural activity indicated by localized increases in blood oxygenation [3-5].The fMRI experiments generates vast quantity of data which are mostly rich and complex in nature and handled by computer based techniques. In fMRI analysis the voxels which have more information about the mental task are using as features and the analysis finds the voxels in a particular region of interest (ROI).

In biomedical studies huge number of fMRI data collected for individual subject and there are hundreds of subjects involved [6]. Each fMRI image contain hundreds of thousands of voxels results over-fitting of classification algorithm i.e., the classifier can work properly on training data but fails on new data [7]. The feature selection plays the crucial role in classification by searching for the best features subset from the original features based on the evaluation criteria [8]. It is vital in cognitive classification as it is computationally infeasible to use all the available features [13].

1.1 Genetic Algorithm

Among different feature selection techniques, the genetic algorithms (GAs) are suitable to solve optimization problem in a high dimensional space [14].The GA is an adaptive global-optimization searching method which is based on Darwin"s fittest principle, which tells an initial population of individuals evolves through the process of natural selection in such a way that the fittest individual has a higher chance of survival. The fitness is the deciding factor which decides whether a chromosome is good or not in a population. GA produces next generation from the current generation using three operators: *reproduction*, *crossover*, and *mutation*. The highly fittest chromosomes move to the next generation. The classification accuracy is returned as a measure of the quality of the transformation matrix, which is used by GA to searches transformation that minimizes the dimensionality of the transformed patterns and maximizes classification accuracy [8].

1.2 Decision Tree

The decision tree (DT) classification algorithm is the most commonly used due to its easy implementation and understanding compared to other classification algorithms [16].

The node of decision tree contains either a single feature f (monolithic trees) or a function of multiple features $f(.)$ (polythetic trees). When an entropy-based splitting criteria is used, the goal is to find cut points of f which minimise the conditional entry on class labels $c = \{1, \dots, C\}$ attached to points in the corresponding sub domains of the discretized variable f'. The entropy of the dataset partitioned by feature f' represented as:

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

whereP_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 those belongs to class c and have value j for feature f'.

The goal of the decision tree growing is to minimise the Eq. (1), which involves recursively selecting features (or discriminant functions) and computing the result of applying different cut-points to them. The edge weight put zero by the feature selection are never included as feature. The learning procedure is divided into two phases i.e., growing phase and pruning phase. In case of growing phase, at each decision node either a single feature of a discriminant function based on linear combination of features is used whereas in the pruning phase, the functional leaves containing the discriminant functions replaced by simple function that predicts the class value [1].

1.3 Classifier Ensemble

In recent years the ensemble technique gaining popularity due to its accuracy, where multiple component learners are trained for a same task and predictions of the component learners are combined to deal with future instances [15]. The classifier ensemble has significant role in classification because it has been observed that classifiers which are not very accurate in classification individually and tend to make mistake on different objects may form a very accurate ensemble [9].

The construction of an ensemble performed using the most popular techniques such as "Bagging" and "Boosting". In case of 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 drawing m examples uniformly (with replacement) from S[17]. It is effective on unstable learning algorithms like decision tree where small changes in the training set resulting large changes in the predictions [18].The ensemble method is divided into two steps, where first step is to imparting training multiple component learners and then combining their predictions [15].

The paper is organized as follows. In Section 2, we describe the related works carried out on fMRI data analysis using genetic algorithms and decision tree ensemble. In Section 3, we depict the overall framework of the proposed technique and our approach for using GA and Ensemble DT. In Section 4, we detail the data preparation and experimental setup. In Section 5, we describe the experimental result and the comparison study.

2 RELATED WORK

In recent years the hybrid techniques gained popularity in cognitive classification due to the complexities involved in fMRI data analysis and limitations of independent classifier. A handful of research works available considering applying Decision Trees and GA as hybrid technique for cognitive classification which are reviewed in this section.

Boehm et al. [11], have demonstrated the effectiveness of genetic algorithm as feature selection technique for classifying cognitive states of brain activities. In their experiment four subjects inside a MRI scanner were passively watching images belonging to five different semantic categories such as: human face, houses, patterns, objects, blank image. The experimental result shown that genetic algorithm together with the one-class neural network compression network can be used to find appropriate features that, on the one hand, increase the accuracy of the classification to close to that obtainable from two class method.

To overcome the problem with single accurate decision tree due to small sample size and large feature dimensionality, Kuncheva et al. [9], have applied the ensemble of decision tree on fMRI data for classification task.

Richiardi et al. [12], have showed the efficiency of ensemble of decision tree in brain decoding which result high-end classification performance with interpretability of results.

3   PROPOSED TECHNIQUE

The block diagram of the proposed technique is shown in the Figure 1.In the proposed technique, the processed fMRI data (.mat) file is supplying as an input to GA for selecting the most promising features from the high dimensional dataset (detail is discussed in Subsection 3.1). The selected features are used to construct a classifier based on ensemble of DTs for classification (c.f., Subsection 3.2). The classification accuracy of the constructed classifier is evaluated using the confusion matrix based on the classification result obtained by test data.
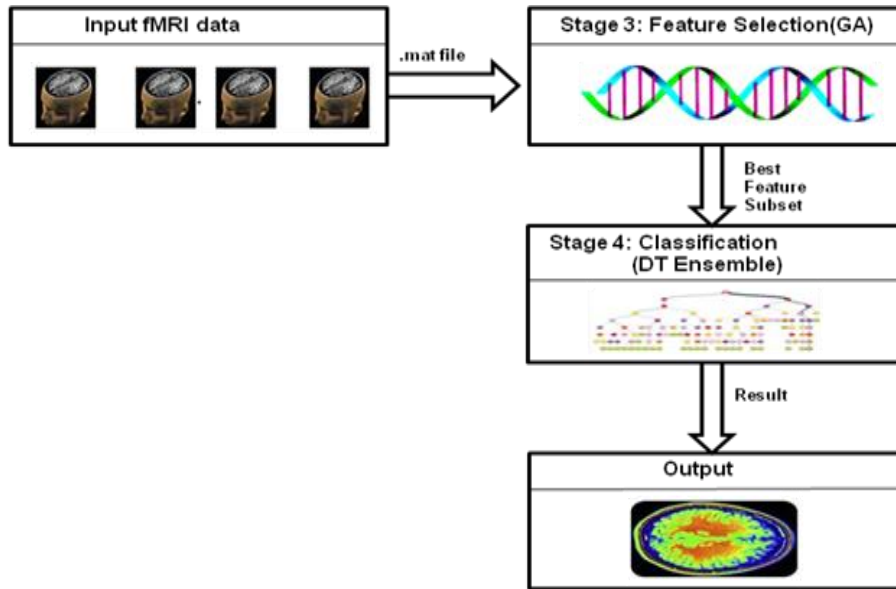
Figure 1: Block diagram of the proposed technique.

3.1 Feature Selection Using Genetic Algorithms

The design of population, fitness function and configuration parameters of the proposed method are explained as below.

The initial population generated by populating a matrix with dimension of population size rows by independent variable (genome length) columns. The values in this matrix are integers which are randomly selected from the processed input data based on ranking, as show in Figure 2.
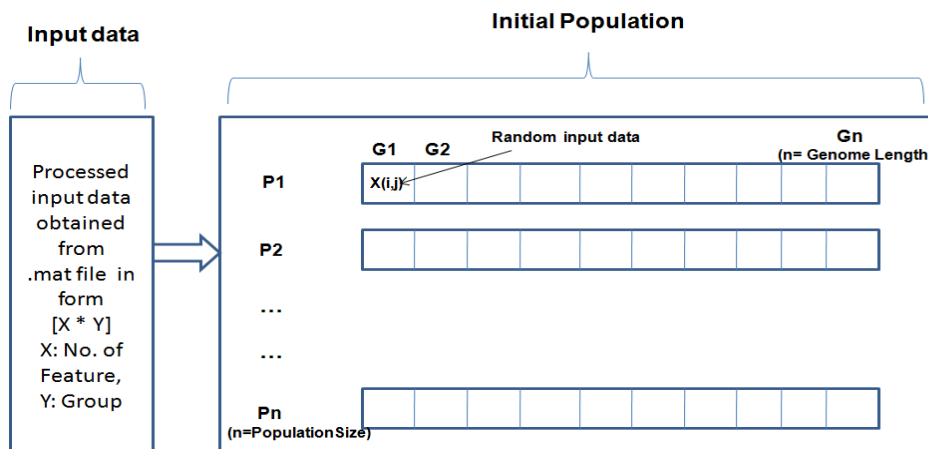
Figure 2: Initial Population Matrix.

4

The fitness of the population is estimated using the fitness function as shown in Figure 3. We have used the fitness function provided by the GA Toolbox, which maximizes the separability of two classes using linear combination of the posterior probability and empirical error rate of linear classifier(classify).
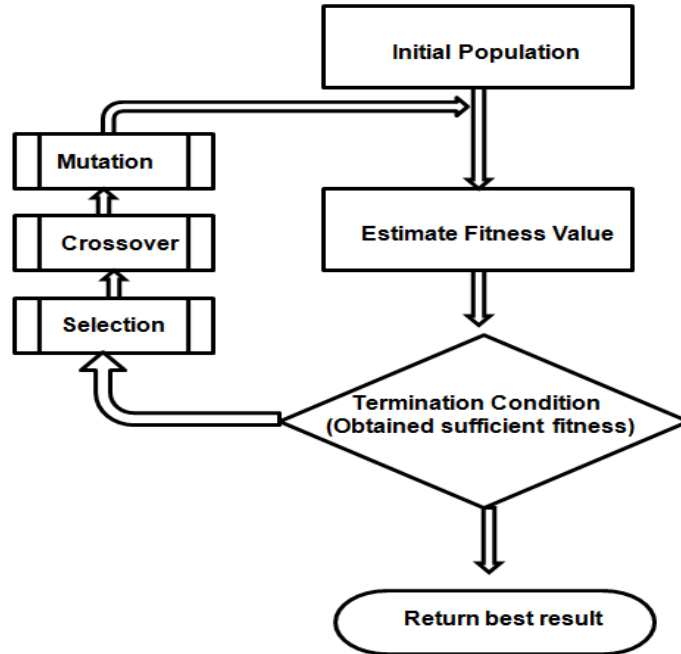

Figure 3: Genetic Algorithm based feature selection.

We have chosen the Roulette wheel selection method where 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 [21-22]. The resulting population is divided into two parts where each pair gone through the crossover (single point, two point and arithmetic) with certain probability P_c [23]. The two point crossover selected, where two points to be selected on the parent strings and all between the two points is swapped between parents to generate child strings. The mutation genetic operator randomly alters the values of genes in parent string. At each step of evolution, the crossover and mutation applied stochastically so their probabilities of occurrence must be set [24]. The uniform mutation selected as the mutation function.

The genetic algorithm parameters configured as shown in Table 1.

Genetic Algorithm Parameter	Value
Population Size	48
Number of Generation	100
Selection function	Roulette
Crossover function	Two point crossover
Mutation function	0.01

Table 1: GA Parameters.

Based on the above initial population and parameters configuration, the input passed to GA toolbox provided genetic algorithm function which returns the best features. The genetic algorithm function runs multiple times to obtain the best set of features which can contribute significantly in subsequent DT ensemble classification.

3.2 Decision Tree Ensemble

In order to overcome the issue with over-fit of learning using single decision tree, the ensemble of decision tree is popularly adopted [19]. The decision tree ensembles are very popular ensembles because decision trees are unstable classifier where the output undergoes significant change in response to small changes in training data. The ensemble of decision tree generally performs better compared to single decision tree due to its representational power. However, it can show poor results on data if it contains many irrelevant attributes [20].

We have used the ensemble of decision trees for classifying the true class labels. The classification result obtained by the voting of the trees. The best features obtained by the GA, input to the decision tree ensemble in the matrix form of [X *X* Y], where X denotes the observations i.e., data of the best features and Y denotes the class labels for the observations. We have selected the number of trees in the range of 200-500 for GA with ensemble DT and the number of trees reduced for ensemble DT without GA due to long execution time because of large sample size [4698 X 2196].

The dataset is divided into two almost equal parts (50-50) where the even samples formed as training dataset and odd samples formed as test dataset as shown in Table 2. The training data is used for training decision trees and the test data is used for predicting the classification accuracy. It estimates the out-of-bag classification error and then comparing with the number of trees grown as shown in Figure 4. Based on the prediction scores, the confusion matrix is made. The classification accuracy is calculated using the confusion matrix. .

4   DATA PREPARATION AND EXPERIMENTAL STUDY

The experiment carried out on the platform of 32 bit, Intel 2.70 GHz processor, 4.00 GB RAM, running under Windows 7 operating system. The programs for the experiment coded using the Matlab R2010a (The Mathworks©). The Matlab provided GA Toolbox functions used for feature selection.

The fMRI data set is collected from the Carnegie Mellon University (CMU)‟s public data repository [10]. The data is taken for a single subject („04847‟) and is partitioned into trials. The experiment consists of set of trials. For some of these intervals, the subject simply rested, or gazed at a fixation point on the screen. In other trials, the subject has shown a picture and a sentence, and instructed to press a button to indicate whether the sentence correctly described the picture. For these trials, the sentence and picture were presented in sequence, with the picture presented first on half of the trials, and the sentence presented first on the other half of the trials. Forty such trials are available the subject. The images were collected every 500 msec. There are 54 trials, 2800 snapshots. The data is stored in a [54X1] cell array with one cell per „trial‟ in the experiment. Each element in the cell array is an NXV array of observed fMRI activations and each array contain 4698 number of voxels (features) per snapshot. The genetic algorithm applied to reduce the number of features (V) of each of NXV array. During the initial population generation we have ignored the Cond=0 which indicates data to ignored and Cond=1 indicates, segment is a rest or fixation interval. The dataset containing total number of features and class label applied to genetic algorithm is [4698 X 2196]. The number of samples used for training and testing is shown in Table 2.

Algorithm	Train	Test
Ensemble DT	2349 (50%)	2349 (50%)
GA + Ensemble DT	210 (50%)	210 (50%)

Table 2 : Dataset for Training and Testing.

6

5   RESULTS

The generation vs. fitness value of 51 generation and the number of trees grown along with the out-of-bag classification error are shown in the Figure 4. The number of trees is selected for ensemble DT was 200. The GA reduces the number of features from 4698 to 420. The important features for Ensemble DT with GA are shown in the Figure 5.
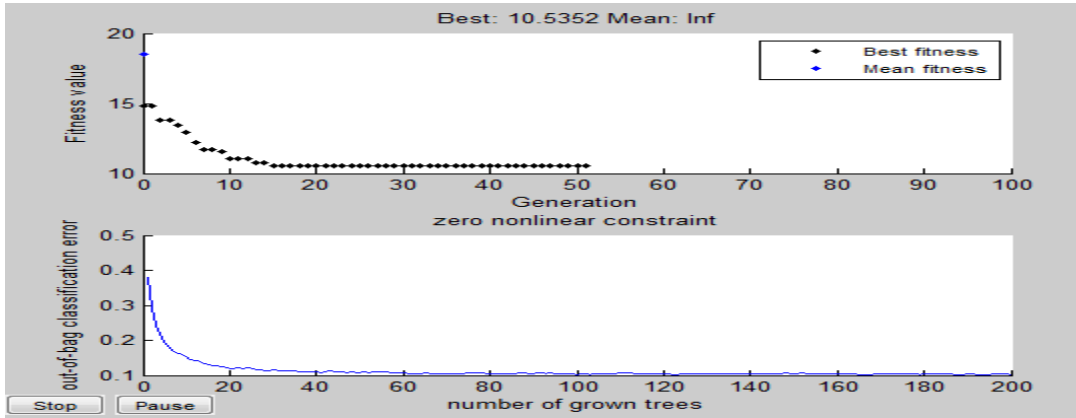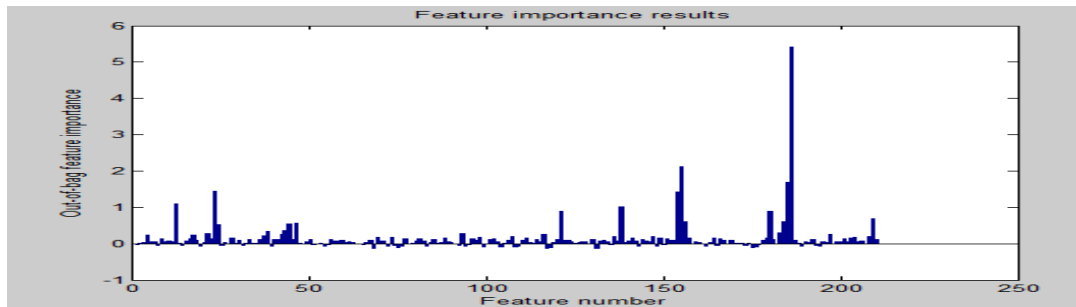

Figure 4: GA fitness value with generation.

Figure 5: Important Features.

5.1 Classification Accuracy

The confusion matrix of the proposed method on training data is listed in Table 3. The element of ith row and j^{th} column denotes the classification accuracy belonging to class i are assigned to class j after the classification. The class „Picture" denotes, pictures before sentences and the class „S" denotes, sentences before picture.

		Predicted	
		Picture	Sentence
Ensemble DT	Picture	820(TP)	280(FN)
	Sentence	614(FP)	482(TN)
GA + Ensemble DT	Picture	886 (TP)	214 (FN)
	Sentence	603 (FP)	493 (TN)

Table 3: Confusion matrix

The performance measurement has been estimated using the following measures:

7

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} * 100\%, \tag{2}$$

$$Sensitivity = \frac{(TP)}{(TP+FN)} * 100\%, \tag{3}$$

$$Specificity = \frac{(TN)}{(TN+FP)} * 100\%, \tag{4}$$

Where,

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

Algorithm	Sensitivity	Specificity	Accuracy
Ensemble DT	74.54	43.97	59.28
GA + Ensemble DT	80.54	44.98	62.79

Table 4: Performance Value.

5.2 Comparison Result

We have compared the result in terms of number of features vs. accuracy for ensemble DT without GA with ensemble DT with GA as shown in Table 5.The computation time of the ensemble DT without GA and ensemble DT with GA is shown in the Table6.

Algorithm	Number of Features	Accuracy
Ensemble DT	4698	59.28
GA + Ensemble DT	420	62.79

Table 5: Comparison Matrix (No. of Features vs. Accuracy).

Algorithm	Sample Size(features and class labels)	Computation Time (second)
Ensemble DT	[4698 X 2196]	634.299
GA + Ensemble DT	[420 X 2196]	55.601093

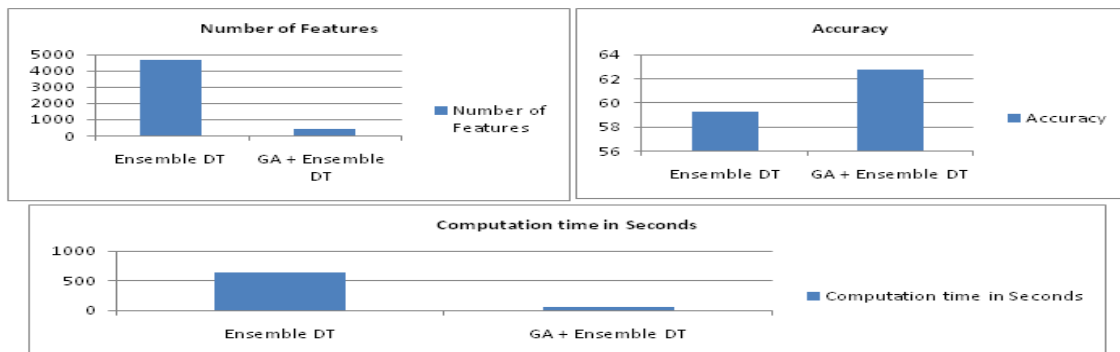Table 6: Comparison Matrix (No. of features Vs Computation Time).

Figure 6: Comparison Graphs.

8

6   CONCLUSION AND FUTURE PERSPECTIVES

In this paper, we have developed a hybrid ensemble technique using genetic algorithm and decision tree for fMRI data classification. The experimental result shown in Table 3, comparison matrix in Tables 4, 5, and 6, highlights the usefulness of the proposed technique in fMRI classification. The key differentiators of the proposed technique are: i) the significant reduction in number of features compared to original data set which is effective in fMRI based data analysis; ii) the faster execution time emphasizes the importance of such techniques to explore further.

The directions for future enhancement of the proposed technique are: i) improvement in performance in terms of accuracy; ii) analyse the result for multiple subject and different cognitive task; iii) another potential area can be explored, creating discriminative graph for the selected features by GA and apply to decision tree ensemble for classification

The fMRI experiments involves different cognitive activities and various instantaneous states resulting high dimensional dataset, therefore, it demands technique that can reduce number of features and to minimize computation cost without compromising with prediction accuracy.

REFERENCES

[1]   J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, D. V. D. Ville, Decoding brain states from fMRI connectivity graphs. *NeuroImage*, 56, 2, 616–626, 2011.

[2]   J.D. Haynes, G. Rees, Decoding mental states from brain activity in humans. *Nature Review Neuroscience*, 7, 7, 523–534, 2006.

[3]   S. Supek, R. Magjarevic, Neurodynamicmeasures of functional connectivity and cognition. *Medical and Biological Engineering*, 49, 507–509, 2011.

[4]   J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, P. Pietrini, Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2429, 2001.

[5]   C. Chen, C. W. Tyler, H. A. Baseler, Statistical properties of BOLD magnetic resonance activity in the human brain. *NeuroImage*, 20, 2, 1096–1109, 2003.

[6]   J. Fan, R. Li, Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery. *In International Congress of Mathematicians*, III, 595–622, 2006.

[7]   B. Jin, A. Strasburger, S.J. Laken, F.A. Kozel, K.A. Johnson, M.S. George, X. Lu,Feature Selection for fMRI-based Deception Detection. *BMC Bioinformatics*, 10, suppl. 9, S15, 2009.

[8]   A. Kharrat, A Gasmi, K.Messaoud, M.B. Benamrane, N. Abid, M, A hybrid approach for automatic classification of brain MRI using genetic algorithm and support vector machine. *Leonardo Journal of Science*, 9, 17, 71–82, 2010.

[9]   L. I. Kuncheva, J. J. Rodriguez, Classifier Ensembles for fMRI data Analysis: An Experiment. *Magnetic Resonance Imaging*, 28, 4, 583–593, 2010.

[10] http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/

[11] O. Boehm, D. R. Hardoon, L. M. Manevitz, Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms. *International Journal of Machine Learning and Cybernetics*, 2, 3, 125–134, 2011.

[12] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, D. V. D. Ville, Brain decoding of fmri connectivity graphs using decision tree ensembles. *In Proc. IEEE Int. Symposium on Biomedical Imaging,* 1137–1140, 2010.

[13] R. Kohavi, G. H. John, Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97, 1–2, 1997.

[14] D. Cordes, G. Herzmann, R. Nandy, T. Curran, Optimization of contrast detection power with probabilistic behavioural information. *NeuroImage*, 60, 3, 1788–1799, 2012.

[15] Z.H. Zhou, W.Tang, Z.H. Zhou, W. Tang, Selective Ensemble of Decision Trees. *Lecture Notes in Artificial Intelligence*, 467–483, 2003.

[16] M.N. Anyanwu, S.G. Shiva, Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security*, 3, 3, 230–240, 2009.

[17] T.G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, Boosting, and Randomization. *Machine Learning*, 40, 2, 139–157, 2000.

[18] D. Opitz, R. Maclin, Popular ensemble methods: an empirical study. *Journal of Artificial Intelligence Research*, 11, 169–198, 1999.

[19] J. Paul, M. Verleysen, P. Dupont, The stability of feature selection and class prediction from ensemble tree classifiers. *ESANN*, 263–268, 2012.

[20] M. Gashler, C. G. Carrier, T. Mortinez, Decision tree ensembles: small heterogeneous is better than large homogeneous. *ICMLA*, 900–905, 2008.

[21] M. K. Rafsanjani, S. Eskandari, The effect of a new generation based sequential selection operator on the performance of genetic algorithm. *Indian Journal of Science and Technology*, 5, 12, 2012.

[22] D. Sharma, V. Singh, C. Sharma, GA based scheduling of FMS using Roulette Wheel selection process. *SocProS 2011*, 131,931–940, 2012.

[23] S. Sefiance, M. Benbouziane, Portfolio selection using genetic algorithm. *Journal of Applied Finance & Banking*, 2, 4, 143–154, 2012.

[24] L. Scrucca, GA: a package for genetic algorithms in R. *Journal of Statistical Software*, 53, 4, 2013.