JCSE Journal of Computing Science and Engineering

Regular Paper

Journal of Computing Science and Engineering,
Vol. 17, No. 3, September 2023, pp. 135-143

An Efficient Attention Deficit Hyperactivity Disorder (ADHD) Diagnostic Technique Based on Multi-Regional Brain Magnetic Resonance Imaging

Vasily Sachnev*
School of Information, Communication and Electronics Engineering, Catholic University of Korea, Bucheon, Korea
bassvasys@hotmail.com

B. S. Mahanand
Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, JSS Science and Technology University, Mysuru, India
bsmahanand@gmail.com

Abstract
In this paper, an efficient technique for the diagnosis of attention deficit hyperactivity disorder (ADHD) was proposed. The proposed method used features/voxels extracted from structural magnetic resonance imaging (MRI) scans of seven brain regions and efficiently classified three subtypes of ADHD: ADHD-C, ADHD-H, and ADHD-I, as well as the typically developing control (TDC). Training and testing data for experiments were obtained from ADHD-200 database, and 41,721 features/voxels were extracted from sMRI by using region-of-interest (ROI). The proposed ADHD diagnostic technique built an efficient ADHD classifier in two steps. In the first step, a proposed regional voxels selection method (rVSM) selected an optimal set of features/voxels from seven brain regions available in ADHD-200, i.e., the Amygdala, Caudate, Cerebellar Vermis, Corpus Callosum, Hippocampus, Striatum, and Thalamus. In the second step, voxels/features selected by rVSM were used together to form a unified set of voxels. The unified set of voxels was used by a multi-region voxels selection method to train an efficient classifier using the extreme learning machine (ELM). Finally, the proposed method selected a unique set of voxels from the seven brain regions and built a final ELM classifier with maximum accuracy. Experiments clearly indicated that the proposed method produced better results than existing methods.

Category: Bioinformatics

Keywords: Attention deficit hyperactivity disorder; ADHD-200; MRI; Voxels selection method; Extreme learning machine

I. INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a widely spread neuropsychiatric disorder, which affects 5%–7% of people between 7 and 21 years of age [1]. Clinicians diagnose three main types of ADHD: hyperactive (ADHD-H), abnormal impulsive (ADHD-I), or combined (ADHD-C) by analyzing patients' behaviors. However,

Open Access http://dx.doi.org/10.5626/JCSE.2023.17.3.135 http://jcse.kiise.org

Received 03 September 2023; **Accepted** 09 September 2023
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

the origin of ADHD is not fully understood, and it may have genetic [2] and/or environmental factors associated with it. Originally, ADHD diagnosis was based on symptomatic behaviors. However, recent studies involved analyzing brain structures by using various imaging techniques such as magnetic resonance imaging (MRI) and single-photon emission computed tomography (SPECT) [3, 4]. These imaging techniques identify abnormal brain areas, and give more information about the physiological origin of ADHD.

Magnetic resonance imaging (MRI) provides accurate information about the shape and volume of the brain. Recent advances in ADHD research based on MRI discovered abnormalities in different brain regions. For example, subtle changes were identified in the amygdala and thalamus areas [5], cerebral volume [6], temporoparietal lobes, corpus callosum, basal ganglia and frontostriatal areas [7-9]. It is also reported that significant changes occur in the thalamus, amygdala, caudate, striatum, hippocampus, and basal ganglia brain regions. The brain regions associated with cognition and human emotions such as the amygdala, caudate and hippocampus [10-12] are extensively researched in the literature. According to recent studies, the amygdala accumulates emotions and memories associated with fear [12]. The caudate is resposible for many high-level functions, such as planning of movements, learning, memory, rewards, motivations, and emotions [13]. The hippocampus is responsible for long-term memory, releasing hormones, emotional response, and many others [14].

ADHD diagnosis based on MRI is widely studied and reported in the literature. Sachnev et al. [15] presented a cognitive neuro-fuzzy interface system (McFIS) to solve the four-class ADHD classification problem—typically developing control (TDC) vs. ADHD-C vs. ADHD-H vs. ADHD-I. Later, in [16], the authors improved classification performances by using a features selection method and voxels from the Hippocampus. Mahanand et al. [17] used a project-based learning meta-cognitive radial basis function network (PBL-McRBFN) method to analyze the amygdala and cerebellar vermis brain regions. Qureshi et al. [18] utilized an extreme learning machine (ELM) for ADHD diagnosis on a reduced number of samples. Dai et. al. [19] used multi kernel learning (MLK) to classify the three subtypes of ADHD. From the literature it is observed that amygdala, cerebellar vermis, and hippocampus play a significant role in ADHD; however, other brain areas may also play a significant role for the diagnosis of ADHD. Hence, a new ADHD diagnosis method based on the analysis of multiple brain regions need to be developed.

In this paper, a combination of seven brain areas (amygdala, caudate, cerebellar vermis, corpus callosum, hippocampus, striatum, thalamus) forms an extremely large features space (n=41,721). Building a classifier which can efficiently utilize such a large number of features is computationally intensive and also there may be a possibility of feature redundancy. Hence, a two-step features selection procedure, which can efficiently select a suitable set of features to build an efficient classifier, is presented in this paper. The proposed efficient ADHD diagnostic technique unifies (step 1) a "regional voxels selection method (rVSM)" and (step 2) a "multi-region voxels selection method (mrVSM)". The rVSM picks the voxels/features in each examined brain area (amygdala, caudate, cerebellar vermis, corpus callosum, hippocampus, striatum, thalamus) with probable influence on ADHD. The mrVSM uses information about the most significant voxels in each brain region to build a unified set of features, containing features from all seven examined brain regions. The unified set of features is used to build a final ADHD classifier for better classification performance.

The paper is organized as follows: Section II presents the ADHD-200 dataset used in this work. The proposed efficient ADHD diagnostic technique is presented in Section III. Section IV presents the experimental results. Finally, conclusions are drawn in Section V.

II. ADHD-200

ADHD-200 is a publicly available dataset consisting of both structural and functional MRI of ADHD patients and TDC [20]. Dataset consists of 581 TDC and 360 ADHD patients. Among the 360 ADHD patients, 210 have combined inattentive and hyperactive, 137 were inattentive and 13 were hyperactive. The 941 available samples were divided into 770 training samples and 171 testing samples. The training set contained 487 TDC, 161 ADHD-C, 11 ADHD-H, and 111 ADHD-I; comprising of 290 females and 480 males. The testing set contained 94 TDC, 49 ADHD-C, 2 ADHD-H, and 26 ADHD-I; comprising of 65 females and 106 males.

A. Region-of-Interest

The ADHD-200 consortium [20] suggested a few ways to preprocess the structural MRI scans. In this work, Burner pipeline was used which consisted of three major steps: (1) statistical parametric mapping (SPM) [21] separated gray matter and white matter; (2) Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) [16] normalizes data; and (3) each image was transformed into a space of population averages. Region-of-interest (ROI) method was used in this work to extract a set of voxels/features from each MRI scan. Pickatlas tool [22] was used to generate ROI masks. A total of 41,721 voxels/features were extracted from regions which include: amygdala (n=1,050), caudate (n=3,904), cerebellar vermis (n=6,358), corpus callosum

(n=8,536), hippocampus (n=6,076), striatum (n=9,359), and thalamus (n=6,438).

III. PROPOSED EFFICIENT ADHD DIAGNOSTIC TECHNIQUE

The proposed ADHD classifier was employed to efficiently classify four classes (TDC, ADHD-C, ADHD-I, ADHD-H) by using a large set of 41,721 voxels/features. Such classification problem was hard, and the choice of suitable classifier was a big challenge. Hence, a new systematic way to build a suitable ADHD classifier is needed.

In this work, a proposed efficient ADHD diagnostic technique was used to reduce the complexity of the classification problem by selecting a smaller set of more relevant voxels/features for better classification. Voxels/features reduction was processed in two steps:

Step 1: the rVSM reduced voxels/features in each brain region individually and

Step 2: the mrVSM, where voxels/features selected by rVSM were used as a unified set for further analysis (Fig. 1).

In the first step, all MRI scans from ADHD-200 were processed by ROI and voxels were extracted. The rVSM was used in each brain region separately to pick the top 10% of voxels from each brain region for further analysis. Details about the rVSM are presented in Section III-B.

In the second step the combined set of voxels selected by rVSM (i.e., amygdala, caudate, cerebellar vermis, corpus callosum, hippocampus, striatum, and thalamus) is used in the mrVSM to build a further unified set of voxels collected from all seven brain regions. The unified set was then used to build a final ADHD classifier by using the ELM [23]. Details about the mrVSM are presented in Section III-C.

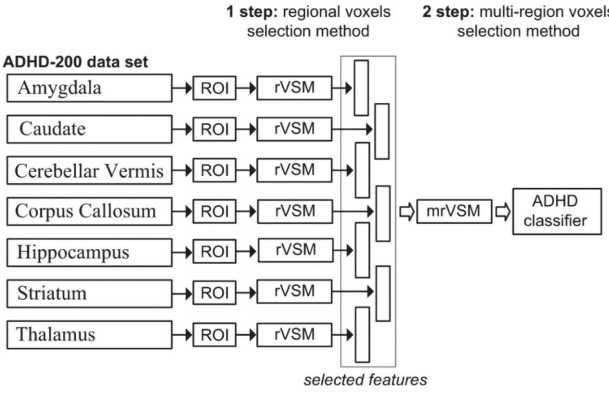

Fig. 1. Framework of the proposed efficient attention deficit hyperactivity disorder (ADHD) diagnostic technique.

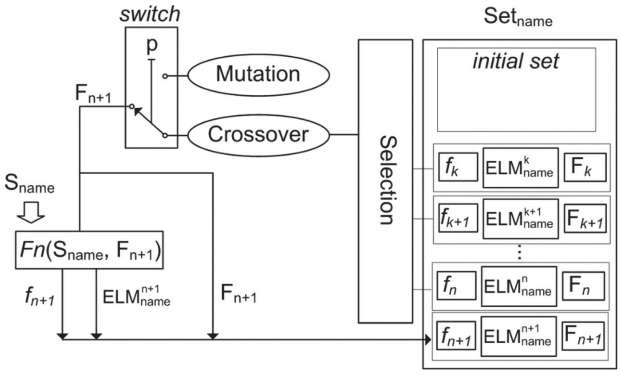

Fig. 2. Framework of the voxels selection method.

A. Voxel Selection Method

A voxel selection method or VSM is a key component of the proposed rVSM and mrVSM. The VSM is a simplified and improved version of the well-known genetic algorithm designed to select voxels/features responsible for ADHD. The presented VSM is an improved version of the binary coded genetic algorithm (BCGA) [24]. The VSM does not use generations, and all generated solutions may have chance to contribute further. Such improvement significantly speeds up the optimization process and can be implemented in a two-step design (Fig. 1). The framework of the VSM is presented in Fig. 2.

The proposed VSM produces a set of ELM classifiers (or ADHD classifiers), Set_{name}, by using a voxels/features set S_{name}:

$$Set_{name} = VSM(S_{name}), \quad (1)$$

where S_{name} is a set of voxels features extracted from MRI. Notation "name" links to amygdala (amy), caudate (cau), cerebellar vermis (cer), corpus callosum (cor), hippocampus (hip), striatum (str), thalamus (tha) for rVSM, and "unified" for mrVSM. Set_{name} collected all created ELM classifiers (or ADHD classifiers) for further analysis.

Voxels selection method VSM contains four main operations: (1) Fitness function (Fn), (2) Selection method, (3) Binary operators (Crossover and Mutation), and (4) Switch.

1) Fitness function

Fitness function or Fn selected voxels/features from S_{name}, (1) to form a set for training data, (2) to build an ELM classifier and (3) to compute measure (fitness value *f*). The detailed framework of the Fitness function is presented in Fig. 3. In Fn each voxel/feature from S_{name} has a corresponding binary coefficient "1" or "0" from a binary vector *F*. The coefficients "1" in *F* refer to selected

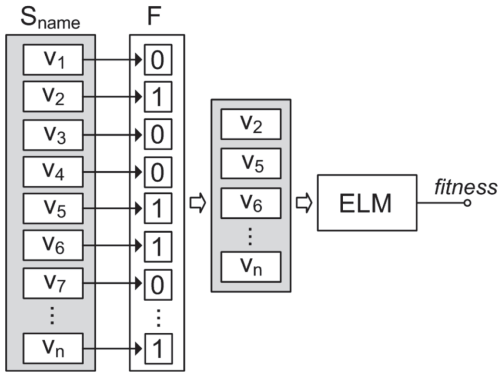

Fig. 3. Fitness function, *fitness* = *Fn*(*Set*$_{name}$, *F*).

features, whereas coefficients "0" refer to skipped features. Selected features were used to train a single ELM classifier. For example, in Fig. 3, features 2,5,6,..,*n* were selected because the binary vector *F* has coefficient "1" in these positions. The overall testing accuracy of the created ELM classifier is used as a fitness. Fitness function (Fn) produced the ELM classifier and corresponding fitness value.

2) Selection method

Selective method was designed to select two items from *Set*$_{name}$ for further processing. In the Selection method each item must have non-zero chance (probability) to be picked according to its fitness value. The items with higher fitness have higher chance, whereas the items with lower fitness have lower chance to be picked. The selected items were used in the Crossover to build another, maybe better, solution.

The proposed "Selection method" was based on the geometric ranking method [25]. The geometric ranking method sorts all items in descending order in respect of the corresponding fitness value (overall testing accuracy of the ELM classifier). The probability of selection for any item *j* is calculated as follows:

$$P_j = q'(1-q)^{r_j-1}, \quad (2)$$

where

$$q' = \frac{q}{1-(1-q)^N},$$

q' is a normalization parameter, r_j is the rank of the *j*-th item in the partially ordered set, and *N* is the population size. The detailed explanation of the geometric ranking method is given in [25]. In this research the parameter *q* is 10^{-3}.

The binary operators, crossover, and mutation, play a vital role in the proposed voxels selection method.

3) Crossover and Mutation

"Crossover" is a natural process that shares genes among generations, where the old generations transfer genes to a new generation with chance to achieve new properties for survival. Researchers try to mimic this natural crossover procedure in order to solve various engineering problems, where individuals with better fitness have higher chance to contribute to the next generation. In the proposed VSM, the crossover operator created one new binary vector *F* from two binary vectors selected from *Set*$_{name}$ by "Selection method."

The proposed "statistical sample balanced crossover" was designed to enhance the performance of the VSM. The framework of the proposed crossover is presented in Fig. 4. The statistical samples balanced crossover enabled a systematic way to mix genes within the VSM framework. In the first step, the statistical samples balanced crossover collected indices of the binary "1" in both binary vectors F_1 and F_2 (see index #1 and index #2 in Fig. 4). Next, the extracted indices formed a set "joint indices" for further steps. The joint indices may have duplicates if some features/voxels are present in both binary vectors (in Fig. 4, voxels #1, #5, #9 and #11 are duplicates). In the statistical samples balanced crossover each index from the set of joint indices was coupled with randomly generated value R:$R_i \in [0,\ 1]$. In the next step, a decision vector *D* was created, where $D_i = 1$ if corresponding random value $R_i > 0.5$ and $D_i = 0$ otherwise. Each $D_i = 1$ in the decision vector *D* defines a set of selected indices, which is [1,4,6,9,9,11] in the presented example (see Fig. 4). In the last step, all duplicated indices must be removed. After removing all duplicates, the final set of indices [1,4,6,7,9,11] formed a new binary vector *F*.

The proposed statistical sample balanced crossover was designed specifically for the VSM and has few significant advantages compare to other possible crossovers. First, the proposed crossover holds and processes only binary coefficients "1" which define the selected voxels/features. Second, a random decision strategy made a balanced choice of the binary coefficients "1." Third, the removing duplicates strategy created two possibilities: the duplicated index was either removed (index #5), or selected (index #1, #9, and #11) to form a new binary vector *F*. As a result, proposed statistical sample balanced crossover significantly speeds up the convergence of the voxels selection method and guarantees high performance.

"Mutation" is a second key component in the evolution of natural genes. Mutation makes random modification in the genes, which may or may not cause new properties. Mostly mutations were useless and cannot help individuals to survive, but sometimes (with very low probability) mutation may produce a massive and significant improvement. Such improvement was not possible to achieve by crossover since crossover manipulates only those genes available in the old generation and a novel gene with new properties cannot be created.

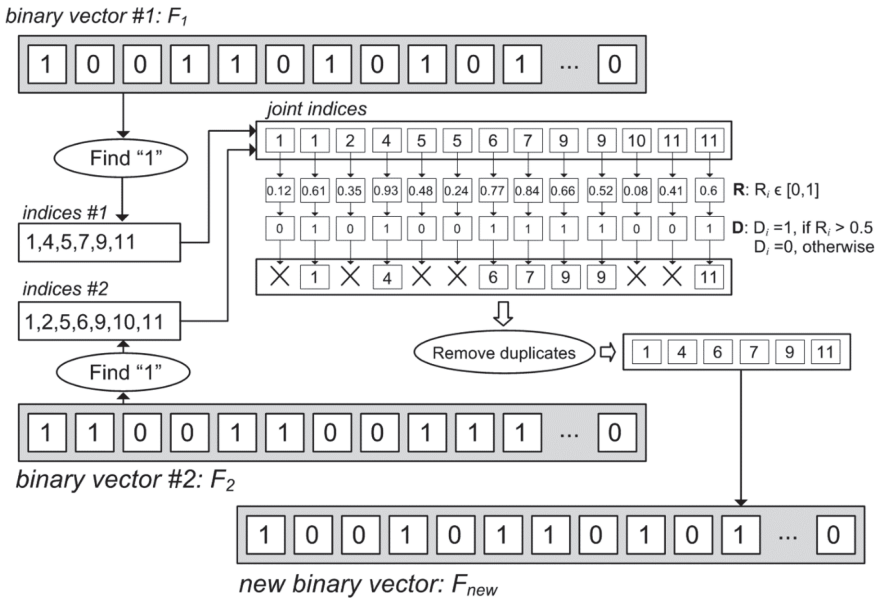

Fig. 4. Proposed statistical balanced samples crossover.

In the proposed VSM, "mutation" operator generates a binary vector F with 50–200 binary coefficients "1" randomly allocated.

4) Switch

"Switch" operator picks "Crossover" or "Mutation" with fixed probability p or $(1–p)$ respectively. In the proposed VSM p=0.8, which means a chance of crossover is 80%, and the chance of mutation is 20%. Every time when VSM generated a new solution, the "Switch" operator triggered a choice.

The VSM is an iterative process to create a single binary vector F (by using Crossover or Mutation) and build a new ELM classifier with corresponding fitness, per iteration. VSM started from creating an initial set of ELM classifiers, which formed an initial Set_{name}. Each VSM iteration started from the "Selection method", where 2 binary vectors F were selected. In the next step, the "Switch" picks "Crossover" or "Mutation". If "Crossover" was picked, a new binary vector F_{new} was created from the two binary vectors chosen by the Selection method (see details above). If "Mutation" was picked, a new binary vector F_{new} was created randomly. In the next step, the "Fitness function" created a new ELM classifier with corresponding fitness from the new binary vector F_{new}. The created ELM classifier, corresponding binary vector F and fitness were then inserted into the Set_{name}. The VSM stopped when the number of items in the Set_{name} reached a termination threshold.

B. Regional Voxels Selection Method

Roughly rVSM can be presented as follows:

$$Set_{name} = VSM(S_{name}),$$

$$S_{name}^{Top10} = Top10(Set_{name}),$$

where VSM is the VSM presented in Section III-A; S_{name} is a set of features extracted from the brain region "name" (i.e., amygdala, caudate, cerebellar vermis, corpus callosum, hippocampus, striatum, thalamus); $Set_{name} = \{[F_1, ELM_{name}^1, f_1], [F_2, ELM_{name}^2, f_2], ..., [F_n, ELM_{name}^n, f_n]\}$ is the set of ELM classifiers unified with corresponding binary vector F and fitness value f (see details in Section III-A); $Top10$ is a function to define the top 10% of ELM classifiers in terms of the fitness value f. Each S_{name}^{Top10} collects the corresponding voxels/features with a logical "1" in each F in the top 10%.

The initial set for rVSM was a set of 10 ELM classifiers with randomly generated binary vectors F, where 50–200 binary "1" were allocated randomly. The rVSM processed 100 VSM steps and generated a final set of 110 ELM classifiers. The rVSM was performed separately for all available brain regions, i.e., amygdala, caudate, cerebellar vermis, corpus callosum, hippocampus, striatum, and thalamus (or S_{amy}, S_{cau}, S_{cer}, S_{cor}, S_{hip}, S_{str}, S_{tal}). The function "$Top10$" selected 11 ELM classifiers with best training accuracy and formed a new set of features from their corresponding F vectors: S_{amy}^{Top10}, S_{cau}^{Top10}, S_{cer}^{Top10}, S_{cor}^{Top10}, S_{hip}^{Top10}, S_{str}^{Top10}, S_{tal}^{Top10}. The created sets of features were then used by the mrVSM.

The rVSM significantly reduced the total number of voxels/features for further analysis in mrVSM. Features reduction results are: the feature space for amygdala was reduced from 1,050 voxels to 275 voxels in S_{amy}^{Top10}; for

caudate from 3,904 voxels to 725 in S_{cau}^{Top10}; for cerebellar vermis from 6,358 voxels to 599 in S_{cer}^{Top10}; for corpus callosum from 8,536 voxels to 509 in S_{cor}^{Top10}; for hippocampus from 6,076 voxels to 180 in S_{hip}^{Top10}; for striatum from 9,359 voxels to 213 in S_{str}^{Top10}; for thalamus from 6,438 voxels to 348 in S_{tal}^{Top10}; giving a total of 2,849 voxels which forms $S_{unified}$.

The overall testing accuracy of the best ELM classifier for amygdala was 56.8%, for caudate 50.1%, for cerebellar vermis 60.3%, for corpus callosum 59.1%, for hippocampus 57.4%, for striatum 59.8%, and for thalamus 58.92%.

C. Multi-Region Voxels Selection Method

In the proposed efficient ADHD diagnostic technique, the mrVSM repeated voxel selection method VSM for set $S_{unified}$ created by rVSM with minor modifications. Roughly, mrVSM is presented as follows:

$$Set_{final} = VSM(S_{unified}),$$

$$M_{best} = Best(Set_{final}),$$

where VSM is the voxel selection method; $S_{unified} = S_{amy}^{Top10} \cup S_{cau}^{Top10} \cup S_{cer}^{Top10} \cup S_{cor}^{Top10} \cup S_{hip}^{Top10} \cup S_{str}^{Top10} \cup S_{tal}^{Top10}$ is a set of 2,849 voxels/features selected by rVSM from all seven examined brain regions, i.e., amygdala, caudate, cerebellar vermis, corpus callosum, hippocampus, striatum, thalamus; $Set_{final} = \{[f_1, ELM_{unified}^1, F_1], [f_2, ELM_{unified}^2, F_2], ..., [f_n, ELM_{unified}^n, F_n]\}$ is the set of ELM classifiers coupled with corresponding binary vector F and fitness value f (see details in Section III-A); function "Best" picks the ELM classifier with the best training efficiency M_{best}, which is the final ADHD classifier of the proposed method.

D. Extreme Learning Machine

The presented ADHD classification problem was sparse (41,721 features), multiclass (4 classes) and unbalanced (ADHD classes have very different number of samples). As a result, the choice of suitable machine learning techniques to solve the ADHD classification problem was limited. The proposed VSM implemented in rVSM significantly reduced the feature space from 41,721 to 2,849, which made the choice of the proper machine learning tool easier.

An ELM with Gaussian hidden neurons was used as a machine learning tool to solve the ADHD classification problem. The ELM is a classical one hidden layer feed-forward neural network with unique way to assign input and output neurons' weights. In ELM the input weights were assigned randomly, and output weights were calculated analytically [23]. The bias of the hidden neurons is assigned randomly.

In the ELM framework the training data is presented as follows: $\{(X^1, c^1), ..., (X^t, c^t), ..., (X^N, c^N)\}$, where vector X^t presents the m-dimensional input features of t-th sample, $c^t \in \{1, 2, 3, ..., C\}$ is a class label for multiclass classification problem. The class label is coded by values y^t as follows:

$$y_k^t = \begin{cases} 1, & if\ c^t = k \\ -1, & otherwise \end{cases} \qquad k = 1, 2, ..., C.$$

In the ADHD classification problem there four classes, or C=4. For example, in $y_k^t = [-1, 1, -1, -1]$, coefficient 1 links to the 2nd class, i.e., ADHD-C.

For further details about ELM please refer to [13].

IV. EXPERIMENTAL RESULTS

The proposed efficient ADHD diagnostic technique has been extensively tested and 941 subjects available in the ADHD-200 dataset were classified. The proposed ADHD diagnostic technique selected a set of features by using rVSM and mrVSM and builds an ELM classifier with optimum classification performance.

The classification performance of the proposed ADHD diagnostic technique was evaluated using the concept of confusion matrices. A confusion matrix is a matrix where each column links to predicted class label and each row links to actual class label. The confusion matrix size is: "Number of classes" × "Number of classes" or 4 × 4. Thus, correctly classified samples were allocated along the main diagonals, and incorrectly classified samples are allocated in other matrix cells. Training and testing confusion matrices are presented in Tables 1 and 2.

For example, the classification success or failure of the

Table 1. Training confusion matrix

	TDC	ADHD-C	ADHD-H	ADHD-I
TDC	279	113	8	86
ADHD-C	14	129	3	15
ADHD-H	0	0	11	0
ADHD-I	2	2	2	105

Table 2. Testing confusion matrix

	TDC	ADHD-C	ADHD-H	ADHD-I
TDC	66	14	1	13
ADHD-C	12	25	1	10
ADHD-H	1	0	0	1
ADHD-I	6	4	0	16

class 1 (or TDC) is presented in the first row of the "Training confusion matrix." The first number in the first-row links to correctly classified samples (where the actual class label is the same as the predicted class label), which equal to 279. The second number in the first row (113) was linked to misclassification, where the actual class label is 1 (TDC) but the predicted class label was 2 (ADHD-C). The third number in the first row (8) also linked to the misclassification, where the actual class label is 1 (TDC) but the predicted class label was 3 (ADHD-H). The fourth number in the first row (86) was also linked to the misclassification, where the actual class label is 1 (TDC) but the predicted class label is 4 (ADHD-I). The second, third and fourth rows present the classification distribution of ADHD-C, ADHD-H, and ADHD-I. Thus, the confusion matrix is an efficient and very representative way of displaying the classification performances of the machine learning tools for multiclass classification problems.

The performance of the proposed ADHD diagnostic technique has been evaluated by overall training accuracy η^{train} and testing accuracies η^{test}:

$$\eta^{test} = \frac{1}{N_{test}} \sum_{i=1}^{4} n_i^{test} \times 100\% ,$$

$$\eta^{train} = \frac{1}{N_{train}} \sum_{i=1}^{4} n_i^{train} \times 100\% ,$$

where n_i^{test} and n_i^{train} are number of correctly classified samples of each class i; N_{test} = 171 and N_{train} = 770 are total number of testing and training samples.

Overall training accuracy is η^{train} = 68.14%.

Overall testing accuracy is η^{test} = 66.08%.

The proposed ADHD diagnostic technique chose 511 voxels from seven examined brain regions: amygdala (11 voxels), caudate (30), cerebellar vermis (81), corpus callosum (108), hippocampus (77), striatum (140), and thalamus (64).

A. Comparison with Existing Methods

The ADHD classification problem is well studied in the literature. Researchers often conduct experiments on the dataset by unifying all ADHD types into one, which turns the multiclass ADHD classification problem into a two-class problem: ADHD versus TDC. Another way to simplify data is to reduce the number of samples, where a fewer number of samples may result in higher classification accuracy. However, the four-class classification (ADHD-I vs. ADHD-H vs. ADHD-C vs. TDC) was much harder to solve and required the development of more sophisticated machine learning techniques.

In [15], authors used the Genetic Algorithm coupled with the meta-cognitive neuro-fuzzy interface system (McFIS) for accurate diagnosis of three types of ADHD and TDC. Authors used all available samples for analysis and reported a testing accuracy of around 56%. Sachnev and Suresh [24] introduced an efficient feature searching approach and ADHD classifier to solve the four-class ADHD classification problem with all samples available in the ADHD-200 dataset. Testing accuracy was reported as 58.6%. Authors in [15] and [24] used features from the Hippocampus only for experiments. In [26], authors extracted features from the cingulate cortex, prefrontal cortex, and visual cortex brain regions for reduced set of samples and build an ADHD classifier to solve the four-class classification problem by using a well-known deep learning technique. Authors reported testing accuracy of around 35.19%.

Qureshi et al. [18] solved the three-class classification problem (i.e., TDC vs. ADHD-I vs. ADHD-C) for a reduced number of samples (159 patients). Authors reported testing accuracy of about 60.78%. Later, Sachnev et al. [27] used multi-region risk-sensitive cognitive ensembler to solve the same three-class classification problem with all available ADHD-200 samples. Authors reported testing accuracy of about 86%.

Fair comparison is possible only if methods have same initial data to build the classifier, such as all available samples from ADHD-200 database (770 samples for training and 171 samples for testing), and the same classes for classification, such as the four classes (ADHD-I vs. ADHD-H vs. ADHD-C vs. TDC). Thus, comparison is possible only with methods presented in [15] and [24]. The proposed efficient ADHD diagnostic approach showed a 10 % and 7.4% improvement in terms of overall testing accuracy compared to [15] and [24], respectively. Methods of [18, 26] used reduced number of samples, and experiments with the complete set of samples are more challenging and result in a lower classification accuracy. Methods of [18, 27] excluded the ADHD-H samples from the database and solved the simplified ADHD classification problem (TDC vs. ADHD-I vs. ADHD-C), which is much easier than the four-class classification problem (i.e., TDC vs. ADHD-I vs. ADHD-H vs. ADHD-C).

V. CONCLUSION

The proposed efficient ADHD diagnostic technique efficiently classified three types of ADHD and TDC. The proposed rVSM and mrVSM selected a total of 2,849 features from a large feature space of 41,721. The reduced set of features was then used to build an efficient ADHD classifier by employing ELM. The proposed VSM identified the significant voxels which may be responsible for ADHD from the seven brain regions which included amygdala (11 voxels), caudate (30), cerebellar vermis (81), corpus callosum (108), hippocampus

(77), striatum (140), and thalamus (64). Experiment results indicated the better performance of the proposed ADHD diagnostic technique in comparison to existing techniques presented in the literature.

ACKNOWLEDGMENTS

This work was supported by The Catholic University of Korea, Research Fund 2020.

Conflict of Interest(COI)

The authors have declared that no competing interests exist.

REFERENCES

1. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association, 2000.
2. T. Banaschewski, K. Becker, S. Scherag, B. Franke, and D. Coghill, "Molecular genetics of attention-deficit/hyperactivity disorder: an overview," *European Child & Adolescent Psychiatry*, vol. 19, pp. 237-257, 2010. https://doi.org/10.1007/s00787-010-0090-z
3. H. Schneider, J. F. Thornton, M. A. Freeman, M. K. McLean, M. J. van Lierop, and J. Schneider, "Conventional SPECT versus 3D thresholded SPECT imaging in the diagnosis of ADHD: a retrospective study," *The Journal of Neuropsychiatry and Clinical Neurosciences*, vol. 26, no. 4, pp. 335-343, 2014. https://doi.org/10.1176/appi.neuropsych.12110280
4. D. G. Amen, T. A. Henderson, and A. Newberg, "SPECT functional neuroimaging distinguishes adult attention deficit hyperactivity disorder from healthy controls in big data imaging cohorts," *Frontiers in Psychiatry*, vol. 12, article no. 725788, 2021. https://doi.org/10.3389/fpsyt.2021.725788
5. I. Ivanov, R. Bansal, X. Hao, H. Zhu, C. Kellendonk, L. Miller, et al., "Morphological abnormalities of the thalamus in youths with attention deficit hyperactivity disorder, *American Journal of Psychiatry*, vol. 167, no. 4, pp. 397-408, 2010. https://doi.org/10.1176/appi.ajp.2009.09030398
6. M. V. Cherkasova and L. Hechtman, "Neuroimaging in attention-deficit hyperactivity disorder: beyond the frontostriatal circuitry," *The Canadian Journal of Psychiatry*, vol. 54, no. 10, pp. 651-664, 2009. https://doi.org/10.1177/070674370905401002
7. J. N. Giedd and J. L. Rapoport, "Structural MRI of pediatric brain development: what have we learned and where are we going?," *Neuron*, vol. 67, no. 5, pp. 728-734, 2010. https://doi.org/10.1016/j.neuron.2010.08.040
8. A. M. H. Onnink, M. P. Zwiers, M. Hoogman, J. C. Mostert, C. C. Kan, J. Buitelaar, and B. Franke, "Brain alterations in adult ADHD: effects of gender, treatment and comorbid depression," *European Neuropsychopharmacology*, vol. 24, no. 3, pp. 397-409, 2014. https://doi.org/10.1016/j.euroneuro.2013.11.011
9. E. Perlov, A. Philipsen, L. T. van Elst, D. Ebert, J. Henning, S. Maier, E. Bubl, and B. Hesslinger, "Hippocampus and amygdala morphology in adults with attention-deficit hyperactivity disorder," *Journal of Psychiatry and Neuroscience*, vol. 33, no. 6, pp. 509-515, 2008.
10. H. Suzuki, K. N. Botteron, J. L. Luby, A. C. Belden, M. S. Gaffrey, C. M. Babb, et al, "Structural-functional correlations between hippocampal volume and cortico-limbic emotional responses in depressed children," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 13, pp. 135-151, 2013 https://doi.org/10.3758/s13415-012-0121-y
11. V. Vuontela, S. Carlson, A. M. Troberg, T. Fontell, P. Simola, S. Saarinen, and E. T. Aronen, "Working memory, attention, inhibition, and their relation to adaptive functioning and behavioral/emotional symptoms in school-aged children," *Child Psychiatry & Human Development*, vol. 44, pp. 105-122, 2013. https://doi.org/10.1007/s10578-012-0313-2
12. K. J. Ressler, "Amygdala activity, fear, and anxiety: modulation by stress," *Biological Psychiatry*, vol. 67, no. 12, pp. 1117-1119, 2010. https://doi.org/10.1016/j.biopsych.2010.04.027
13. C. A. Seger and C. M. Cincotta, "The roles of the caudate nucleus in human classification learning," *Journal of Neuroscience*, vol. 25, no. 11, pp. 2941-2951, 2005. https://doi.org/10.1523/JNEUROSCI.3401-04.2005
14. K. S. Anand and V. Dhikav, "Hippocampus in health and disease: an overview," *Annals of Indian Academy of Neurology*, vol. 15, n. 4, pp. 239-246, 2012. https://doi.org/10.4103/0972-2327.104323
15. V. Sachnev, "An efficient classification scheme for ADHD problem based on binary coded genetic algorithm and McFIS," in *Proceedings of 2015 International Conference on Cognitive Computing and Information Processing (CCIP)*, Noida, India, 2015, pp. 1-6. https://doi.org/10.1109/CCIP.2015.7100690
16. J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95-113, 2007. https://doi.org/10.1016/j.neuroimage.2007.07.007
17. B. S. Mahanand, R. Savitha, and S. Suresh, "Computer aided diagnosis of ADHD using brain magnetic resonance images," in *AI 2013: Advances in Artificial Intelligence*. Cham, Switzerland: Springer, 2013, pp. 386-395. https://doi.org/10.1007/978-3-319-03680-9_39
18. M. N. I. Qureshi, B. Min, H. J. Jo, and B. Lee, "Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study," *PloS One*, vol. 11, no. 8, article no. e0160697, 2016. https://doi.org/10.1371/journal.pone.0160697
19. D. Dai, J. Wang, J. Hua, and H. He, "Classification of ADHD children through multimodal magnetic resonance imaging," *Frontiers in Systems Neuroscience*, vol. 6, article no. 63, 2012. https://doi.org/10.3389/fnsys.2012.00063
20. ADHD-200 Consortium, "The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience," *Frontiers in Systems Neuroscience*, vol. 6, article no. 62, 2012. https://doi.org/10.3389/fnsys.2012.00062
21. K. J. Friston, S. J. Kiebel, T. E. Nichols, and W. D. Penny, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. San Diego, CA: Academic Press, 2007.
22. J. A. Maldjian, P. J. Laurienti, R. A. Kraft, and J. H. Burdette, "An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets," *Neuroimage*, vol. 19, no. 3, pp. 1233-1239, 2003. https://doi.org/10.1016/S1053-8119(03)00169-1

23. G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 2006. https://doi.org/10.1016/j.neucom.2005.12.126
24. V. Sachnev and S. Suresh, "An ADHD diagnostic approach based on binary-coded genetic algorithm and extreme learning machine," *Journal of Computing Science and Engineering*, vol. 10, no. 4, pp. 111-117, 2016. https://doi.org/10.5626/JCSE.2016.10.4.111
25. S. Suresh, S. N. Omkar, V. Mani, and T. G. Prakash, "Lift coefficient prediction at high angle of attack using recurrent neural network," *Aerospace Science and Technology*, vol. 7, no. 8, pp. 595-602, 2003. https://doi.org/10.1016/S1270-9638(03)00053-1
26. D. Kuang, X. Guo, X. An, Y. Zhao, and L. He, "Discrimination of ADHD based on fMRI data with deep belief network," in *Intelligent Computing in Bioinformatics*. Cham, Switzerland: Springer, 2014, pp. 225-232. https://doi.org/10.1007/978-3-319-09330-7_27
27. V. Sachnev, S. Suresh, N. Sundararajan, B. S. Mahanand, M. W. Azeem, and S. Saraswathi, "Multi-region risk-sensitive cognitive ensembler for accurate detection of attention-Deficit/Hyperactivity disorder," *Cognitive Computation*, vol. 11, pp. 545-559, 2019. https://doi.org/10.1007/s12559-019-09636-0

Vasily Sachnev https://orcid.org/0000-0001-7063-5069

Vasily Sachnev received his B.S. and M.S. degrees in electrical engineering from the Komsomolsk-na-Amure State Technical University, Russia, in 2002 and 2004, Ph.D. degree in Multimedia Security Laboratory at the Center of Information Security and Technology (CIST), Graduate School of Information Management and Security, Korea University, Seoul, Korea in 2009. Since 2010, he is a faculty member in the Catholic University of Korea. His research interests include machine learning, bioinformatics, digital watermarking, steganography, and image processing.

B. S. Mahanand https://orcid.org/0000-0001-8189-6889

B. S. Mahanand is a professor and Head of the Department of Information Science and Engineering; he carries 20 years of teaching and research experience at Sri Jayachamarajendra College of Engineering (SJCE), JSS Science and Technology University, Mysuru, India. He was Raman post-doctoral fellow at Harvard Medical School, Harvard University, USA. Dr. Mahanand has published more than 50 research articles in various peer-reviewed journals and reputed conferences. His research interests are in the areas of artificial intelligence, machine learning and medical image analysis.