Cognitive Computation (2019) 11:545–559
https://doi.org/10.1007/s12559-019-09636-0

Multi-Region Risk-Sensitive Cognitive Ensembler for Accurate Detection of Attention-Deficit/Hyperactivity Disorder

Vasily Sachnev[1] · Sundaram Suresh[2] · Narasimman Sundararajan[2] · Belathur Suresh Mahanand[3] · Muhammad W. Azeem[4] · Saras Saraswathi[5]

Received: 25 December 2017 / Accepted: 10 March 2019 / Published online: 30 March 2019

Abstract
In this paper, we present a multi-region ensemble classifier approach (MRECA) using a cognitive ensemble of classifiers for accurate identification of attention-deficit/hyperactivity disorder (ADHD) subjects. This approach is developed using the features extracted from the structural MRIs of three different developing brain regions, viz., the amygdala, caudate, and hippocampus. For this study, the structural magnetic resonance imaging (sMRI) data provided by the ADHD-200 consortium has been used to identify the following three classes of ADHD, viz., ADHD-combined, ADHD-inattentive, and the TDC (typically developing control). From the sMRIs of the amygdala, caudate, and hippocampus regions of the brain from the ADHD-200 data, multiple feature sets were obtained using a feature-selecting genetic algorithm (FSGA), in a wraparound approach using an extreme learning machine (ELM) basic classifier. An improved crossover operator for the FSGA has been developed for obtaining higher accuracies compared with other existing crossover operators. From the multiple feature sets and the corresponding ELM classifiers, a classifier-selecting genetic algorithm (CSGA) has been developed to identify the top performing feature sets and their ELM classifiers. These classifiers are then combined using a risk-sensitive hinge loss function to form a risk-sensitive cognitive ensemble classifier resulting in a simultaneous multiclass classification of ADHD with higher accuracies. Performance evaluation of the multi-region ensemble classifier is presented under the following three scenarios, viz., region-based individual (best) classifier, region-based ensemble classifier, and finally a multiple-region-based ensemble classifier. The study results clearly indicate that the proposed "multi-region ensemble classification approach" (MRECA) achieves a much higher classification accuracy of ADHD data (normally a difficult problem because of the variations in the data) compared with other existing methods.

Keywords Attention-deficit/hyperactivity disorder (ADHD) · Neuroimaging structural magnetic resonance imaging (sMRI) · Extreme learning machine (ELM) · Genetic algorithm (GA)

Introduction

Attention-deficit/hyperactivity disorder (ADHD) [1] is an ongoing pattern of neuro-developmental disorder that is both pervasive and prevalent. The symptoms include inattention, hyperactivity, and impulsivity that are not consistent with a child's developmental age. According to Children and Adults with Attention-Deficit/Hyperactivity Disorder (CHADD), a non-profit organization that provides support and education for those who are affected, ADHD affects 11% of school-age children in the USA (around 5.1 million), resulting in severe functional impairments such as failure in school, disruption in family, stress, and failed relationships [2]. These disorders often persist into adulthood for 75% of these patients [3, 4] who are subjected to increased risks of depression, antisocial behavior, delinquency, and employment issues and are prone to accidents and substance abuse. The economic cost to individuals, families, and society due to ADHD patients ranges between 143 and 266$ billion per year, incurred mainly for education, health care, loss of income, and productivity [5].

In the current clinical view, ADHD is classified into three behavior-based subtypes, namely, hyperactive-impulsive type

✉ Vasily Sachnev
bassvasys@hotmail.com

[1] Department of Information, Communication and Electronics Engineering, Catholic University of Korea, Seoul, South Korea

[2] School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore, Singapore

[3] Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, India

[4] Psychiatry Division, Sidra Medicine, Doha, Qatar

[5] Department of Children's Clinical Management Group, Sidra Medicine, Doha, Qatar

(ADHD-H), hyperactive-inattentive type (ADHD-I), and the combined type (ADHD-C) [6, 7]. The hyperactive-impulsive type is an early-onset disorder that occurs in the nursery or primary developmental age that mainly affects behavior and is present three times more in boys as compared with girls. The ADHD-inattentive type occurs later, at primary or high school age, and affects boys and girls equally. The combined type is a combination of both the inattentive and hyperactive-impulsive types [6, 7]. These diagnostic categories are based on symptomatic behaviors and are not indicative of the underlying disorders that might have physiological origins. Although ADHD studies have traditionally been based on clinical symptoms, more recent studies indicate that structural and functional abnormalities in the brain (assessed by structural and functional magnetic resonance imaging, sMRI, and fMRI) may appear due to ADHD.

Recent studies show that the abnormal volumes observed in certain brain regions can be indicative of ADHD disorder and can help with the early evaluation of neuropsychiatric disorders and early diagnosis of ADHD [8, 9]. Developmental delays and structural alterations in the brain regions of interest (ROIs) such as the amygdala, caudate, and hippocampus (hereafter referred to as ROI_3) are thought to be affected in ADHD children and youth [10–14]. These regions are responsible for attention, cognition, emotion, and sensorimotor functions. Specifically, the amygdala plays a primary role in the processing of memory, decision-making, and emotional reactions [11, 12]. The caudate nucleus plays an important role in various motor and non-motor functions such as procedural learning, associative learning, inhibitory control of action, and reward system [13]. The hippocampus plays an important role in the consolidation of information from short-term memory, long-term memory, and spatial memory [14]. Since research shows that both the intra-region variability and sub-regional variability in the brain contribute to the abnormalities in the brain of ADHD patients, they have to be captured and studied simultaneously [10–14].

In order to study the above, in recent years, neuroscientists are using computational methods for analyzing brain images obtained using sMRI. These methods may help to obtain valuable insights into the pathological mechanisms related to ADHD and provide possible target locations for an early intervention. Morphometric analysis of brain regions using anatomic MRI data indicate that ADHD subjects have differences in volume in the right globus pallidus, anterior frontal regions, and cerebellum and they exhibit loss of normal asymmetry in the caudate [9, 15–19]. Extra studies focus on the similarities between Parkinson's disease and ADHD [20] and the connectivity in between electroencephalogram and MRI for ADHD [21].

A brief review of ADHD detection using structural image-based models that were developed using sMRI data is given below. These models help us to classify ADHD subtypes and differentiate patients from the normal subjects, referred to as typically developing children (TDC) [22, 23].

Chang et al. [24] have used structural image-based models to distinguish ADHD patients from typically developing children (TDC). They used isotropic local binary patterns on three orthogonal planes (LBP-TOP) to extract features and have used support vector machines (SVM) to classify the ADHD types. Kobel et al. [25] have applied structural imaging techniques such as voxel-based morphometry (VBM), diffusion tensor imaging (DTI), and magnetization transfer imaging (MTI) that showed anatomical differences in brain regions of ADHD patients. Cao et al. [26] implemented a graphical analysis of large-scale brain systems for searching biomarkers responsible for ADHD using structural MRI. Rangarajan et al. [27] discovered potential biomarkers in the hippocampus area using a projection-based learning algorithm for a meta-cognitive radial basis function network (PBL-McRBFN) for ADHD vs. TDC diagnosis. Sabuncu [28] exploited multivariate pattern analysis (MVPA) for the diagnosis of Alzheimer's, schizophrenia, autism, and ADHD using structural MRI. Iannaccone et al. [23] proposed an ADHD vs. TDC classifier based on SVM implemented for structural MRI. Mahanand et al. [29] have proposed a meta-cognitive radial basis function network (PBL-McRBFN) for the diagnosis of ADHD and its subtypes. Sato et al. [30] used three feature extraction techniques and ten different pattern recognition techniques for the diagnosis of ADHD vs. TDC and ADHD-I vs. combined ADHD. Eloyan et al. [31] exploited singular value decompositions (SVDs), CUR decompositions, random forest, gradient boosting, bagging, voxel-based morphometry, and support vector machines (SVM) for ADHD vs. TDC diagnosis. Colby et al. [32] proposed multiple support vector machine recursive feature elimination (SVM-RFE) algorithm and feature subset selection by optimizing the expected generalization performance of a radial basis function kernel SVM (RBF-SVM) for ADHD vs. TDC by using structural MRI. Peng et al. [33] exploited the extreme learning machine (ELM) with F score and SFS feature selection methods for ADHD vs. TDC diagnosis. Sachnev et al. [34] used a binary-coded genetic algorithm (BCGA) with Meta-Cognitive Neuro-Fuzzy Inference System (McFIS) in the hippocampal area for classifying TDC and 3 subtypes of ADHD.

Based on the review of the above existing studies on ADHD, it may be seen that many studies have been conducted in classifying ADHD patients vs. normal persons (TDC). It is seen that such problems have been treated mainly as a two-class classification problem. Only a few studies exist in the literature that focusses on classifying TDC and three subtypes of ADHD (ADHD-H, ADHD-I, and ADHD-C). The problem of classifying ADHD subtypes is much more complex in nature and requires more sophisticated machine learning tools to build an efficient classifier. MRI data, available for the classification of ADHD and its subtypes, is sparse and unbalanced, making the problem more difficult for the machine learning methods.

To overcome the above difficulties, in this paper, we present a multiple-region ensemble classification approach (MRECA) for an accurate diagnosis of ADHD and its subtypes using the sMRI of the three most affected brain regions viz., the amygdala, the caudate, and the hippocampus. Regions of interest (ROIs) for these regions are defined using the Wake Forest University Pick-atlas [35]. MRECA efficiently classifies the TDC and its two subtypes, ADHD-I and ADHD-C (ADHD-H type has been excluded in this study due to the availability of only a small number of samples). A specially tailored "feature-selecting genetic algorithm (FSGA)," based on binary selections procedures [36], is used for the feature selection process. FSGA is then coupled with an ELM classifier in a wraparound scheme for identifying the TDC and two ADHD subtypes ADHD-C and ADHD-I. ELM has been used to evaluate the importance of the chosen features. Better features would create a better ELM classifier that can result in higher training and testing accuracies. In this study, extreme learning machine was chosen as the base learner because of its fast training (learning) scheme compared with other existing machine learning classifiers [37–40]. Each ELM classifier that is obtained using the unique feature set chosen by the FSGA is referred to as the "region-based individual classifier."

The proposed FSGA-ELM generates three sets of ELM classifiers, viz., HipC, AmyC, and CauC for the hippocampus, amygdala, and caudate regions, respectively. In the next step, a risk-sensitive cognitive ensemble classifier is used to select a set of ELM classifiers with higher classification accuracies to construct an ensembler, which is a fusion of the selected ELM classifiers. Such a selection is processed using the HipC, AmyC, and CauC separately to build what is referred to as the *region-based cognitive ensemble classifiers*: hippocampus ensemble classifier HipEC = ξ (HipC), the amygdala ensemble classifier AmyEC = ξ (AmyC), and the caudate ensemble classifier CauEC = ξ (CauC). The risk-sensitive cognitive ensemble classifier is denoted with the symbol ξ (see Fig. 1). The union of several ELM classifiers utilizes the benefits of multiple sets of features and covers more relevant data needed for an accurate classification of ADHD.

Selection of ELM classifiers for the ensemble is made by another classifier-selecting genetic algorithm (CSGA); the fusion is accomplished by a modified version of the earlier ensemble approach proposed in Sachnev [41] using a risk-sensitive hinge loss function. In this paper, the risk-sensitive cognitive ensemble classifier has been developed under two different scenarios. They are as follows: (1) to build a *region-based ensemble classifier* and (2) to build a *multiple-region-based ensemble classifier*. The *Multiple-region ensemble classifier* approach or MRECA processes all ELM classifiers from HipC, AmyC, and CauC and builds a fusion of all the ELM classifiers from the three regions (see Fig. 1). Detailed performance evaluation of MRECA clearly shows its superior performance compared with existing methods that are currently available for ADHD diagnosis.

This paper is organized as follows. The "Data Set Description" section gives the details of the ADHD-200 dataset used in this study. The "Method" section presents the "multiple-region ensemble classifier approach for diagnosis TDC, ADHD-C, and ADHD-I." It contains the description of ROI (region of interest)–based feature extraction, feature-selecting genetic algorithm (FSGA), classifier-selecting genetic algorithm (CSGA), and the risk-sensitive cognitive

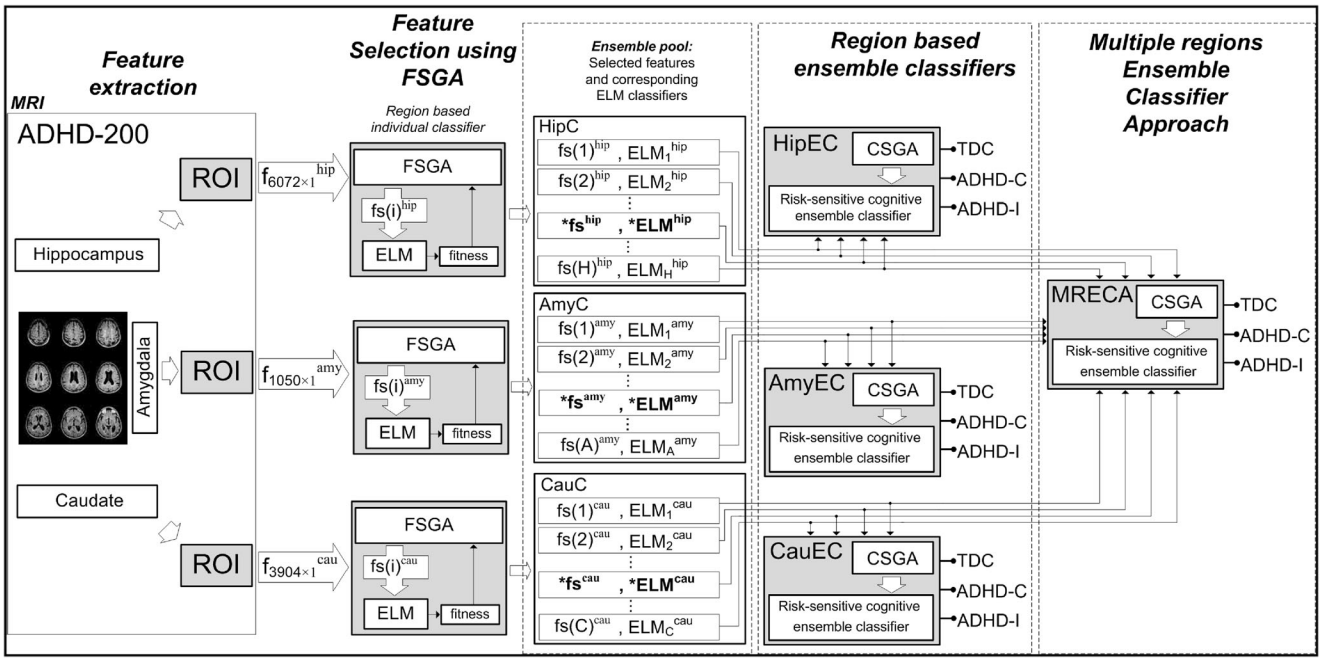

Fig. 1 A framework for the multiple-region ensemble classification approach (MRECA)

ensemble of ELM Classifiers. The "Results" section presents the results of the MRECA performance evaluation. It includes the performance results for (1) the region-based individual classifier, (2) the region-based ensemble classifier and, (3) the multiple-region ensemble classifier. These results have been compared with existing methods. A preliminary approach to search for voxels responsible for ADHD is presented in this section. The "Conclusions" section summarizes the conclusions from this study.

Data Set Description

ADHD-200 Consortium is a collaborative effort to get a better understanding of the neural basis of ADHD and it invites ideas in neuroscience that are related to ADHD diagnosis [42], (http://fcon1000.projects.nitrc.org/indi/adhd200/results.html), [19]. It provides a single source of functional MRI (fMRI) and sMRI data from ADHD patients and normal TDC subjects. This data helps one to investigate multiple brain regions and to understand the relationships underlying the pathophysiology of ADHD. sMRI has a high spatial resolution and can detect minute abnormalities in the shapes and sizes of the brain regions. Brain functions are highly dependent on the integrity of the brain structures. Morphometric techniques provide data on the volume or shape of gray matter structures in ADHD-affected brain regions. Research based on structural information can help to study the brain morphological differences in ADHD compared with typically developing children. Hence, in this study, we use ADHD-200 sMRI data to implement our machine learning algorithm in classifying ADHD under three classes, viz., TDC, ADHD-C, and ADHD-I.

In this study, sMRI images were used to extract 6072, 1050, and 3904 features (voxels, ROI_3), from the hippocampus, amygdala, and caudate, respectively (see the "Multi-Region Ensemble Classifier Approach for Diagnosis of ADHD-C and ADHD-I" section for method of extraction). The ADHD-200 dataset includes 776 subjects in the training set and 197 in the test set. Among the 776 in the training set, only 770 subjects were used in this study and the rest were excluded since they did not pass the quality check. Among the 197 in the test set, only 171 subjects were used in this study as the remaining 26 subjects were from Brown University, which never released their labels.

Thus, the ADHD-200 dataset contains sMRI of 941 samples consisting of 581 TDC subjects and 360 ADHD subjects. Among the ADHD subjects, 13 are ADHD-H, 137 are ADHD-I, and 210 are ADHD-C. It is to be noted that there are a huge imbalance and sparsity between the number of samples available for TDC and each of the ADHD subtypes (especially ADHD-H), making accurate classification very difficult. Hence, we excluded the ADHD-H subtypes from this study and used only TDC, ADHD-I, and ADHD-C subtypes to obtain 759 training samples and 169 testing samples.

Method

Multi-Region Ensemble Classifier Approach for Diagnosis of ADHD-C and ADHD-I

Most of the current ADHD studies solve the following two-class problem viz., classifying whether a person is normal (TDC) or is affected by ADHD. Very few studies have focused on the more complex three-class problem of simultaneously classifying the two ADHD subtypes, ADHD-C and ADHD-I and TDC. With this objective, we present the multi-region ensemble classifier approach named as MRECA, a machine learning algorithm that can simultaneously and efficiently solve the abovementioned three-class classification problem. MRECA broadly consists of four major steps as shown in Fig. 1 and explained below.

- Step 1: Region of interest–based feature extraction. In this step, using voxel-based morphometry and sMRI images, we obtain the corresponding sets of features referred to as f^{hip}, f^{amy}, and f^{cau}, from the hippocampus, amygdala, and caudate regions, respectively.
- Step 2: Feature selection. From the features obtained in step 1, we select a smaller set of features using a specially tailored feature-selecting genetic algorithm (FSGA). FSGA uses an extreme learning machine (ELM) classifier in a wraparound method and the corresponding classifiers are called as HipC, AmyC, and CauC.
- Step 3: In this step, a region-based ensemble classifier is built using a "classifier-selecting genetic algorithm" (CSGA). Using a risk-sensitive hinge loss function and the three sets of features and the corresponding classifiers obtained from step 2, the classifiers are combined to build three risk-sensitive cognitive ensemble classifiers: ξ (HipC), ξ (AmyC), and ξ (CauC). These ensemble classifiers are referred to as region-based ensemble classifiers.
- Step 4: Finally, in a similar fashion, the three region-based risk-sensitive cognitive ensemble classifiers obtained in step 3 are combined to develop the final multiple-region ensemble classifier. These four steps are explained in detail below.

Region of Interest–Based Feature Extraction

For the initial preprocessing of sMRI data, we refer to the Neuro Bureau initiative [36] that provides preprocessed versions of the ADHD-200 data. Among the different pipelines

supported by the Neuro Bureau initiative, we focused on the data computed by the Burner pipeline (http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:BurnerPipeline). The Burner pipeline was managed by Carlton Chu using the Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) toolbox from statistical parametric mapping (SPM). The Burner pipeline includes modulated and normalized gray matter maps. There are three steps in the Burner pipeline. In the first step, structural images were segmented into gray matter and white matter probability maps using SPM8's unified segmentation procedure [43]. Next, gray matter and white matter probability maps are rigidly aligned (translation and rotation) and an inter-subject normalization is performed using the DARTEL approach [44]. Images are then iteratively registered to the group average (population template) and the template is iteratively updated. Lastly, the registration parameters were applied to each gray matter probability map to transform them into the space of population averages. Modulation is then applied to conserve the global tissue volumes after the normalization. The resulting gray matter probability maps were used in our study.

A feature extraction method based on ROI is employed in this study [29]. The amygdala, caudate, and hippocampus ROIs were defined using the Wake Forest University Pickatlas [35]. Gray matter tissue probability values were extracted as features from the probability maps using the generated ROI masks. Spatial information of each feature is preserved and then ordered in the feature vector. A total of 1050, 3904, and 6072 features for the amygdala, caudate, and hippocampus were extracted and used in selecting the ROI_3 features.

Feature Selection Using a Wraparound Approach

In the second step, a specially tailored "feature-selecting genetic algorithm (FSGA)" is used to select the best set of features that produce higher classification accuracies, using an ELM classifier for classifying the subtypes TDC, ADHD-C, and ADHD-I (Fig. 1). *HipC*, *AmyC*, and *CauC* are the sets of all ELM classifiers generated by the FSGA as explained below. These classifiers are referred to as the "region-based individual (best) ELM classifier".

Feature Selection Using FSGA

FSGA is based on the well-known standard genetic algorithm (GA) [45] that is widely used to solve complex feature selection problems. A wrapper method for feature selection is used where ELM evaluates the selected features [46]. Using iterative random search techniques, GA improves the fittest members of a population (solution strings) by using genetic operators such as crossover and mutation. A problem specific fitness function is used to evaluate each string, which in this case is the classification accuracy. GA starts with an initial generation of $p = 200$ random strings. It iteratively updates this set of 200 solutions by using a modified single-point crossover (70% of current solutions) and binary mutation (30%) operators. Each generated solution is then used with an ELM classifier and is evaluated by the fitness g, which in this case is the overall testing accuracy of that ELM classifier (Fig. 2). GA generates new solutions until the maximum number of generations (termination criteria) is reached (300 for FSGA). This set of 200 solutions are processed using a selection procedure based on the well-known geometric ranking method [47], where the population is arranged in descending order of its fitness values (g). The probability of the selected solution j is calculated as follows:

$$Ps_j = q'(1-q)^{(r_j-1)} \quad (1)$$

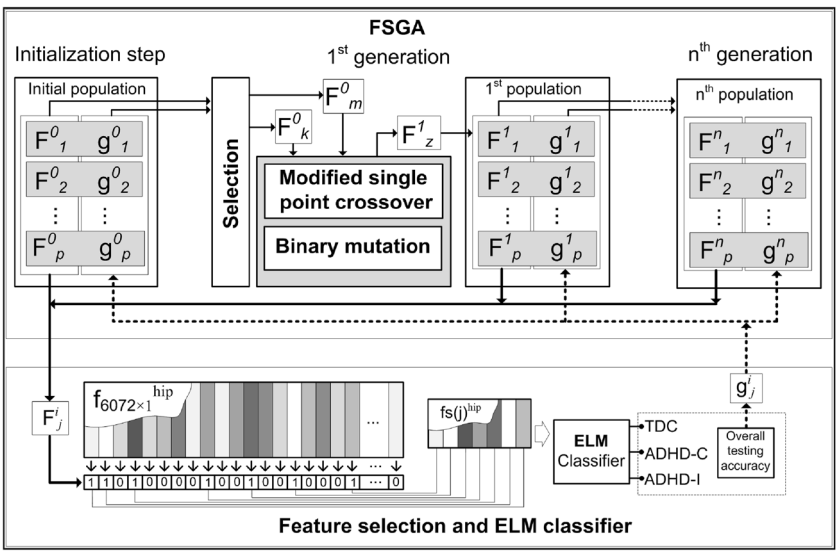

Fig. 2 Feature selection using FSGA coupled with ELM

where $q' = q/(1-(1-q)^p)$ is the selection probability, r_j is the rank of the jth solution in the partially ordered set, and p is the population size (http://neurobureau.projects.nitrc.org/ADHD200/Introduction.html) ($p = 200$ for FSGA and CSGA). In this work, the parameter q is set as 10^{-3} for "FSGA" (see Fig. 1). (For more details, refer to [47].)

FSGA String Representation

The solution string is an array of binary values F where the binary digits represent either the presence or absence of that feature. A value of "1" means that the particular feature is participating and "0" means that feature is absent. A typical string representing 6072 features of the hippocampus is shown in Fig. 2. The choice of the string representation depends on the optimization problem, the GA framework, and the genetic operators used.

Modified Single-Point Crossover Operators

Performance of a genetic algorithm depends heavily based on the particular optimization problem being solved, the given data, and the choice of a proper crossover operator. Choice of crossover operators for binary strings that exist in the literature [45] is very limited. Most of these operators use only simple binary logical operations. In our studies, the binary string should represent a relatively smaller number of chosen features (around 20–200 features from a set of 1050, 3904, and 6072 features for the three regions). Thus, the crossovers based on simple binary logical operations will significantly modify the new solutions and, in some cases, may generate solutions with all zeros, which is unacceptable. As a result, the performance of GA could be significantly affected. To overcome this problem, we developed a modified single-point crossover method to efficiently solve the ADHD feature selection problem (Fig. 3).

The proposed single-point crossover operator uses two solutions from the generation i (parent 1: p_1 and parent 2: p_2) as inputs and produces two new solutions for the next generation $i+1$ (offspring of_1 and of_2). This operator analyzes both the parents p_1 and p_2 and collects all the "common" features as $C = p_1 \cup p_2$ and "non-common" features as $\overline{C} = p_1 \cup p_2 - p_1 \cap p_2$. The "common" features C are then moved to both the offspring solutions without modifications. The "non-common" features $\overline{C}$ are then divided randomly into 2 parts $\overline{c}_1$ and $\overline{c}_2$. The random split is limited to the range of 0.25–0.75.

Finally, the first offspring solution of_1 (Fig. 3) is the combination of the "common" features C and the "non-common" features from "part 1" $\overline{c}_1$ or $of_1 = C \cup \overline{c}_1$; the second offspring solution of_2 is the combination of the "common" features C and the "non-common" features from "part 2" $\overline{c}_2$ or $of_2 = C \cup \overline{c}_2$ (Fig. 3).

Illustrating this with an example, assume that the sets of chosen features from the first and second solution are $p_1 = \{2, 3, 4, 6, 7, 10, 11, 12\}$ and $p_2 = \{1, 2, 3, 5, 6, 9, 10, 12\}$, respectively. The random split is chosen as 0.333. The number of chosen features from the first and second solutions is 8. The "common" features set is $C = p_1 \cup p_2 = \{2, 3, 6, 10, 12\}$; the "non-common" set is $\overline{C} = \{1, 4, 5, 7, 9, 11\}$. The number of elements in "non-common" set is 6. Then, the number of randomly selected features in "part 1" is $floor\ (6 \times 0.333) = 2$; the number of features in "part 2" is 6–2 = 4. The "part 1" set of randomly selected 2 features among 6 from the "non-common" set is $\overline{c}_1 = \{1, 7\}$; "part 2" is a set of remaining

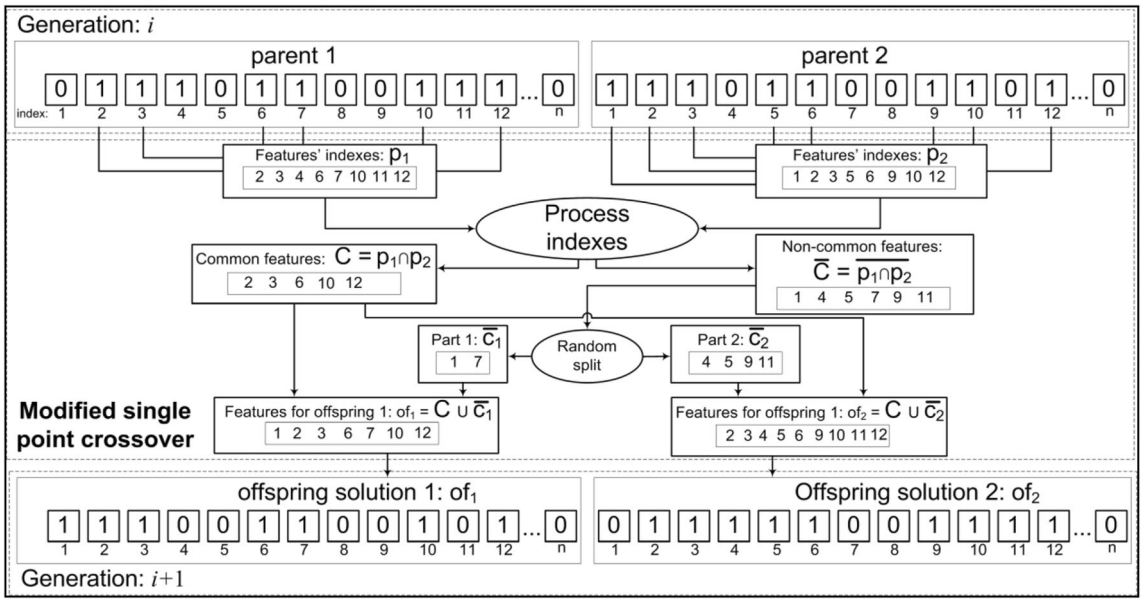

Fig. 3 Modified single-point crossover for binary data

features from the "non-common" set or $\overline{c}_2 = \{4, 5, 9, 11\}$. Features for offspring 1 of_1 is a combination of features from the "common" set $C = \{2, 3, 6, 10, 12\}$ and "part 1" $\overline{c}_1 = \{1, 7\}$ or $of_1 = C \cup \overline{c}_1 = \{1, 2, 3, 6, 7, 10, 12\}$. Features for offspring 2 of_2 is a combination of features from the "common" set $C = \{2, 3, 6, 10, 12\}$ and "part 2" $\overline{c}_2 = \{4, 5, 9, 11\}$ or $of_2 = C \cup \overline{c}_2 = \{2, 3, 4, 5, 6, 9, 10, 11, 12\}$.

When both the parent solutions do not have any common features, all the features belonging to the parents will be in the "non-common" set. Then, the offspring solutions will be created only from the set of the "non-common" features. The modification for both the offspring solutions compared with the original parent features will be very significant. Alternatively, when most of the features belong to the "common" set, the modifications in the offspring solutions will be minor. During the first iteration, when most of the features are different, the search will be chaotic and random. Near the end of the iterations, when most of the features are common, the search will be focused on finding the optimal set of features.

Modified Binary Mutation Operator

For solving the ADHD problem, a modified version of the standard binary mutation operator has been integrated into the FSGA. Most of the formal binary mutations operators presented in the literature keep a normal distribution of binaries 1 and 0 in the offspring solutions. This creates a significant number of binary coefficient 1 in the offspring. A larger set of selected features used to train the ELM classifier causes significant performance loss compared with using a smaller set. The proposed modified binary mutation operator picks one parent solution randomly from the current generation i, counts the number of binary coefficients of "1" (N), and generates a new offspring solution with randomly allocated N binary coefficients "1." Thus, the number of features in the new offspring solution for the next generation remains the same as compared with the parent solution.

Based on our studies, a combination of the proposed modified single-point crossover and binary mutation operators shows better performance for ADHD classification compared with other methods in the literature.

Extreme Learning Machine Classifier

ADHD diagnosis using ROI_3 data is a three-class problem involving thousands of features in a high-dimensional feature space. The objective of any classifier is to approximate the functional relationship between the features to their corresponding class labels. In this study, a Gaussian activation function is used in the hidden layer of an ELM classifier to approximate the classification decision boundary. Selection of ELM is mainly based on its training speed compared with other classifiers. ELM is a single hidden layer feed-forward neural network where the input weights are randomly assigned and output weights are analytically estimated [38–40]. The hidden neurons have randomly assigned bias values and input weights. ELM provides a unified framework to solve both regression and classification problems. The classification problem is defined as follows:

Given a training dataset with N samples, $\{(X^1, c^1), \ldots, (X^t, c^t), \ldots, (X^N, c^N)\}$, where $X^t \in R^m$ are the m-dimensional input features of t sample and $c^t \in \{1, 2, \ldots, K\}$ is it tth class label. The coded class label y^t is obtained using:

$$y_k^t = \begin{cases} 1, \text{if } c^t = k \\ -1, \text{otherwise} \end{cases} k = 1, 2, ..., K \tag{2}$$

The classification problem is that of approximating the decision function (F_{ELM}) that maps the input features to the coded class labels, i.e., $F_{\text{ELM}} : x \rightarrow y$. The neurons in the outermost layers are linear, while the hidden layer neurons use the Gaussian activation functions. ELM is used to solve the problem with m input neurons and K output neurons. However, the performance of the ELM is dependent on the randomly chosen centers and hidden neuron's bias values, especially in an application like the ADHD classification problem, where the input feature dimension is large [37].

Single ELM and ensemble of ELM have been widely studied recently [48–52]. Ensemble unifies a set of classifiers by using either votes [52] or weights [50, 51]. Authors usually have to solve two problems: (1) pick the best set of classifiers for ensemble and (2) assign appropriate weights for ensemble. In this paper, a classifier-selecting genetic algorithm is used to pick the best set of classifiers for further ensemble and risk-sensitive cognitive ensemble classifier approach to assign weights.

Next, we describe the region-based ensemble classifier.

Region-Based Ensemble Classifier

In a particular region, a union or ensemble of individual classifiers may produce a better classification performance compared with a single individual classifier. Each classifier uses a unique reduced set of features for training. It has been shown in [41] that a combination (ensemble) of many base learners trained using different sets of features results in better classification accuracies. The classification of the different subtypes of ADHD and TDC needs more discriminative features for accurate classification. Hence, in this section, we describe a methodology for forming the region-based ensemble classifier, which is an ensemble of ELM classifiers with their corresponding selected features. The CSGA is used for selecting the classifiers from a big pool, along with their discriminant features. The ensemble of classifiers is formed using a risk-sensitive loss function. In the following sections, both these approaches are explained in more detail.

Classifier-Selecting Genetic Algorithm

The framework of CSGA is similar to that of FSGA. Instead of the features that are used in FSGA, the CSGA uses a pool of ELM classifiers (referred to an ensemble pool in Fig. 1) generated by HipC, AmyC, and CauC in the previous stage. All generated ELM classifiers were unified together and ranked based on their training accuracies. It was noted that the ELM classifiers generated by HipC and AmyC have similar training accuracies, whereas the accuracies obtained for CauC are relatively lower. Next, all those ELM classifiers generated by HipC, AmyC, and CauC producing accuracies in the range of 55–62% are collected to form the ensemble pool of classifiers. Thus, the "ensemble pool" consists of 2148 ELM classifiers (see Fig. 1) and this includes 1045 ELM classifiers generated by HipC ($H = 1045$), 1000 ELM classifiers generated by AmyC ($A = 1000$), and 103 ELM classifiers generated by CauC ($C = 103$). These sets of ELM classifiers are then used for each three of the region-based ensemble classifiers for the hippocampus, amygdala, and caudate. For the final fusion, the total ensemble pool of 2148 selected ELM classifiers is used by the multiple-region ensemble classification approach (MRECA).

Each of the three region-based ensemble classifiers uses its own CSGA for searching for the best set of ELM classifiers for further fusion. CSGA for HipEC creates a binary vector F with 1045 binary coefficients, where binary coefficients "1" indicate the chosen ELM classifier (similarly for the other two classifiers). The selected ELM classifiers are then processed and combined using a risk-sensitive ensemble classifier approach described below.

Region-based ensemble classifier selected 52, 45, and 41 ELM classifiers to build "HipEC," "AmyEC," and "CauEC," respectively. Fifty-two selected ELM classifiers of "HipEC" use 922 features from the 6072 features of the hippocampus. Similarly, 45 and 41 ELM classifiers of "AmyEC" and "CauEC" use 766 and 823 features of the amygdala and caudate, respectively. This is summarized in Table 2. Note that each ELM classifier uses its own features set and some of the same features may appear in the features sets of other ELM classifiers.

Risk-Sensitive Cognitive Ensemble Classifier Approach

The risk-sensitive cognitive ensemble classifier approach is used to build a fusion of the ELM classifiers for the three region-based ensemble classifiers (i.e., "HipEC," "AmyEC," and "CauEC"). This ensemble approach is using a modified version of the approach proposed earlier by Sachnev et al. (see [41]). A modified RSHL designed for multiclass classification problems is presented below.

Risk-sensitive hinge loss (RSHL) error function enables an efficient fusion of individual classifiers into one ensemble group that yields improved classification performance. RSHL for an ensemble classifier was earlier used in [41] for solving complex classification problems where the traditional machine learning techniques failed to correctly classify the data due to a large number of features, sparsity, and imbalances in the data. This problem arises in the case of ADHD data too.

In this paper, a modified version of the cognitive ensemble approach developed earlier in [41] and herein referred to as the risk-sensitive cognitive ensemble classifier, is used to diagnose ADHD accurately (Fig. 1).

In the ensemble part, each chosen ELM classifier is penalized by the corresponding weight a (Eq. (5)). The sum of all the penalized outputs of the chosen ELM classifiers is the final output of the proposed risk-sensitive ensemble. The method for computing the corresponding weights a is a key component of any ensemble technique [41]. RSHL is used to compute these weights in this study and a brief description of the same is given below.

All the ELM classifiers in a pool contain the outputs $\hat{y}$, which is a matrix with size $3 \times N$, where N is the number of training samples. H, A, and C are the number of ELM classifiers for the three regions.

Let $\hat{y} = \hat{y}^{\text{hip}} \cup \hat{y}^{\text{amy}} \cup \hat{y}^{\text{cau}}$ is the set of ELM classifiers' outputs selected by CSGA, where $\hat{y}^{\text{hip}} = [\hat{y}_1^{\text{hip}}, \hat{y}_2^{\text{hip}}, \ldots, \hat{y}_{H_1}^{\text{hip}}]$, $\hat{y}^{\text{amy}} = [\hat{y}_1^{\text{amy}}, \hat{y}_2^{\text{amy}}, \ldots, \hat{y}_{A_1}^{\text{amy}}]$, and $\hat{y}^{\text{cau}} = [\hat{y}_1^{\text{cau}}, \hat{y}_2^{\text{cau}}, \ldots, \hat{y}_{C_1}^{\text{cau}}]$ are the sets of H_1, A_1, and C_1 chosen ELM classifiers trained by using features from ROI_3. Assume that $M_1 = H_1 + A_1 + C_1$ then $a = [a_1, a_2, \ldots, a_{M_1}]$ are the weights to penalize ELM models.

- Initialization:
- Generate initial weights randomly: $a^0 = [a_1^0, a_2^0, \ldots, a_{M_1}^0]$
- Compute first initial ensemble classifier output as follows:

$$\hat{y}^0 = a_1^0\ \hat{y}_1^0 + \ldots + a_l^0\ \hat{y}_l^0 + \ldots + a_{M_1}^0\ \hat{y}_{M_1}^0 \tag{3}$$

- Update weights:
- Update the weights using risk-sensitive hinge loss error function (RSHL) as follows:

$$a_l = a_l^{'} + \eta \sum_{i=1}^{3} e(i)_l \hat{y}(i)_l + \beta\delta \tag{4}$$

where

$$\delta = \eta \sum_{i=1}^{3} e^{'}(i)_l \hat{y}(i)_l + \beta\delta^{'} \tag{5}$$

Here, $\eta = 10^{-6}$ and $\delta = 10^{-7}$ are the learning rate constants, δ is the adaptive factor for the current iteration, $\delta^{'}$ is the adaptive factor counted in the previous iteration, $a^{'}$ is the weights counted

in the previous iteration, a_l is the weight of the lth classifier for the current iteration and $e(i)_l$ is the error of the lth classifier for the current iteration, $i = 1, 2, 3$ is the class counter, and $\hat{y}(i)_l$ is the output of the lth classifier in respect of the class counter.

- Compute the ensemble classifier output using updated weights.

$$\hat{y}^s = a_1^s\ \hat{y}_1^s + ... + a_l^s\ \hat{y}_l^s + ... + a_{M_1}^s\ \hat{y}_{M_1}^s \tag{6}$$

- Compute the overall training efficiency.
- Repeat step "update weights" until there is no performance improvement during the last 100 iterations.

Thus, the ensemble of ELM classifiers realized as above is our region ensemble classifiers "HipEC," "AmyEC," and "CauEC" which are better than each individual ELM classifier.

Multi-Region Ensemble Classifier Approach

Multi-region ensemble classifier approach (MRECA) is the integration of all the three region-based ensemble classifiers, where all the available 2148 ELM classifiers are processed together in a wraparound approach. Note that MRECA does not use the ELM classifiers chosen by the region-based ensemble classifiers HipEC, AmyEC, and CauEC. The CSGA is again used in the MRECA framework. It contains a binary string representation vector F with 2148 binary coefficients, where binary coefficient "1" highlights the chosen ELM classifiers. Selected ELM classifiers are fused together by using the risk-sensitive cognitive ensemble classifier approach. MRECA searches for the best set of the ELM classifiers with the best overall testing and training accuracies. The experimental results and performance evaluation for MRECA are given in the next section.

Results

Performance Evaluation of MRECA and Discussion

In this section, the performance evaluation of the MRECA approach for the diagnosis of ADHD and its subtypes is presented under three scenarios that were previously discussed. In addition, we try to identify voxels in these three regions, which might be responsible for affecting ADHD patients.

Scenario 1: Performance of Region-Based Individual Classifier Region-Based Individual (Best) Classifiers Performance of region-based individual (best) classifiers was arrived at for each of three regions separately. The best possible single ELM classifier results are shown in (Fig. 1) and Table 1. The best ELM classifiers $*\mathbf{ELM}^{hip}$, $*\mathbf{ELM}^{amy}$, and $*\mathbf{ELM}^{cau}$ were trained based on selected best features $*\mathbf{fs}^{hip}$, $*\mathbf{fs}^{amy}$, and $*\mathbf{fs}^{cau}$ that gave the highest classification performance in that region. Table 1 presents these experimental results and contains the confusion matrices along with the overall testing and training accuracies for these classifiers.

ELM classifier for the caudate shows the lowest overall classification accuracies of 57.05% for training and 50.88% for testing. The best overall training and testing accuracy of 62.18% and 57.39%, respectively, is obtained for the hippocampus. Experiments involving the amygdala show an overall training accuracy of 58.49% and testing accuracy of 56.8%. Confusion matrices for all the three experiments show significant misclassification in all the three classes (Table 1). The study in the hippocampus shows slightly better results compared with the experiments in the amygdala and caudate.

Based on this study, it may be inferred that a single standalone ELM classifier is not sufficient to classify ADHD subtypes with higher accuracy.

Scenario 2: Performance of Region-Based Ensemble Classifiers FSGA and a risk-sensitive hinge loss function (RSHL) were used to build an ensemble of ELM classifiers for each of the three regions (Fig. 1). Table 2 shows the experimental results of "region-based ensemble classifiers" for the hippocampus (HipEC), amygdala (AmyEC), and caudate (CauEC) and includes confusion matrices along with their overall training and testing accuracies.

Organization of Table 2 is similar to that of Table 1. For the three regions considered in this study, the region-based ensemble classifiers improve the overall training and testing performances of ADHD classification by 12.52%, 14.06%, and 14.49% for training and by 13.61%, 20.12%, and 17.16% for testing (compared with the region-based individual best classifier), respectively.

With the region-based classifiers, the confusion matrices display significant improvements compared with the previous experiments done with individual classifiers. Misclassification between classes is significantly reduced. Misclassification between TDC, ADHD-C, and ADHD-I is still significant in each experiment. Similar to the region-based individual classifiers, region-based ensemble classifier for the hippocampus (HipEC) shows higher training and testing accuracies compared with others. However, misclassification between classes is still significant and needs further improvement.

Scenario 3: Performance of MRECA In this study, a fusion of all the three brain regions-based classifiers is studied. Multiple regions ensemble classifier is the combination of FSGA and the ensemble classifier approach based on the risk-sensitive hinge loss function (Fig. 1) for all of the three regions put together. In this experiment, 2148 ELM classifiers have been selected from the set of the single ELM classifiers generated from scenario 1 for arriving at the ensemble.

Table 1 Experimental results for the best region-based individual classifiers

Brain region/best classifier	ADHD subtypes	Training confusion matrix			Testing confusion matrix			Training accuracy (%)	Testing accuracy (%)	Selected features/all features
		TDC	ADHD-C	ADHD-I	TDC	ADHD-C	ADHD-I			
Amygdala ($^{*}ELM^{amy}$)	TDC	*251*	110	126	*61*	17	16	58	56.8	122/1050
	ADHD-C	31	*104*	26	12	*21*	16			
	ADHD-I	11	11	*89*	5	7	*14*			
Caudate ($^{*}ELM^{cau}$)	TDC	*252*	129	106	*53*	25	16	57	50.9	101/3904
	ADHD-C	26	*106*	29	19	*19*	11			
	ADHD-I	22	14	*75*	9	3	*14*			
Hippocampus($^{*}ELM$)	TDC	*281*	104	102	*62*	17	15	62	57.4	98/6072
	ADHD-C	30	*110*	21	14	*24*	11			
	ADHD-I	20	10	*81*	10	5	*11*			

The numbers marked in italics highlight the number of correctly classified samples for TDC, ADHD-I and ADHD-C

The '*' symbol indicates that the 'best' classifiers were trained based on selected best features

MRECA selected 39 ELM classifiers from the hippocampus, 25 ELM classifiers from the amygdala, and 10 ELM classifiers from the caudate to build an ensemble classifier resulting in maximum training and testing overall accuracies.

The study results are presented in Table 3. Organization of Table 3 is similar to that of Tables 1 and 2 except that it presents the results only for the best multiple regions ensemble classifier. MRECA significantly outperforms "region-based individual classifier" and the "region-based ensemble classifier" for the ADHD diagnosis. MRECA achieves a maximum overall training accuracy of 82.74% and a maximum overall testing accuracy of 81.24%. MRECA obtains better results compared with the "region-based single classifier" by 25.25%, 25.69%, and 20.56% in terms of overall training accuracies and by 24.44%, 30.36%, and 23.85% in terms of overall testing accuracies for the three regions ROI3. MRECA shows better results compared with "region-based ensemble classifier" by 11.73%, 11.63%, and 6.07% in terms of overall training accuracies and by 10.83%, 10.24%, and 6.69% in terms of overall testing accuracies for ROI_3.

The confusion matrices generated with MRECA are shown in Table 3. Misclassification between subtypes of ADHD (ADHD-C and ADHD-I) is negligible. Only 7 ADHD-C among 161 were classified as ADHD-I (see Table 3 "Training confusion matrix"). However, TDC samples are still misclassified. Fifty-five TDC samples were classified as ADHD-C and 64 samples were classified as ADHD-I. Besides, 4 ADHD-C samples and 1 ADHD-I sample were classified as TDC samples. Similar observations can be made about the experiment with testing samples. Misclassification between all 3 classes for testing samples is slightly higher compared with the experiments with training samples. "Testing confusion matrix" shows a similar imbalance between all classes.

Next, a study was undertaken to evaluate the improvements obtained by the modified crossover and mutation operators in CSGA along with that of a risk-sensitive ensemble classifier. The results of the study are summarized in the next section.

Table 2 Experimental results for the Region-based ensemble classifiers

Brain region/best classifier	ADHD subtypes	Training confusion matrix			Testing confusion matrix			Training accuracy (%)	Testing accuracy (%)	Selected features/all features
		TDC	ADHD-C	ADHD-I	TDC	ADHD-C	ADHD-I			
Amygdala (AmyEC)	TDC	*317*	81	89	*76*	8	10	71	70.4	766^{*}/1050
	ADHD-C	9	*123*	29	14	*27*	8			
	ADHD-I	5	7	*99*	7	3	*16*			
Caudate (CauEC)	TDC	*292*	95	100	*74*	13	7	71.1	71	823^{*}/3904
	ADHD-C	8	*142*	11	10	*32*	7			
	ADHD-I	3	2	*106*	9	3	*14*			
Hippocampus (HipEC)	TDC	*341*	70	76	*72*	7	15	76.7	74.6	922^{*}/6072
	ADHD-C	10	*137*	14	6	*38*	5			
	ADHD-I	4	3	*104*	7	3	*16*			

The numbers marked in italics highlight the number of correctly classified samples for TDC, ADHD-I and ADHD-C

The '*' symbol indicates that the 'best' classifiers were trained based on selected best features

Table 3 Experimental results for MRECA

Brain region/ensemble classifier	ADHD subtypes	Training confusion matrix			Testing confusion matrix			Training accuracy (%)	Testing accuracy (%)	Selected features/all features
		TDC	ADHD-C	ADHD-I	TDC	ADHD-C	ADHD-I			
Fusion of the amygdala, caudate, and hippocampus (MRECA)	TDC	*368*	55	64	*83*	5	6	82.74	81.2	3498*/11026
	ADHD-C	4	*150*	7	5	*39*	5			
	ADHD-I	1	0	*110*	6	3	*17*			

The numbers marked in italics highlight the number of correctly classified samples for TDC, ADHD-I and ADHD-C

The '*' symbol indicates that the 'best' classifiers were trained based on selected best features

Effect of Risk-Sensitive Cognitive Ensemble Classifier and Modified Operators in CSGA

For the CSGA, the maximum and average testing accuracies are expected to increase from one generation to the next until the maximum is reached as illustrated in Fig. 4 a. Results for a classifier-selecting genetic algorithm with classical binary crossover (uniform crossovers) and bit inversion mutation is given in Fig. 4 b. This shows minor improvement in the accuracies for subsequent generations. CSGA, which is based on the proposed modified single-point crossover and modified binary mutation operator (Fig. 4 a). shows significant improvement when compared with CSGA with the classical operators.

Performance evaluation of the risk-sensitive ensemble classifier is given in Fig. 4 c. In this figure, the fusion of the best MRECA, which are 39 ELM classifiers from the hippocampus, 25 ELM classifiers from the amygdala, and 10 ELM classifiers from the caudate, is presented. Risk-sensitive cognitive ensemble classifier starts from a very low testing accuracy of 35% and steadily and significantly improves until it reaches the maximum accuracy of 81.24%. Overall training accuracy shows a similar trend.

Performance Comparison with Other Existing Methods

Many of the findings that use MRI data come from studies that use a variety of data sets for different brain regions and apply a wide range of analytical tools. This makes it difficult to analyze and compare the results and draw meaningful conclusions [53–55].

Only one study in the literature that uses ADHD-200 hippocampus data for a four-class classification problem has been reported by Sachnev et al. (see [34]). In this work, the authors combined the binary-coded genetic algorithm with a Meta-Cognitive Neuro-Fuzzy Inference System (McFIS) to classify TDC, ADHD-C, ADHD-H, and ADHD-I with a testing accuracy of 56%. Mahanand et al. classified 3 subtypes of ADHD (ADHD-C, ADHD-H, ADHD-I) using a meta-cognitive radial basis function network (PBL-McRBFN) coupled with chi^2 feature selection technique [23]. Their method deals with the amygdala and the cerebellar vermis brain areas. The authors reported an overall testing accuracy of around 62%. Qureshi et al. [56] proposed an ELM based classifier for three classes (ADHD-C, ADHD-H, ADHD-I) with a reduced number of ADHD patients (159 patients). The authors reported about 60.78% of testing accuracy. Dai et al. [57] classified 3 subtypes of ADHD using multi-kernel learning (MKL) with accuracy 54.1%.

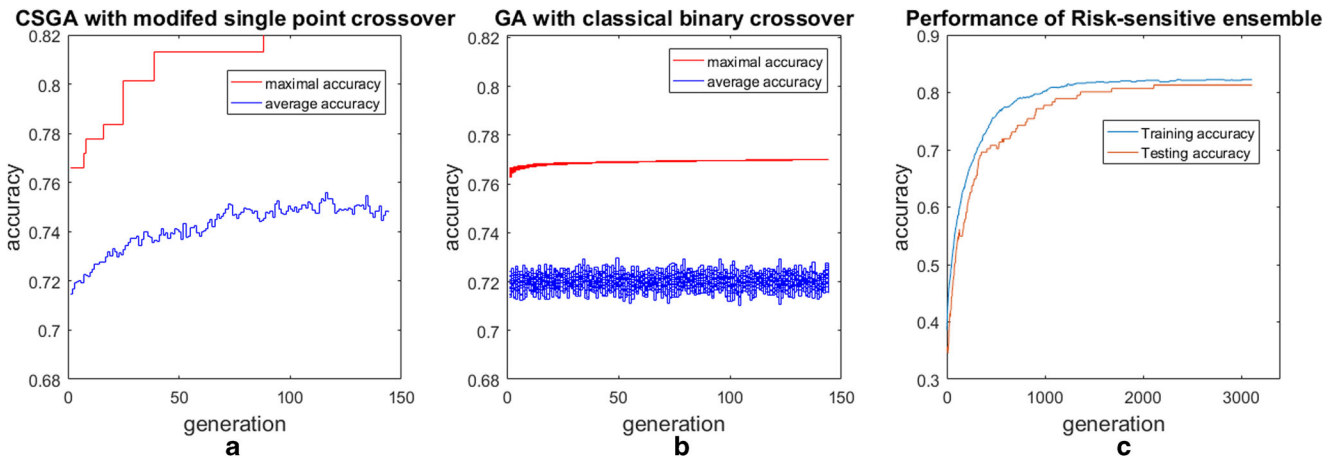

Fig. 4 **a** (left), **b** (middle) Generalization performance of the CSGA and **c** (right) generalization performance of risk-sensitive ensemble classifier

Table 4 Summary of comparison of our performance with existing techniques

No. of classes	Classifier	Features		Overall accuracies	
		Description	No. of features	Training (%)	Testing (%)
Three-class	PBL-McRBFN [29]	ROI over the amygdala + the cerebellar vermis	7412	65.00	62.00
	H-ELM [56]	RFE-SVM over the whole brain	320	–	60.78
	MKL [57]	The whole brain	–	–	53.2
Three*-class	MRECA	FSGA over the hippocampus + the amygdala + the caudate	11,026/3498**	82.74	81.24

The 3-class classification includes ADHD-C, ADHD-H, and ADHD-I

The 3*-class classification includes TDC, ADHD-C, and ADHD-I

**features chosen by MRECA

The proposed MRECA has been tested using 3 classes (TDC, ADHD-C, and ADHD-I) which is different compared with other work done with ADHD data. Hence, a direct comparison of our results with other projects in literature is not possible. However, some comparative results are shown in Table 4.

Statistical Analysis

In this section, we performed statistical tests with the seven proposed ADHD diagnosis methods in this paper: (1) 3 region-based individual classifiers for ELM^{hip}, ELM^{amy}, and ELM^{cau}, (2) 3 region-based ensemble classifier for HipEC, AmyEC, and CauEC, and 3) the multiple regions ensemble classifier MRECA. Results of the *t* test are shown in Table 5.

t tests show a significant advantage of the multiple-region ensemble classifier (MRECA) compared with the other six methods. Pairwise *t* test measures the similarity in a couple of classification methods. *t* test MRECA vs. other methods show numbers close to 0, which indicates a significant difference in terms of classification performance, i.e., MRECA can correctly classify more samples compared with other methods.

Finally, *t* test indicates the importance of the ensemble hippocampus, amygdala, and caudate for more accurate ADHD diagnosis.

Search for Voxels (Regions) Responsible for ADHD

As a continuation of the above studies, an attempt to evaluate the impact of the individual voxels (the locations in the regions) from the MRI images for the ADHD diagnosis is presented.

In general, ADHD research is based on a statistical analysis of the MRIs of the subjects under study. Researchers try to develop techniques which can identify the micro-regions of the human brain (voxels) responsible for ADHD. In this study, machine learning has been used for such an analysis. During the training procedure, machine learning techniques accumulate the hidden knowledge about ADHD-200 in their network structures. Thus, machine learning may efficiently classify three types of ADHD patients and TDC. However, machine learning techniques can not directly evaluate the importance of features (voxels) used for training. Hence, a new machine learning technique with the ability to evaluate the impact of different features (voxels) is needed.

In this paper, MRECA has provided such a mechanism for combining the feature selection technique and the machine learning method (classifier) together. Such a combination searches for an optimal set of features (voxels) for accurate classification. This combination of feature selection and machine learning classifier can be a very efficient approach to evaluate the impact of individual voxels for ADHD diagnosis. In this new approach, ten multiple regions ensemble classifiers with efficiencies closer to the maximum that can be achieved

Table 5 Pairwise *t* test for the seven examined ADHD diagnosis methods

	ELM^{hip}	ELM^{amy}	ELM^{cau}	HipEC	AmyEC	CauEC	MRECA
ELM^{hip}	x	0.2413	0.3522	0.777	0.3513	0.1784	0.052
ELM^{amy}		–	0.75	0.11	0.0141*	0.0138*	0.019*
ELM^{cau}			–	0.2398	0.0532	0.021*	0.0026*
HipEC				–	0.3993	0.2452	0.0159*
AmyEC					–	0.7075	0.3112
CauEC						–	0.6547
MRECA							–

**t* test rejects the null hypothesis at the default 5% significance level

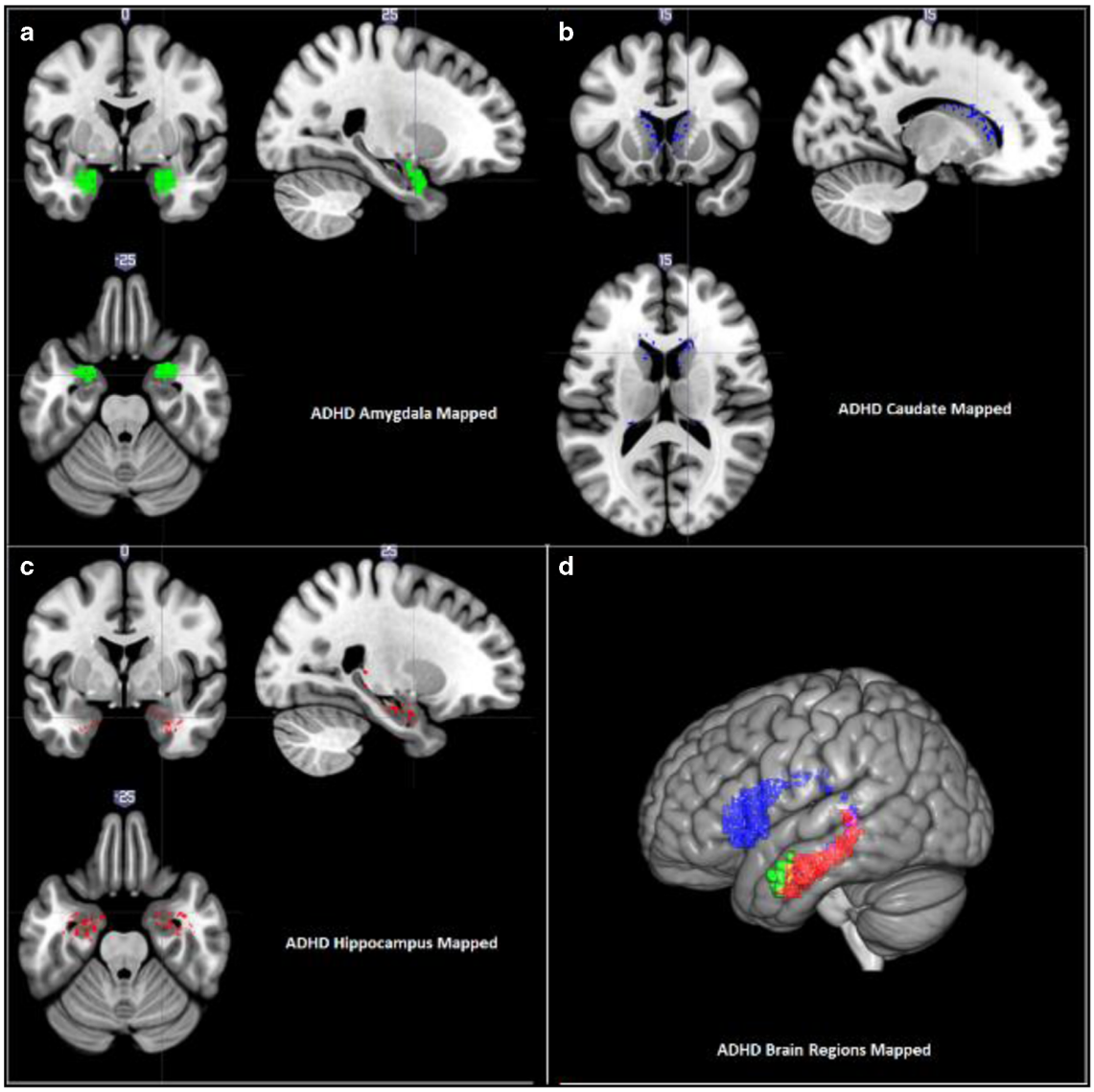

Fig. 5 Discovered voxels responsible for ADHD in **a** the amygdala (green), **b** caudate (blue), **c** hippocampus (red), and **d** the consolidated view of voxels marked for all three regions

have been used for a voxels impact analysis. Note that each ELM classifier has been trained using an individual set of voxels chosen by the FSGA.

TOP-10 classifiers unify the 39 ELM classifiers from the hippocampus, 306 ELM classifiers from the amygdala, and 64 ELM classifiers from the caudate area. Voxels analysis is based on counting the number of times when each voxel was chosen by the FSGA in the set of 393 ELM classifiers.

Figure 5 maps back the more frequently occurring voxels to the brain regions (1529 voxels from the hippocampus, 864 voxels from the amygdala, and 1098 voxels from the caudate). The regions selected are responsible for the attention, cognition, emotion, and sensorimotor functions. Specifically, voxels from the amygdala play a primary role in the processing of memory, decision-making, and emotional reactions. Voxels from caudate nucleus play an important role in various motor and non-motor functions such as procedural learning, associative learning, inhibitory control of action, and reward system. Voxels from the hippocampus play an important role in the consolidation of information from short-term memory, long-term memory, and spatial memory.

Conclusions

Currently, clinical diagnosis for early identification of ADHD subtypes poses real challenges when the clinical criteria do not fully meet or when there are comorbid conditions complicating the picture. In this paper, a multiple-region ensemble classification approach (MRECA) for an accurate ADHD diagnosis is presented. To improve the classification accuracy, two

new methods for feature selection referred to as a feature-selecting GA and a classifier-selecting GA which are used to select the higher performing one from a collection of classifiers have been developed. The feature-selecting GA has been modified with an improved crossover operator and the mutation operator to specifically handle the ADHD sub-classification issues. For selecting the classifiers, a risk-sensitive hinge loss function is used to form the cognitive ensemble of classifiers. Performance of the proposed MRECA approach is evaluated using the ADHD-200 consortium data under seven different scenarios. Based on the studies, it is seen that the proposed multiple-region ensemble classification approach for ADHD diagnosis shows higher training (82.74%) and testing (81.24%) accuracies. The statistical tests performed on the results show that MRECA does better than other methods. Based on the selected features, the brain regions most affected by ADHD have been identified indicating their importance for the early identification of ADHD. Our studies clearly indicate that sMRI data from the three brain regions, namely, the amygdala, caudate, and hippocampus, can be used in unison to diagnose ADHD subtypes.

Acknowledgments The authors would like to thank the ADHD-200 consortium and the Neuro Bureau for making the MRI data available. We would like to thank the ADHD-200 Pre-processed initiative and Dr. Carlton Chu for the Burner pipeline.

Funding information This work was supported by Catholic University of Korea, Research Funds 2016, National Research Foundation of Korea, grant #2011-0013695.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Research Involving Human Participants and/or Animals/Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. *Diagnostic and Statistical Manual of Mental Disorders*, fifth edn. (American Psychiatric Association).
2. Polanczyk G, DeLima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatr. 2007;164:942–8.
3. Biederman J. Attention-deficit/hyperactivity disorder: a life-span perspective. J Clin Psychiatry. 1998;59(7):4–16.
4. Frodl T, Skokauskas N. Meta-analysis of structural MRI studies in children and adults with attention-deficit hyperactivity disorder indicates treatment effects. Acta Psychiatr Scand. 2012;125:114–26.
5. Doshi JA, Hodgkins P, Kahle J, Sikirica V, Cangelosi MJ, Setyawan J, et al. Economic impact of childhood and adult attention-deficit/hyperactivity disorder in the United States. J Am Acad Child Adolesc Psychiatry. 2012;51(10):990–1002.
6. Davenport N, Karatekin C, White T, Lim KO. Differential fractional anisotropy abnormalities in adolescents with ADHD or schizophrenia. Psychiatry Res. 2010;181:193–8.
7. Selikowitz M. ADHD. Oxford press; 2009.
8. Fuente ADL, Xia S, Branch C, Li X. A review of attention-deficit/hyperactivity disorder from the perspective of brain networks. Front Hum Neurosci. 2013;7(192):1–6.
9. Wolosin SM, Richardson ME, Hennessey JG, Denckla MB, Mostofsky SH. Abnormal cerebral cortex structure in children with ADHD. Hum Brain Mapp. 2009;30(1):175–84.
10. Giedd JN, Rapoport JL. Structural MRI of pediatric brain development: what have we learned and where are we going? Neuron. 2010;67:728–34.
11. Whalen PJ, Phelps EA. The human amygdala. Guilford Press, 2009.
12. Markowitsch H. Differential contribution of right and left amygdala to affective information processing. Behav Neurol. 1998;11(4):233–44.
13. Nestler EJ, Hyman SE, Holtzman DM, Malenka RC. Molecular neuropharmacology: a foundation for clinical neuroscience (3d ed.). McGraw Hill; 2009.
14. Martin JH Neuroanatomy: text and atlas, 4th edn. McGraw Hill; 2003.
15. Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Vaituzis AC, Dickstein DP, et al. Quantitative brain magnetic resonance 560 imaging in attention-deficit hyperactivity disorder. Arch Gen Psychiatry. 1996;53:607–16.
16. Kyeong S, Park S, Cheon K-A, Kim J-J, Song D-H, Kim E. A new approach to investigate the association between brain functional connectivity and disease characteristics of attention-deficit/hyperactivity disorder: topological neuroimaging data analysis. PLoS One. 2015;10(9):e0137296.
17. Nachamai M. Sub-type discernment of attention-deficit hyperactive disorder in children using a cluster partitioning algorithm. 2016; Indian J Sci Technol **9**(8).
18. Wang J-B, Zheng L-J, Cao Q-J, Wang Y-F, Sun L, Zang Y-F, et al. Inconsistency in abnormal brain activity across cohorts of ADHD-200 in children with attention-deficit hyperactivity disorder. Front Neurosci. 2017;11:320.
19. Bellec P, Chu C, Chouinard-Decorte F, Benhajali Y, Margulies DS, Craddock RC. The Neuro Bureau ADHD-200 preprocessed repository. *NeuroImage*. 2017;**144**(*Part B*):*275*–86.
20. Madureira DQM, Carvalho LAV, Cheniaux E. Focus modulated by mesothalamic dopamine: consequences in Parkinson's disease and attention-deficit hyperactivity disorder. Cogn Comput. 2010;2(1):31–49.
21. Zou L, Xu S, Ma Z, Lu J, Su W. Attentional automatic removal of artifacts from attention-deficit hyperactivity disorder electroencephalograms based on independent component analysis. Cogn Comput. 2013;5(2):225–33.
22. Anderson A, Douglas PK, Kerr WT, Haynes VS, Yuille AL, Xie J, et al. Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. NeuroImage. 2014;19(102).
23. Iannaccone R, Hauser TU, Ball J, Brandeis D, Walitza S, Brem S. Classifying adolescent attention-deficit/hyperactivity disorder (ADHD) based on functional and structural imaging. Eur Child Adolesc Psychiatry. 2015;24(10):1279–89.
24. Chang CW, Ho CC and Chen JH. ADHD classification by a texture analysis of anatomical brain MRI data, Front Syst Neurosci. 2012; 6.
25. Kobel M, Bechtel N, Specht K, Klarhofer M, Weber P, Scheffler K, et al. Structural and functional imaging approaches in attention-deficit/hyperactivity disorder: does the temporal lobe play a key role? Psychiatry Res. 2010;183:230–6.

26. Cao M, Shu N, Cao Q, Wang Y, He Y. Imaging functional and structural brain connectomics in attention-deficit/hyperactivity disorder. Mol Neurobiol. 2014;50(3):1111–23.
27. Rangarajan B, Suresh S, Mahanand BS. Identification of potential biomarkers in the hippocampus region for the diagnosis of ADHD using PBL-McRBFN approach, Proceeding of 2014 13th International Conference on Control Automation Robotics and Vision (ICARCV) 2014; 2, 17–22.
28. Sabuncu MR, Konukoglu E. Clinical prediction from structural brain MRI scans: a large-scale empirical study. Neuroinformatics. 2015;13(1):31–46.
29. Mahanand BS, Savitha R, Suresh S. Computer aided diagnosis of ADHD using brain magnetic resonance images. Advances in Artificial Intelligence. 2013; 386–395.
30. Sato JR, Hoexter MQ, Fujita A, Rohde LA. Evaluation of pattern recognition and feature extraction methods in ADHD prediction, Front Syst Neurosci. 2012; **6**.
31. Eloyan A, Muschelli J, Nebel MB, Liu H, Han F, Zhao T, Barber AD, Joel S, Pekar JJ, Mostofsky SH, Caffo B. Automated diagnoses of attention-deficit hyperactive disorder using magnetic resonance imaging., Front Syst Neurosci 2012; **6**.
32. Colby JB, Rudie JD, Brown JA, Douglas PK, Cohen MS, Shehzad Z. Insights into multimodal imaging classification of ADHD. Front Syst Neurosci (59) (6) 1–18.
33. Peng X, Lin P, Zhang T, Wang J. Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS One. 2013;8:e79476.
34. Sachnev V. An efficient classification scheme for ADHD problem based on binary-coded genetic algorithm and McFIS, Proceeding on 2015 International Conference on Cognitive Computing and Information Processing (CCIP) 2015; 1–6.
35. Maldjian J, Laurienti P, Kraft R, Burdette J. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. NeuroImage. 2003;19(3):1233–9.
36. Murugan MS, Suresh S, Ganguli R, Mani V. Target vector optimization of composite box beam using real-coded genetic algorithm: a decomposition approach. Struct Multidiscip Optim. 2007;33:131–46.
37. Suresh S, Babu V, Sundararajan N. Image quality measurement using sparse extreme learning machine classifier, 2006 9th International Conference on Control, Automation, Robotics and Vision 2006; pp. 1–6.
38. Huang G-B, Zhu QY, Siew CK. Extreme learning machine: theory and application. Neurocomputing. 2006;70:489–501.
39. Huang G-B. An insight into extreme learning machines: random neurons, random features and kernels. Cogn Comput. 2014;6:376–90.
40. Huang G-B. What are extreme learning machines? Filling the gap between frank Rosenblatt's dream and John von Neumann's puzzle. Cogn Comput. 2015;7:263–78.
41. Sachnev V, Savitha R, Suresh S, Kim HJ, Hwang HJ. A cognitive ensemble of extreme learning machines for steganalysis based on risk-sensitive hinge loss function. Cogn Comput. 2014;7(1):103–10.
42. Milham MP, Fair D, Mennes M, Mostofsky SH. The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Front Syst Neurosci. 2012;6(62):1–5.
43. Friston K, Ashburner J, Kiebel S, Nichols T, Penny W (eds.). Statistical parametric mapping: the analysis of functional brain images. Academic Press; 2007).
44. Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage. 2007;38(1):95–113.
45. Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence. U Michigan Press; 1975.
46. Suresh S, Saraswathi S, Sundararajan N. Performance enhancement of extreme learning machine for multi-category sparse data classification problems. Eng Appl Artif Intell. 2010;23(7):1149–57.
47. Suresh S, Omkar SN, Mani V, Prakash TNG. Lift coefficient prediction at high angle of attack using recurrent neural network. Aerosp Sci Technol. 2003;7:595–602.
48. Rong H-J, Jia Y-X, Zhao G-S. Aircraft recognition using modular extreme learning machine. Neurocomputing. 2014;128:166–74.
49. Rong H-J, Ong Y-S, Tan A-H, Zhu Z. A fast pruned-extreme learning machine for classification problem. Neurocomputing. 2008;72:1–3.
50. Liu N, Wang H. Ensemble Based extreme learning machine. IEEE Signal Processing Letters. 2010;17(8):754–7.
51. Yu Q, van Heeswijk M, Miche Y, Nian R, He B, Séverin E, et al. Ensemble delta test-extreme learning machine (DT-ELM) for regression. Neurocomputing. 2014;129(10):153–8.
52. Cao J, Lin Z, Huang G-B, Liu N. Voting based extreme learning machine. Inf Sci. 2012;185(1):66–77.
53. Qureshi MNI, Min B, Jo HJ, Lee B. Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study. PLoS One. 2016;6(4):249–68.
54. Dai D, Wang J, Hua J, He H. Classification of ADHD children through multimodal magnetic resonance imaging. Front Syst Neurosci. 2012;6(63):1–8.
55. Oishi K, Faria AV, Yoshida S, Chang L, Mori S. Quantitative evaluation of brain development using anatomical MRI and diffusion tensor imaging. Int J Dev Neurosci. 2013;31(7):512–24.
56. Davatzikos C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. NeuroImage. 2004;23(1):17–20.
57. Chiapponi C, Piras F, Piras F, Fagioli S, Caltagirone C, Spalletta G. Cortical grey matter and subcortical white matter brain microstructural changes in schizophrenia are localized and age independent: a case-control diffusion tensor imaging study. PLoS One. 2013;8(10):e75115.