Importance of phenotypic information in ADHD diagnosis

B. Rangarajan
School of Computer Engineering
Nanyang Technological University
Singapore
Email: badrinarayanan@ntu.edu.sg

K. Subramaian
School of Computer Engineering
Nanyang Technological University
Singapore
Email: skartick@ntu.edu.sg

S. Suresh
School of Computer Engineering
Nanyang Technological University
Singapore
Email: ssundaram@ntu.edu.sg

***Abstract*—Attention Deficit Hyperactivity Disorder (ADHD) is one of the widely researched neuro-developmental disorder. This paper highlights the importance of phenotypic information in the diagnosis of ADHD, in addition to Magnetic Resonance Image (MRI) based features. In this study, features from amygdala region of the brain is extracted using region of interest based feature extraction technique. These features are used in combination with phenotypic information for effective discrimination of ADHD from typically developing controls. Further, to improve the effectiveness of the detection system, SBMLR based feature selection technique is used to select the most discriminative features. Performance of the proposed ADHD diagnosis is evaluated on a benchmark ADHD-200 consortium database. The performance evaluation on two state-of-the-art classification techniques clearly highlight the importance of phenotypic information in the detection of ADHD.**

***Index Terms*—Adhd, phenotypes, amygdala, meta-cognitive radial basis function network, extreme learning machines, region of interest.**

I. Introduction

Statistics show that more than 5% of the children worldover are getting affected by the Attention Deficit Hyperactivity Disorder (ADHD) [14], [23]. ADHD is a developmental disorder that could be better controlled when treated at an early age. Though there is no evidence for the cause of ADHD, it has been shown in the literature that multiple factors such as other cognitive disorders, age, gender, genetics, and environment might contribute to this disorder [6].

Studies indicate that there exist relation between the volume of brain and presence of ADHD [10]. As a result, structural MRI based approaches have been used widely in early detection of ADHD and other cognitive disorders. In the neurobiology of pediatric cognitive disorders, it has been shown that several regions of the brain gets affected due to ADHD [9]. Among these regions, 'amygdala' region has been widely used in detection of ADHD [22], [7], [12]. This is due to the fact that amygdala which controls the emotional learning in the brain and its adaptivity is the key to understanding attention [24], [29], [20]. Hence in this study, we employ features extracted from amygdala region of the brain in detection of ADHD. In order to extract features from amygdala, a region of interest based feature extraction technique is employed [16]. Further, to identify the most discriminant feature, a Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation (SBMLR) [8] based feature selection technique is employed.

Recently, the MRI data has been extensively studied using several machine learning approaches for several cognitive disorders. Mahanand B.S. et.al uses a region of interest based feature extraction technique of MRI and classified them using PBL-McRBFN in [16]. In the pilot study, two regions were taken (amygdala and cerebellar vermis) from the ADHD-200 data set. Badrinarayanan et. al. uses PBL-McRBFN to indentify the potential biomarkers in the hippocampus region in order to diagnose ADHD. The study focuses on choosing few features in the hippocampus using Chi2 based features selection technique and classifies them using PBL-McRBFN and SVM. The comparative study shows that PBL-McRBFN outperforms SVM in identifying the potential biomarkers in the hippocampus region. In [28], Vigneshwaran et. al. presents an approach to diagnose autism spectrum disorder using voxel based morphometry based features and classified them using PBL-McRBFN classifier. In [17], Mahanand et. al. uses Integer-Coded Genetic Algorithm (ICGA) to select the features and self-adaptive resource allocation network (SRAN) to classify the alzheimer's disease. A comparative study with SVM and ELM is presented and the study shows that the proposed ICGA-SRAN helps in better classification. In [27], Tenev A. et.al proposes a machine learning approach for classifying adults diagnosed with ADHD. Here, two resting conditions and two neuropsychological tasks are used for classification. Support vector machine is used to classify these four different set of samples separately and combined using a voter function which is derived from a Karnaugh map. Qui M. et.al has highlighted the changes of brain structure and function in ADHD children in [25]. In this study, 15 ADHD patients and 15 control participants were taken into consideration. Functional MRI was performed and anatomical volume of the white matter is studied. Further, a short review on structural and function MRI studies in ADHD is presented which highlights the importance of volume in diagnosing ADHD. In [21], Peng X. et.al. studies the classification of ADHD using MRI data using ELM. This study uses the samples from Peking University imaging site of the ADHD-200 data set. Upon selecting the features, using sequential feature selection, the samples were classified using extreme

978-1-4799-7171-8/15/$31.00 ©2015 IEEE

learning machine and support vector machines.

Few studies in literature uses phenotypic information in ADHD diagnosis. Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD is studied in [1]. This study eliminates the samples that map to hyperactive patients. Kernel Principal Component Analysis for dimensionality reduction in fMRI-based diagnosis of ADHD is studied in [26]. Resting state fMRI of thalamus region coupled with the phenotypic information is studied in [19].

In addition to MRI based information, other environmental factor might be helpful in identifying ADHD, as shown by [6].Hence, in this study, in addition to structural MRI, phenotypic information, which provies behavioral features and other related information, is considered. These information include demographics, age, gender, presence of other disorders, behavioral measures, and intelligence quotient.

In order to evaluate the advantage of including phenotypic information, ADHD-200 consortium data set is considered. The data set provides 941 samples containing both ADHD and Typically Developing Controls (TDC). However, only a subset of this data set contain phenotypic information and only those samples are chosen in this study. Six different phenotypic information are considered: diagnostic status, dimensional ADHD symptom measures, age, sex, IQ and lifetime medication status. More information on phenotypic data selection is presented in the Section II-C. To evaluate the performance, two state-of-the-art classifiers are employed: meta-cognitive radial basis function network (PBL-McRBFN) [4] and extreme learning machine (ELM) [13]. Firstly, the features extracted from amygdala alone are considered and in the second test, the phenotypic information appended with amygdala are considered. The performance comparison clearly highlights around 30% increase in classification accuracy upon considering phenotypic information.

This paper is organized as follows. In the section II, the ADHD-200 data set is described, followed by the feature extraction technique and the description of the classifiers. Experimental results are evaluated in the sections III-A and III-B. Sections IV and V discusses few highlights of the paper and conclusions from the study.

II. Methods and Materials

This section starts with data set description (Ref: II-A) followed by region of interest based feature extraction (Ref: II-B), Phenotypic data selection (Ref: II-C) and SBMLR based feature selection (Ref: II-D). Finally, we explain the overview of ELM (Ref: II-E) and PBL-McRBFN (Ref: II-F) classifiers.

A. Dataset

The data used in paper are publicly available in the NITRC ADHD-200 consortium [18]. This data set is a collection 941 subjects. It includes many phenotypic information including age, gender, dexterity, IQ measure, medical co-morbidity status. The classifications are provided below.

- Typically developing control (TDC) or normal subjects.
- ADHD = Combination of ADHD-Combined, ADHD-Hyperactive/Impulsive and ADHD-Inattentive.

Not all the regions of the brain and not all the voxels of the regions contribute to the abnormality. The main objective is to extract the appropriate features, develop a model and classify them to a particular class label. In this study, we have combined all the sub-types and classifying them as 'ADHD'.

B. Data Preprocessing and region of interest based feature extraction

All MRI data were processed with the Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) [2] based on the Statistical Parametric Mapping (SPM) [11] software package in the Burner pipeline from the ADHD-200 consortium. The Burner pipeline includes normalized grey matter maps generated using SPM. There are three steps in Burner pipeline. Firstly, T1 images are segmented into grey matter and white matter probability maps using SPM. Images are then iteratively registered to the group average which is updated iteratively. Lastly, the registration parameters are applied to each image to transform each image into the space of population average. Modulation is applied to conserve the global tissue volumes after normalization. Voxel Based Morphometry (VBM) is used to segment the patient's MRI. When combined with the MRI templates and Region Masks, these segments are transformed into voxels of the region that is considered. We have extracted Amygdala [15] region of interest in this study. Please refer to [16] to understand more about region of interest based feature extraction.

C. Phenotypic data selection

Since this study claims to show the importance of phenotypic information in ADHD diagnosis, we have chosen those samples that has the phenotypic information populated by the ADHD-200 data set. Among the 941 samples, the phenotypic information is given only for 502 samples. The phenotypic information considered are gender, age, handedness, ADHD measure, ADHD index, inattentive measure, hyper/impulsive measure, IQ measure, verbal IQ, performance IQ and Full4IQ. All these information are provided as a part of the ADHD-200 dataset. We have used a stratified 10-fold cross validation to support our hypothesis by assigning 75% of the samples for training and 25% for testing which maps to 377 samples for training and 125 samples for testing.

D. Feature selection using Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation (SBMLR)

SBMLR [8] is a state-of-the-art feature selection method which realizes a sparse feature selection by using a Laplace prior. Here, the logistic regression is used to build a regularized regression model. SBMLR is shown to provide better classification accuracy with biomedical data [8]. Also, SBMLR is shown to improve the generalization performance [30]. In this study, we have employed SBMLR to select the most discriminative features of amygdala.

E. ELM Classifier Overview

ELM [13] is a single hidden-layer feed-forward neural network. The hidden neuron has bias and input weights. They are randomly assigned and the output weights are computed analytically. Given a training data set with N samples, $\{(X^1, c^1), ..., (X^t, c^t), ..., (X^N, c^N)\}$, where $X^t \epsilon \Re^m$ are the m-dimensional input features of t sample and $c^t \epsilon \{1, 2, ..., C\}$ is its class label. The coded class label $\mathbf{y}^t$ are obtained using:

$$\mathbf{y}_k^t = \begin{cases} 1, & \text{if } c^t = k \\ -1, & \text{otherwise} \end{cases} \quad k = 1, 2, ..., C \tag{1}$$

The classification problem is assumed as approximating the decision function (F) that maps the input features as the coded class labels, i.e., $F : \Re^m \rightarrow C^c$. The neurons in the initial and the terminating layers are linear, while the hidden layer has the Gaussian activation function. The ELM used to solve the problem has m input neurons and C output neurons. Let the number of hidden neurons be L. Then, the predicted class label y_k^t of the k-th hidden neuron of the t-th sample is computed as:

$$\hat{\mathbf{y}}_k^t = \sum_{j=1}^{L} \beta_{kj} . h_j^t, k = 1, ..., C \tag{2}$$

where β_{kj} is the weight connecting the j (hidden neuron) and k (output neuron). Equation 2 can be written in the matrix form as $\hat{Y} = \boldsymbol{\beta} H$, where H is the hidden-layer output matrix as shown below:

$$H(\mathbf{A}, B, X) = \begin{pmatrix} G(\mathbf{a}_1, b_1, X_1^t) & \cdots & G(\mathbf{a}_1, b_1, X_N^t) \\ \vdots & \vdots & \vdots \\ G(\mathbf{a}_L, b_L, X_1^t) & \cdots & G(\mathbf{a}_L, b_L, X_N^t) \end{pmatrix} \tag{3}$$

where, $A = \{a_1 \dots a_L\}$ is the weight vector, $B = \{b_1 \dots b_L\}$ is the threshold and $X = \{x_1 \dots x_L\}$ is the input. Thus, for a given training sample set $\{(X^1, \mathbf{y}^1), ..., (X^N, \mathbf{y}^N)\}$, and the number of hidden neurons L, the learning algorithm of ELM is summarized below:

- Randomly generate hidden neuron centers, $\mathbf{A} \epsilon \Re^{L \times m}$ and the Gaussian width of the hidden neurons, $B \epsilon \Re^{(L \times l)}$
- Compute the hidden-layer output matrix $\mathbf{H}^\dagger$.
- Calculate the output weights: $\boldsymbol{\beta} = \mathbf{Y}\mathbf{H}^\dagger$ where $\mathbf{H}^\dagger$ is the Moore-Penrose generalized inverse of responses of the hidden neurons.

However, the performance of the ELM is dependent on the randomly chosen centers and hidden neuron bias, especially in applications like ADHD detection, where the dimension of the input feature is large.

F. PBL-McRBFN Classifier Overview

Consider a set of training data samples, where $\mathbf{x}^t \in \Re^m$ is an m-dimensional input of the t^{th} sample, and $c^t \in (1, n)$ is its class label, and n is the total number of classes. McRBFN architecture is same as in [3] and has both cognitive component and the meta-cognitive component. We discuss these components in the following sections.

1) Cognitive component of McRBFN: The cognitive component of McRBFN is a radial basis function network. The hidden layer of the network holds the Gaussian activation function. For a given input $\mathbf{x}^t$, the predicted output $\widehat{y}_j^t$ is given as

$$\widehat{y}_j^t = \sum_{k=1}^{K} w_{kj} h_k^t, \qquad j = 1, \cdots, n \tag{4}$$

where w_{kj} is the weight connecting the k^{th} hidden neuron to the j^{th} output neuron and h_k^t is the response of the k^{th} hidden neuron to the input $\mathbf{x}^t$ is given by

$$h_k^t \quad = \quad exp\left(-\frac{\|\mathbf{x}^t - \mu_k^l\|^2}{(\sigma_k^l)^2}\right) \tag{5}$$

where $\mu_k^l \in \Re^m$ is the center and $\sigma_k^l \in \Re^+$ is the width of the k^{th} hidden neuron. Here, the superscript l represents the corresponding class of the hidden neuron.

Projection based learning algorithm helps to calculate the optimal network output parameters. For t consecutive samples, the error function is

$$J(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^{t} \sum_{j=1}^{n} \begin{cases} 0 & \text{if } y_j^i \widehat{y}_j^i > 1 \\ \left(y_j^i - \widehat{y}_j^i\right)^2 & \text{otherwise} \end{cases} \tag{6}$$

The resultant optimal output weights $\mathbf{W}^*$ corresponding to the error function $J(\mathbf{W}^*)$ is represented as

$$\sum_{k=1}^{K} a_{kp} w_{kj} \quad = \quad b_{pj}, \tag{7}$$

which is a system of linear equation in the form of $\mathbf{W}^* = \mathbf{A}^{-1}\mathbf{B}$. Here, a_{kp} is the projection matrix and b_{pj} is the output matrix. The derivation of this equation can be found in [3].

2) Meta-cognitive component of McRBFN: The meta-cognitive component uses various learning strategies and self-regulated thresholds to model the dynamics of the cognitive component. It helps to monitor and control the cognitive component. The meta-cognitive component uses predicted class label ($\widehat{c}^t$), maximum hinge loss (E^t), confidence of classifier ($\hat{p}(c^t|\mathbf{x}^t)$) and class-wise significance (ψ_c) as the measure of knowledge in the new training sample. Four different learning strategies are used namely; sample delete, neuron growth, sample reserve and parameter update. These strategies specializes in improving the generalization ability of the classifier. Please refer to the neuron growth architecture on [3] to further understand the output weights and the learning strategies of PBL-McRBFN.

In the next section, we evaluate the performance of amygdala with and without the phenotypic information and show the importance of phenotypes in diagnosing ADHD.

III. Experimental Results

In this section, we study the performance of ELM and PBL-McRBFN classifier in distinguishing ADHD patients from Typically Developing Controls (TDC). Figure 1 shows the schematic of the ADHD diagnosis including phenotypic information of the samples provided by the ADHD-200 data

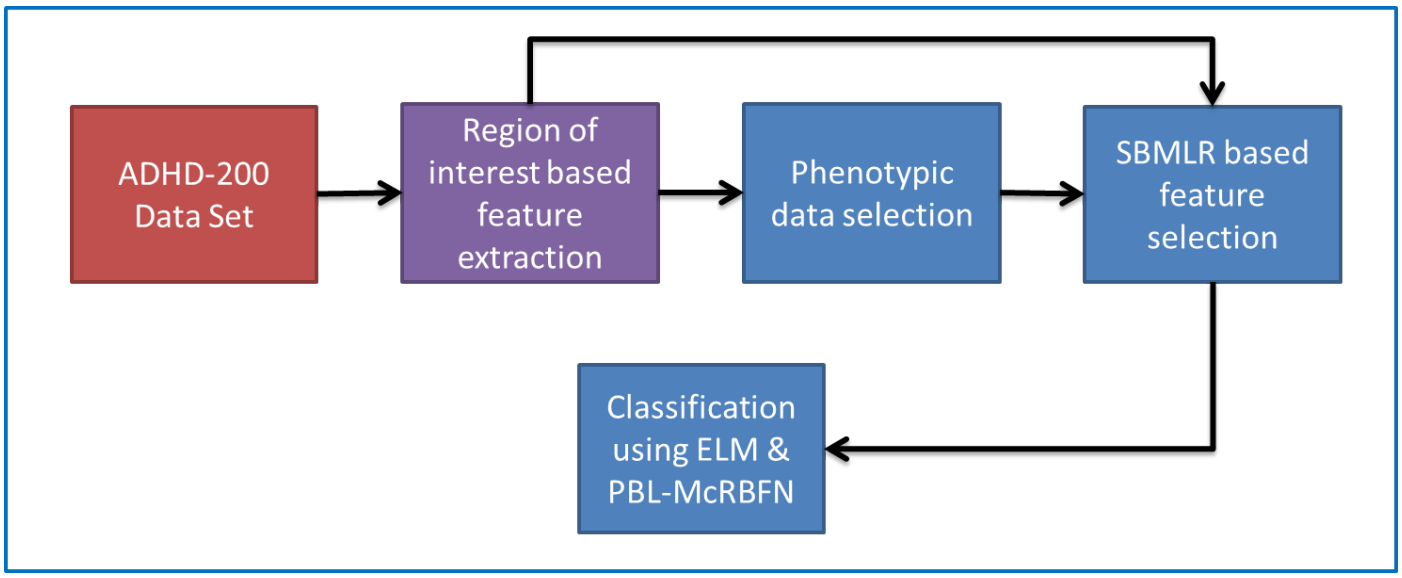

Fig. 1. Schematic of ADHD diagnosis including phenotypic information

set. As shown in the schematic, the process starts with data preprocessing and region of interest based feature extraction followed by phenotypic data selection. Refer II-C for more information on data selection. Once the region based features are extracted and phenotypic data are selected, we employ the Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation based feature selection technique to the 502 samples and obtained four most discriminative features. Initially, we use only these four features that represent amygdala to classify as ADHD or TDC and finally we append the 11 phenotypic information as features to classify the samples as ADHD and TDC. For our analysis, we have created a stratified 10-fold validation of the ADHD-200 data set with phenotypic information which provides 377 samples for training and remaining 125 samples for testing. In the section III-A, we evaluate the performance of PBL-McRBFN and ELM for the 10-fold data without the phenotypes. Then in the section III-B, we repeat the same experiment with the phenotypes appended to the feature set.

A. *Classification of ADHD without phenotypes*

In this section, we evaluate the classification accuracy of amygdala region without phenotypic information. The table I shows the 10-fold cross validation of amygdala using ELM and PBL-McRBFN. It shows that the overall testing accuracy ranges from 57% to 64% using both ELM and PBL-MCRBFN.

B. *Classification of ADHD with phenotypes*

In this section, we evaluate the performance of amygdala using the phenotypic information. Here, the MRI features is appended with the phenotypic information in order to test our claim. It is clear from the table II that by including the phenotypes, the overall testing accuracies of both ELM and PBL-McRBFN outperforms the analysis conducted in the section III-A by almost 30%.

IV. Discussion

[5]. The phenotypes considered here includes age, gender, IQ measures and ADHD measures provided by the ADHD-200 data set. These not only helps us to extend our analysis by accurately diagnosing ADHD from TDC, but it also helps in understanding the diagnosis as a whole. In order to showcase the importance of phenotypic information, this study considers the features of amygdala alone initially. Then the study is extended by appending the phenotypic information to the features (amygdala). Section III-A provides us the diagnosis of ADHD using amygdala taken as a region-of-interest. It shows that the average testing accuracy of PBL-McRBFN and ELM does not go beyond 61.1% and 59.01% respectively. From the section III-B, it shows that the average testing accuracy of PBL-McRBFN and ELM considering the features of amygdala and the phenotypic information reaches 97.54% and 98.28% respectively. This increase in the classification accuracy clearly indicates the importance of phenotypic information in diagnosing ADHD. We would like to further enhance this study by including other regions and create a group-wise analysis of patients with respect to phenotypes.

V. Conclusion

In this study, we show the importance of phenotypic information in the diagnosis of ADHD. Using Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation, we have selected the most discriminative features of 'amygdala' from the ADHD-200 data set (which is the chosen region of interest in our study). A ten-fold cross validation study of amygdala with and without phenotypes using ELM and PBL-McRBFN is conducted. Results show that, there is 30% increase in the overall testing accuracy of diagnosis that include phenotypic information. This shows that the phenotypic analysis may help understand the diagnosis of ADHD in the future.

TABLE I
TEN-FOLD CROSS VALIDATION OF AMYGDALA WITHOUT THE PHENOTYPIC INFORMATION

	ELM - Without Phenotypes					PBL-MCRBFN - Without Phenotypes			
	Training accuracy in %		Testing accuracy in %			Training accuracy in %		Testing accuracy in %	
Fold	Overall	Average	Overall	Average	Fold	Overall	Average	Overall	Average
1	71.35	70.71	60.00	58.98	1	97.88	97.91	60.80	58.80
2	71.35	70.56	60.00	58.98	2	96.29	96.34	56.00	55.02
3	73.47	73.19	57.60	57.09	3	99.47	99.50	59.20	57.54
4	73.74	73.51	58.40	57.95	4	95.23	95.24	58.40	58.88
5	73.47	73.03	56.00	55.94	5	96.82	96.72	64.80	65.20
6	69.23	68.50	60.00	59.91	6	95.76	95.50	67.20	66.16
7	71.88	71.43	60.00	59.33	7	95.76	95.81	64.00	63.64
8	70.82	70.18	57.60	56.51	8	95.76	95.58	64.00	61.79
9	70.56	69.93	64.00	63.41	9	95.49	95.68	66.40	64.95
10	73.21	72.63	63.20	61.97	10	99.20	99.14	57.60	59.06
Mean	71.91	71.37	59.68	59.01	Mean	96.77	96.74	61.84	61.10
Std	1.44	1.58	2.35	2.23	Std	1.47	1.48	3.74	3.58

TABLE II
TEN-FOLD CROSS VALIDATION OF AMYGDALA WITH THE PHENOTYPIC INFORMATION

	ELM - With Phenotypes					PBL-MCRBFN - With Phenotypes			
	Training accuracy in %		Testing accuracy in %			Training accuracy in %		Testing accuracy in %	
Fold	Overall	Average	Overall	Average	Fold	Overall	Average	Overall	Average
1	97.61	97.62	99.20	99.14	1	97.61	97.50	97.60	97.53
2	97.61	97.66	96.80	96.90	2	96.82	96.88	98.40	98.28
3	97.88	97.98	100.00	100.00	3	98.67	98.61	98.40	98.39
4	98.14	98.19	96.00	96.15	4	97.88	97.87	96.80	96.67
5	97.61	97.66	97.60	97.65	5	97.61	97.58	97.60	97.53
6	98.41	98.40	98.40	98.51	6	97.08	97.09	96.00	96.04
7	97.61	97.70	96.80	96.78	7	97.88	97.91	93.60	93.45
8	97.61	97.73	99.20	99.14	8	96.82	96.76	100.00	100.00
9	97.35	97.45	100.00	100.00	9	98.41	98.40	99.20	99.14
10	97.88	97.94	98.40	98.51	10	98.94	98.93	98.40	98.39
Mean	97.77	97.83	98.24	98.28	Mean	97.77	97.75	97.60	97.54
Std	0.30	0.28	1.33	1.29	Std	0.70	0.70	1.72	1.74

VI. ACKNOWLEDGMENT

We would like to thank the ADHD-200 consortium for making the MRI data publicly available. The authors wish to extend their thanks to the Ministry of Education (MoE), Singapore, for providing financial support through tier I (No. M4011269) funding to conduct this study.

REFERENCES

[1] A. Anderson, P. K. Douglas, W. T. Kerr, V. S. Haynes, A. L. Yuille, J. Xie, Y. N. Wu, J. A. Brown, and M. S. Cohen, "Non-negative matrix factorization of multimodal mri, fmri and phenotypic data reveals differential changes in default mode subnetworks in adhd," *NeuroImage*, vol. 102, Part 1, no. 0, pp. 207 – 219, 2014, multimodal Data Fusion.

[2] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, 2007.

[3] G. S. Babu and S. Suresh, "Sequential projection-based metacognitive learning in a radial basis function network for classification problems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 2, pp. 194–206, Feb 2013.

[4] G. S. Babu, S. Suresh, and B. S. Mahanand, "A novel PBL-McRBFN-RFE approach for identification of critical brain regions responsible for parkinson's disease," *Expert Systems with Applications*, vol. 41, no. 2, pp. 478–488, 2014.

[5] R. Badrinarayanan, S.Suresh, and B. S. Mahanand, "Identification of potential biomarkers in the hippocampus region for the diagnosis of adhd using pbl-mcrbfn approach," in *13th International Conference on Control, Automation, Robotics and Vision, (ICARCV 2014). (Accepted)*.

[6] T. Banaschewski, K. Becker, S. Scherag, B. Franke, and D. Coghill, "Molecular genetics of attention-deficit hyperactivity disorder: An overview," *European Child and Adolescent Psychiatry*, vol. 19, no. 3, pp. 237–257, 2010.

[7] M. A. Brotman, B. A. Rich, A. E. Guyer, J. R. Lunsford, S. E. Horsey, M. M. Reising, L. A. Thomas, S. J. Fromm, K. Towbin, D. S. Pine, and E. Leibenluft, "Amygdala activation during emotion processing of neutral faces in children with severe mood dysregulation versus adhd or bipolar disorder," *American Journal of Psychiatry*, vol. 167, no. 1, pp. 61–69, 2010.

[8] G. C. Cawley, N. L. C. Talbot, and M. Girolami, "Sparse multinomial logistic regression via bayesian l1 regularisation," in *NIPS*, 2006, pp. 209–216.

[9] S. Cortese, "The neurobiology and genetics of attention-deficit hyperactivity disorder: What every clinician should know," *European Journal of Paediatric Neurology*, 2012.

[10] A. De La Fuente, S. Xia, C. Branch, and X. Li, "A review of attention-deficit/hyperactivity disorder from the perspective of brain networks," *Frontiers in Human Neuroscience*, vol. 7, no. 192, 2013.

[11] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, Eds., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007.

[12] T. Frodl, J. Stauber, N. Schaaff, N. Koutsouleris, J. Scheuerecker, M. Ewers, M. Omerovic, M. Opgen-Rhein, H. Hampel, M. Reiser, H.-J. Mller, and E. Meisenzahl, "Amygdala reduction in patients with adhd compared with major depression and healthy volunteers," *Acta Psychiatrica Scandinavica*, vol. 121, no. 2, pp. 111–118, 2010.

[13] G.-B. Huang, "An insight into extreme learning machines: Random neurons, random features and kernels," *Cognitive Computation*, pp. 1–15, 2014.

[14] A. Krain and F. Castellanos, "Brain development and ADHD," *Clinical Psychology Review*, vol. 26, no. 4, pp. 433–444, 2006.

[15] J. E. LeDoux, "Amygdala," *Scholarpedia*, 2008. [Online]. Available: http://www.scholarpedia.org/article/Amygdala

[16] B. S. Mahanand, R. Savitha, and S. Suresh, *Computer Aided Diagnosis of ADHD Using Brain Magnetic Resonance Images*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2013, vol. 8272, pp. 386–395.

[17] B. S. Mahanand, S. Suresh, N. Sundararajan, and M. Aswatha Kumar, "Identification of brain regions responsible for alzheimer's disease using a self-adaptive resource allocation network," *Neural Networks*, vol. 32, pp. 313–322, 2012.

[18] M. P. Milham, D. Fair, M. Mennes, and S. H. Mostofsky, "The ADHD-200 consortium: A model to advance the translational potential of neuroimaging in clinical neuroscience," *Frontiers in Systems Neuroscience*, vol. 6, no. 62, 2012.

[19] K. L. Mills, D. Bathula, T. G. Costa Dias, S. P. Iyer, M. C. Fenesy, E. D. Musser, C. A. Stevens, B. L. Thurlow, S. D. Carpenter, B. J. Nagel, J. T. Nigg, and D. A. Fair, "Altered cortico-striatal-thalamic connectivity in relation to spatial working memory capacity in children with adhd," *Frontiers in Psychiatry*, vol. 3, no. 2, 2012.

[20] J. Morris, A. Ohman, and R. Dolan, "Conscious and unconscious emotional learning in the human amygdala," *Nature*, vol. 393, pp. 467 – 470, 1998.

[21] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural MRI data," *PLoS ONE*, vol. 8, no. 11, 11 2013.

[22] K. Plessen, R. Bansal, H. Zhu, and et al, "Hippocampus and amygdala morphology in attention-deficit/hyperactivity disorder," *Archives of General Psychiatry*, vol. 63, no. 7, pp. 795–807, 2006.

[23] G. Polanczyk, M. S. de Lima, B. L. Horta, J. Biederman, and L. A. Rohde, "The worldwide prevalence of ADHD: a systematic review and metaregression analysis," *American Journal of Psychiatry*, vol. 164, no. 6, pp. 942–948, 2007.

[24] G. Pourtois, A. Schettino, and P. Vuilleumier, "Brain mechanisms for emotional influences on perception and attention: What is magic and what is not," *Biological Psychology*, vol. 92, no. 3, pp. 492 – 512, 2013.

[25] M. Qiu, Z. Ye, Q. Li, G. Liu, B. Xie, and J. Wang, "Changes of brain structure and function in adhd children," *Brain Topography*, vol. 24, no. 3-4, pp. 243–252, 2011.

[26] G. S. Sidhu, N. Asgarian, R. Greiner, and M. R. G. Brown, "Kernel principal component analysis for dimensionality reduction in fmri-based diagnosis of adhd," *Frontiers in Systems Neuroscience*, vol. 6, no. 74, 2012.

[27] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller, and G. Candrian, "Machine learning approach for classification of adhd adults," *International Journal of Psychophysiology*, vol. 93, pp. 162 – 166, 2014.

[28] S. Vigneshwaran, B. Mahanand, S. Suresh, and R. Savitha, "Autism spectrum disorder detection using projection based learning meta-cognitive rbf network," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, Aug 2013, pp. 1–8.

[29] V. Vuontela, S. Carlson, A.-M. Troberg, T. Fontell, P. Simola, S. Saarinen, and E. Aronen, "Working memory, attention, inhibition, and their relation to adaptive functioning and behavioral/emotional symptoms in school-aged children," *Child Psychiatry and Human Development*, vol. 44, no. 1, pp. 105–122, 2013.

[30] K. Watanabe and T. Kurita, "Locality preserving multi-nominal logistic regression," in *International Conference on Pattern Recognition*, 2008, pp. 1–4.