Modified Cuckoo Search-Cascade Forest (MCS-CF) for Attention Deficit Hyperactivity Disorder (ADHD) Diagnosis

Dr.T.V. Padmavathy[1*], Dr.M. Vinothkumar[2], Dr.M.N. Vimal Kumar[3]

Abstract

Attention deficit hyperactivity disorder (ADHD) is a disease state of the mind which is frequently observed in young children. Different machine learning approaches, which include Deep Neural Networks (DNNs) and it helps in ADHD classification. The following have been recently proposed: ADHD to examine employing functional Magnetic Resonance Imaging (fMRI) information and gcForest to differentiate between ADHD and normal theme, cascade forest is employed to make use of the concatenated feature vector samples in the form of input for classification. But, classification accuracy takes large time consuming. In order to deal with this problem, Modified Cuckoo Search-Cascade Forest (MCS-CF) based feature selection algorithm is suggested which helps in the accuracy improvement of the classifier used in ADHD. This technical work comprises of three important steps, which include (i) Preprocessing of Data and Extraction of Feature (ii) Selection of Feature, and (iii) Classification. At first, preprocessing consists of eliminating the initial 10 images, correction of slice time, correction of motion, normalization, filtration of band pass and smoothening. After this, the process of features extraction from the fMRI data has been carried out and they are of two types, viz., 1-D Functional Connectivity (FC) feature and 3-D Amplitude of Low Frequency Fluctuations (ALFF) feature. In the next step, MCS algorithm utilized for feature selection aids minimizing the feature space that in turn, boosts the prediction accuracy and reduces the time taken in computation. At last, CF is introduced for classification, whose inputs are the feature vector samples selected. As with the cascade forest structure, revised gcForest uses multilevel combinations of decision tree forests to amalgamate the two features together, thereby creating a novel feature vector in a concatenated way for every sample. The public set of test datas of ADHD-200 are used here and the hold-out testing information is computed in this implementation. The classification techniques are evaluated in terms of the metrics namely Accuracy (ACC), Sensitivity (SEN) and Specificity (SPE).

Key Words: Attention Deficit Hyperactivity Disorder (ADHD), Functional Connectivity (FC), Amplitude of Low Frequency Fluctuations (ALFF), Modified Cuckoo Search (MCS), Cascade Forest (CF).

DOI Number: 10.14704/nq.2020.18.7.NQ20196 **NeuroQuantology 2020; 18(7):83-94**

Introduction

Attention deficit hyperactivity disorder (ADHD), as a state in the innermost sensory system is a generic emotional and behavioral condition found in kids, teens, as well as few adult individuals. ADHD is defined by less important absorption, over activity or absence of self-control [1]. Patients exhibit multiple symptoms, like finding it to be attentive, hyperactivity, difficulty in having a controlled behavior, and cognitive irregularity.

Corresponding author: Dr.T.V. Padmavathy
Address: 1*Professor, Department of ECE, RMK Engineering College, Kavarapettai, Tamil Nadu; 2Associate Professor, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore; 3Assistant Professor, Department of ECE, RMD Engineering College, Kavarapettai, Tamil Nadu.
1*E-mail: tvp.ece@rmkec.ac.in
2E-mail: vinojimail@gmail.com
3E-mail: mnv.ece@rmd.ac.in
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 06 June 2020 **Accepted:** 16 July 2020

It is stated that huge people have been suffering and nearly few grown-up patients are still affected in illness, which had been detected in their small age [2]. The ADHD's syndrome slowly reduces or even vanishes as children grow up. In the case of adults, the syndrome's of hyperactivity were decreased, but the syndromes of absence of interest and impulsivity are still there, and few sufferers get even affected by other behavioral mental conditions [3]. But, the etiology is still a mystery in several cases, with no clarity on the diagnostic conditions, several children cannot get timed and right treatment during the early phases of ADHD. Efficient techniques are the need of the hour to help in the ADHD diagnosis [4].

Medical image processing is emerging to be an interdisciplinary research domain, inspiring experts and researchers from different areas such as mathematicians, physicians, engineers, doctors etc. Due to the collaboration of these specialists, the possibility of non-intrusive imaging modalities has emerged true [5-6]. Several research works have proven that ADHD is connected to dysfunction in the brain, and also the patient's brain structure and abnormal functions[7].

Statistical machine learning techniques have presently penetrated fields like Psychiatry, which focuses on the analysis and curing of neuropsychiatric ailments. The presence of massive scale neuro imaging datasets has motivated the authors to design computer-aided tools as well as the processes for knowing the individual's cerebrum and its conditions [8].Few survey make use of structural MRI scans, which yield a non-intrusive approach for getting a picture of brain anatomy's, since neuroimaging technology, functional Magnetic Resonance Imaging (fMRI) were comprehensively employed for ADHD examination. The brain activity is measured by fMRI through the detection of changes corresponding to the flow of blood. This approach depends on the interconnected fact of flow of cerebral blood and neuronal activation. If a region of the brain is being used, it results in maximum flow of blood in that area. Through the detection of particular encephalic areas, such dorsal Anterior Cingulate Cortex (dACC), the Ventro Lateral Prefrontal Cortex (VLPFC) and the putamen, irregular activations of brain can be discovered [9].

But, Deep Neural Networks (DNNs) have extreme number of hyper-parameters and the presentation is hugely dependent on tuning the parameter. In the form of an alternate to DNNs, the deep forest or gcForest is used. It has been found that the deep forest mechanism demonstrates tough competition to deep neural networks. The multiple layer structure is used in deep forest, where every layer has several random forests. It is really a combination of ensembles of decision tree. As opposed to deep neural networks that need training data with large-scale and huge efforts in tuning the hyper-parameter, gcForest is simple in the direction of education and it functions well even in the presence of small-scale training data. These features render gcForest to be anapt classifier for diagnosis of ADHD.

This work explores the usage of gcForest for supporting the ADHD diagnosis. Modified Cuckoo Search-Cascade Forest (MCS-CF) based feature selection algorithm is introduced and it helps in the accuracy improvement of the classifier used in ADHD. This research work comprises of three important steps, which include (i) Preprocessing of Data and Extraction of Feature (ii) Selection of Feature, and (iii) Classification and these stages are carried out sequentially. Feature extraction faces a big challenge due to the high dimensionality of data. Feature selection is an important step as several times there can be repetitive or corrupted data, which could result in degradation of the overall performance of the new model. Optimizing the features and classifiers is a critical step in getting an optimum solution. Preprocessing comprises of eliminating the first ten images, normalization, band pass filtration and smoothening. A revised gcForest technique generally combines 1-D FC feature and 3-D ALFF feature with multi-grained scanning for the generation of a merged and modified feature for the classification step. MCS algorithm utilized for feature selection aids in reducing the feature space, which in turn helps in the prediction accuracy improvement. Also, for balancing of data, Synthetic Minority Over-sampling Technique (SMOTE) integrated with Edited-Nearest Neighbor (ENN) is also proposed for generation of synthetic minority samples'. When compared with the usage of 1-D FC feature or 3-D ALFF feature, the concatenated feature can help in improving the classification execution.

Literature Review

Sen et al [8] presented learning models LEFMS, LEFMF, LEFMSF, which utilize the Support Vector Machine (SVM) learning algorithm for detecting ADHD, and furthermore Autism, utilizing structural texture and features with functional connectivity is obtained from 3-dimensional structural MRI and 4-dimensional resting-state fMRI outputs of the subjects under inspect. Three set of students is investigated: (1) The LEFMS student initially plays out the extraction of the features from the structural MRI images utilizing the texture-based filters delivered by means of an inadequate auto encoder. The resultant features helps in a contribution of a linear SVM classifier. (2) The LEFMF student yields a diagnostic model through initially playing out the calculation of the spatial non-stationary free segments of the fMRI scans, which it endeavors to separate the fMRI output of each subject into the time courses of these conventional spatial parts. Regardless of the method being utilized, definitive arrangements of highlights are transmitted as input to a straight SVM classifier. (3) At last, the general LEFMSF student utilizes the feature collection obtained from the two extraction of feature completed in (1) and (2) as contribution to a SVM classifier, achieving much better exactness ADHD-200 holdout information and ABIDE holdout information. The outcomes uncover that integrating multi-modal features can render a superior grouping exactness for diagnosing ADHD and Autism, which again is a fundamental advance towards PC helped analysis of these psychiatric diseases and others as well.

Chen et al [10] proposed a multichannel Deep Neural Network (mcDNN) classification model, which depends on multiscale cerebrum functional connect me information as well as the exhibition of this model, is shown utilizing ADHD recognition. The mcDNN model enforced the multiscale brain connect me data and Personal Characteristic Data (PCD) as consolidated features for ADHD identification and mainly prescient brain connectome features recognized for the conclusion of ADHD. The mcDNN model was then contrasted and single-channel Deep Neural Network (scDNN) models as well as the exhibition of characterization was evaluated utilizing cross-validation and hold-out validation as far as the measurements, for example, exactness, affectability, particularity, and Area Under the receiver operating characteristic Curve (AUC).

During the cross-approval, the mcDNN model employing ensemble features (mixture of the multiscale brain connectome information and PCD) accomplished the unrivaled presentation in ADHD discovery joined by means of an AUC in examination with scDNN models that utilizes the features of the brain connectome at each single scale and PCD. Through the combination of the multiscale brain connectome information, the mcDNN model helped enhancing the exhibition of ADHD location considerably rather the situation where a solitary scale is utilized.

Riaz et al [11] learned about a SVM machine learning framework by coordinating non-imaging information with imaging information to inspect the efficient connectivity changes among ADHD and controls (not ADHD analyzed). The target of this exploration work is to utilize computational methodologies, which (1) play out the automatic classification of a subject to be ADHD or control, (2) discover the distinctions in FC of these two sets as well as (3) survey the implication of joining non-imaging with imaging information intended for the classification purposes. In the initial step, the functional connectivity of brain regions is controlled through the characterizing the activity of the brain utilizing Affinity Propagation (AP) clustering algorithm.

At that point, Elastic Net based feature selection is utilized for choosing the most discriminative features from the denser functional brain network and non-imaging information combination. At last, a SVM classifier is prepared for classifying the ADHD subjects and the control. The tale structure was surveyed on a straightforwardly accessible ADHD-200 dataset, and the outcomes show that ensemble of non-imaging information helps enhancing the system implemetation. The consequences of the classification perform better when compared with the benchmarks on fedata subsets.

Shao et al [12] exhibited a changed gcForest to separate ADHD subjects and typical controls. There are two sorts of features, which are obtained from the fMRI information, including the 1-D FC feature and 3-D ALFF feature. An altered gcForest strategy which utilizes a fused multi-grained examining structure to consolidate both the features together, and in this way a novel merged feature vector can be produced for all the examples. Additionally, taking the imbalanced trait of ADHD information, synthetic minority over- sampling approach converged with altered closest neighbor and it helps to make the synthetic minority merged feature vector samples for adjusting the information. Lastly, CF helps in linking the feature vector tests to go about as the classification input.

The straightforwardly accessible set of test data ADHD-200 is utilized and its exhibition is assessed on the hold-out testing information.

Zou et al [13] works in MRI filters, which is learned from a deep learning-based ADHD classification technique through 3-D Convolutional Neural Networks (CNNs). As deep neural networks may apply multitudinous parameters, yet the huge number of MRI tests in enormous informational collections is still significantly less in the event that one must take in the Discriminant features from the raw information.

From the start, the valuable 3-D low-level features are separated from functional MRI (fMRI) and basic MRI (sMRI) information. Besides, it is supported by the general strategy of radiologists for analyzing the brain images', a 3-D CNN model is intended to investigate the nearby spatial examples of the MRI features. Lastly, it is discovered that that functionality of the brain and structural information supplement one another, as well as multi-methodology CNN architecture is intended to incorporate the fMRI and sMRI features. Computation gets completed on the hold-out testing information of the ADHD-200 worldwide rivalry uncover that the recently presented multi-modality 3-D CNN plot accomplishes the amazingly phenomenal precision in correlation with other accessible strategies. Additionally, it is prescribed that multi-methodology classification will turn into a potential course to get solid neuroimaging biomarkers of neuro advancement diseases.

Wang and Kamata [14] studied about a 3D-CNN classifier that employs 3D Fractal Dimension Complexity Map (FDCM) for automated diagnosis of ADHD. ADHD classification employs 3D Hausdorff FDCM obtained from Gray Matter (GM) density. We can extract the GM density data from structural MRI data, in order to evaluate the Hausdorff fractal dimension. In the next step, 3D-CNN for features of extraction from FDCM then reviewing ADHD and Typically Developing Controls (TDC). The map performs the extracts the density features at the voxel-level and saves the morphological information on the brain structure. Fractal size could be of much help in explaining the complexness of cortex folds and have significant structural information. 3D CNN is utilized for the extraction of common features associated with ADHD bio-markers from various complexity maps for computing the probability of classification by sending through the soft-max layer. On the hold-out testing data, this model is assessed for the ADHD-200 global contest and its performance is much better than the earlier techniques that depend on structural MRI data.

Proposed Methodology

In this research work, a Modified Cuckoo Search-Cascade Forest (MCS-CF) is proposed employing the improved gcForest to help in the ADHD diagnosis. This technical work comprises of three important steps, which include (i) Preprocessing the Data and Extraction of Feature (ii) Selection of Feature, and (iii) Classification. The first step of preprocessing consists of the initial 10 images; correction of slice time, correction of motion, normalization, filtration of band pass and smoothening. After this, there are two sets of extraction of features from the fMRI data, which include 1-D Functional Connectivity (FC) feature and 3-D Amplitude of Low Frequency Fiuctuations (ALFF) feature. In the second step, MCS algorithm utilized for feature selection aids in the feature space reduction which increases the prediction accuracy level and reduces the time taken in computation. At last, CF is introduced for classification which considers the chosen feature vector samples to be the input.

Data Preprocessing and Feature Extraction

From ADHD-200 Global competition [15], we get the ADHD fMRI data. The traits are done on four set of test datas, which include Peking University (Peking), Kennedy Krieger Institute (KKI), New York University child study center (NYU) and Neuro Image sample (NI). We should note that these set of test datas from earlier centers were gathered with diverse parameter configurations. Table 1 provides a small outline of these set of test datas.

Table 1. Description of the Set of test datas of ADHD-200 Competition

Set of test datas	Peking	KKI	NYU	NI
Training subjects	194	83	216	48
ADHD subjects	78	22	118	25
Control subjects	116	61	98	23
Hold-out testing subjects	51	11	41	25
ADHD subjects	24	3	29	11
Control subjects	27	8	12	14

DPARSF toolbox [16] is used for carrying out the preprocessing of data. Initial 10 images' elimination will be there in this preprocessing, correction of slice time, correction of motion, normalization, filtration of band pass and smoothing. In every fMRI data, the cerebral brain image is divided into 90 sectors of brain. Every sector in the head is utilized for calculating a series of average time of all the voxels. The Coefficient of Pearson Correlation is calculated to create a Functional Connection (FC) matrix [17], for any average time series pair. Figure 1 shows the flowchart of FC matrix extraction. As the FC matrix is a symmetric one, the matrix's lower left triangle helps in forming a feature vector. The feature vector can be attained, concatenation lower left triangle's first-row vector and the last-row vector.

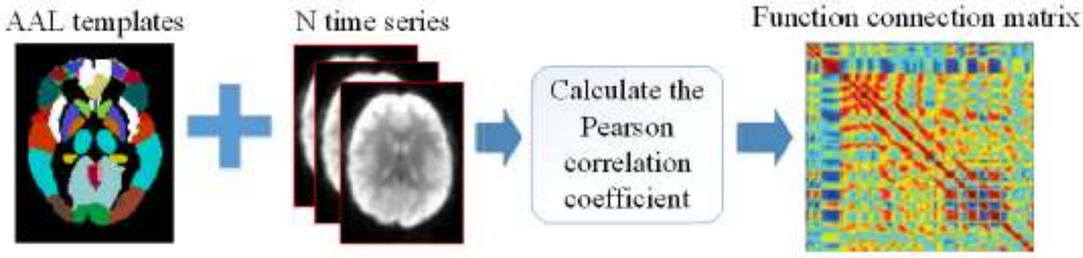

Figure 1. Functional Connection Matrix Acquisition Flowchart

In addition, the fMRI's ALFF image is obtained by REST [18]. Figure 2 shows the processing step for the generation of ALFF. At first, the series of filtered time with respect to the voxel's fMRI and they were translated into frequency domain signals using transformation of fast Fourier for getting the spectrum of power. The averaged square root across 0:01∽0:08Hz at every voxel is considered to be the value of ALFF. At last, for standardization, the ALFF of every voxel is alienated by the universal average ALFF value.

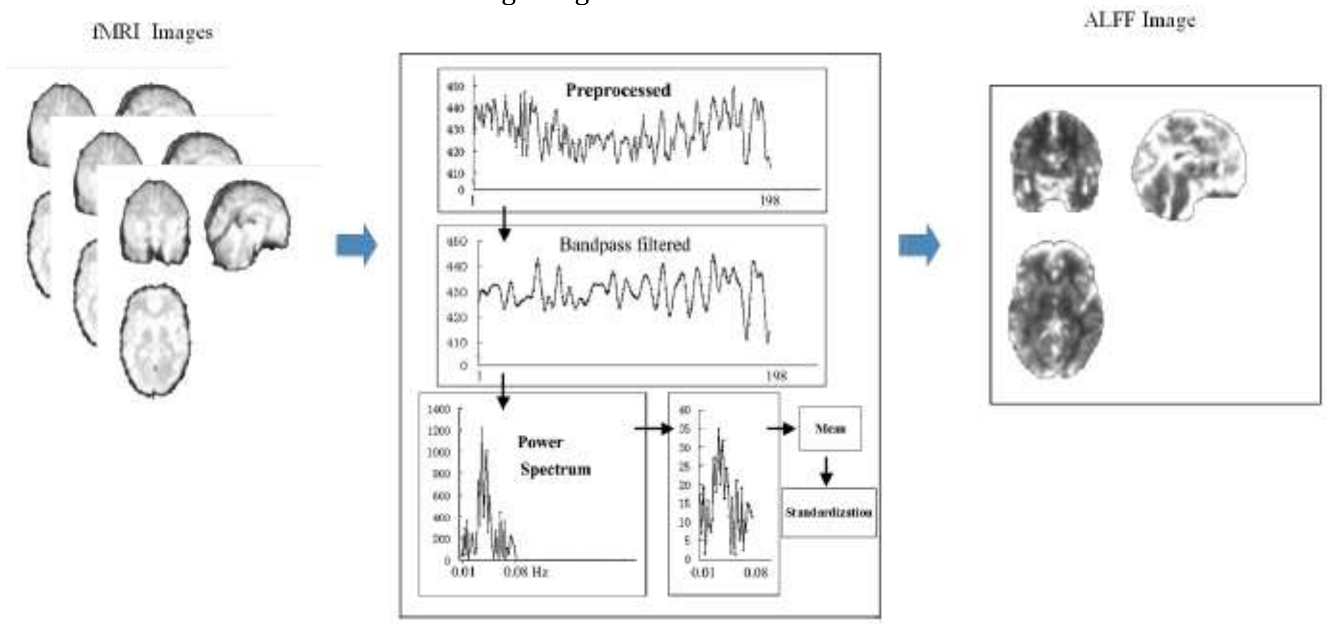

Figure 2. ALFF Image Processing Flowchart

Feature Selection

Feature selection is used in the feature space reduction and it increases the prediction accuracy along with reducing the time taken for computation. This is attained by eliminating the unnecessary, repetitive and noisy features. i.e. it chooses the subset of features, which can help achieving remarkable concert with respect to exactness and computational time.

Cuckoo Search

Cuckoo Search (CS) is a meta-heuristic algorithm influenced by the bird cuckoo; and these include the "Brood parasites" birds. There is no way for forming its own set of features and lays their features which is chosen among number of features. Cuckoo is a popular brood parasite.

Cuckoo Breeding Behaviour

Cuckoos lay their selected features in communal number of features; while they eliminate other selected features for maximizing the emerging possibility of their own selected features. Various species engage the obligate brood parasitism by leaving their selected features from various features of other host birds [19]. Brood parasitism has 3 basic varieties: parasitism with intra specific brood, cooperative breeding, and number of invasion. If the discovered host bird eggs doesn't belongs to them, then they will either get rid of these strange selected features or just discard its shell as well as develops a fresh feature's count elsewhere. The possibility of their selected features being discarded and maximizing their reproductively will be minimized through this activity. Parasitic cuckoos often desire a feature's count where the host bird just places their respective selected features. In general, the cuckoo selected features emerge vaguely earlier when compared with their host selected features. Once hatching its initial cuckoo chick, the initial sense of deed will take part, in order to evict the host selected features, which, in turn, maximizes the cuckoo chick's share of food that is provided by its host bird [20].

L´evy Flights

Different works have revealed that the behavior of flight of various animals and insects may exhibit few common features of L´evy flights. Analysis on human behavior on this kind of foraging patterns also reveals the common characteristics of L´evy flights. Later, this behavior helps in optimization and optimum search, and preface reveals its potential skills.

Cuckoo Search

In a number of features, each selected features indicates a solution and cuckoo selected features indicates a solution in a new and good way. Each feature's count has one selected features of cuckoo in a simplest form, in which each feature's count will have multiple selected features indicating a solution set. For such behavior of breeding, cuckoo search can be idealized and for various optimization problems, we can enforce the same [21].

At a time, one selected features will be laid by every Cuckoo and number of features were chosen without any order in order to dump it.

The best number of features with the high quality of selected features will proceed to the next generations.

The accessible host with various features is fixed and if a host bird recognizes the cuckoo chosen features with the likelihood of p_a=[0,1] then the host bird creature can either reject them or leave them and manufacture another chose features.

As a further estimation, this last supposition can be approximated by a division p_aof the n have number of features are supplanted by new number of features. For an amplification issue, the quality or wellness of an answer can just be relative to the estimation of the goal work. For the usage perspective, chosen features gives an answer, and each cuckoo can lay just one egg (therefore speaking to one solution), here new point can be utilized and possibly better solutions (cuckoos) to supplant a not-so-good solution in the quantity of features. Clearly, this algorithm can be stretched out to the additional confused situation where each number of features has various chosen features speaking to a lot of arrangements, or indicating the multi-destinations [21]. In light of these three guidelines, the elementary strides of the Cuckoo Search (CS) can be condensed as the pseudo code.

Cuckoo Search via Levy Flights
Objective function $f(x), x = (x_1, \ldots\ldots, x_d)^T$
Generate initial population of n host number of features x_i
While (t<Max Generation) or (stop criterion)
Get a cuckoo randomly/generate a solution by Levy flights
and then evaluate its quality/fitness F_i
Choose a number of features among$n(say, j)$ randomly
if $\left(F_i > F_j\right)$
Replace j by the new solution
end
Abandon a fraction (p_a) of worse number of features & generate new solutions
Keep selected features solutions and find the current number of features
end while

Postprocess Results and Visualization

This algorithm utilizes the balanced ensemble of a local random walk and the universal explorative random walk, and a switching parameter p_a controls it. The local random walk can be expressed as

$$x_i^{t+1} = x_i^t + s \otimes H(p_a - \varepsilon) \otimes (x_j^t - x_k^t)$$

where x_j^t and x_k^t refer to two different solutions chosen in random with the help of random permutation, H(u)refers to a Heaviside function, ε stands for a random number obtained from a distribution in a uniform manner, and step size will be represented by s. On the opposite side, the global random walk is performed employing L´evy flights

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda)$$

Where

$$L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin\left(\frac{\pi \lambda}{2}\right)}{\pi} \frac{1}{S^{1+\lambda}}, (s \gg s_0 > 0)$$

The L´evy flight generally yields a random walk whose length of the random step is obtained with the help of an L´evy distribution

$$Levy \sim \frac{1}{S^{\lambda+1}}, (0 < \lambda \leq 2)$$

Here, the steps generally create a process in a random walk, which has the distribution of a power-law step-length and a heavy tail. Few new solutions have to be created using L´evy walk among the best solution. But, a far field randomization helps in recognizing the extensive portion of the new solutions and whose positions must be sufficiently extreme as of the present best solution and it guarantee that the system will not get caught in local optima.

Drawbacks of Cuckoo Search

Boundary problem: The positions of few features may be locations external to the boundary. The bound dealing technique will lead to several features at the same position on the boundary, which is not effective.

CS technique is not ideal; it can easily be trapped into the solution of local optimum and also its convergence is slow.

A Modified Cuckoo Search (MCS) algorithm is proposed, for rectifying these issues.

Modified Cuckoo Search (MCS)

In reality, if a cuckoo's chosen features is fundamentally the same as a host's chosen features, at that point this present cuckoo's chosen features is more disinclined to be found and along these lines the wellness ought to be identified with the distinction in solutions. In this manner, it's a smart way to do a random walk in a one-sided path (biased way) with some random step dimensions. Both, unique and changed code utilizes random step sizes. In the first code, step size is determined utilizing following code articulation:

$$r_1 * number\ of features[permute1[i]][j] - number\ of features[permute2[i]][j]$$

Where r_1 refers to a random number in the range [0,1], the feature's count frames a matrix which has the features picked alongside their factors, permute1 and permute2 show the various rows permutation functions utilized on the matrix of quantity of features.

$$r_1 * number\ of features[sorted[i]][j] - number\ of features[permute[i]][j]$$

The variation is that, sorted function helps here instead of permute1. This capacity arranges the quantity of features matrix utilizing the fitness solutions. This strategies keeps up the selection pressure (the degree to which very fit solutions are picked) towards the better solutions and algorithms must achieve prevalent outcomes [22]. This doesn't infer that high fitness solution will surpass in the populace and the algorithm will be entangled in nearby optima. CS algorithm identifies the best solution x_best at the beginning of each iteration process.

$$\sigma_u = \left\{ \frac{\Gamma(1+\beta) \sin\left(\frac{\pi \beta}{2}\right)}{\Gamma\left[\frac{(1+\beta)}{2}\right] \beta 2^{\frac{(\beta-1)}{2}}} \right\}, \sigma_v = 1$$

where β indicates Lévy dissemination parameter and Γ indicates gamma work. The advancement of cuckoo i starts with the contributor vector v, where $v = x_i^{(t)}$ Step size is registered according to the condition underneath. 89

$$Stepsize^{(t+1)} = 0.01 \frac{u^{(t+1)}}{|v^{(t+1)}|^{\frac{1}{\beta}}} (v - x_{best})$$

Where $u \sim N\ (0, \sigma_u^2)$ $v \sim N\ (0, \sigma_v^2)$ indicates the samples from the respective normal distributions. As to the step size, the suggested parameter β=1.5 according to Levi distribution. The best discovered solution but not in memory individually because of the perfect solution, as per the algorithm, is always saved among the current solutions. This implies that the procedure adopts the property of Markov as the next state relies just on the position of the current work and not the earlier position.

Classification

In the cascade forest structure, modified gcForest uses multilevel combinations of decision tree forests to merge the two features together, and this way, a novel merged feature vector can be designed for every model.

gcForest

Tree-based ensemble machine learning approaches, called random forest [4] offer advantages in handling with issues including nonlinear characterization and over fitting. gcForest produces a deep fprest ensemble and achieves better execution with respect to representation learning and high dimensional information learning issues. gcForest incorporates two significant forms, for example, multi-grained scanning and cascade forest. Given a lot of raw input data, the preparing will be finished by the multi-grained scanning for producing the adjusted and consolidated chosen feature vectors. From there on, the linked chosen feature vectors will be broadcasted into the form of cascade forest, in order to play out the assignment of classification. Accept a set of test data having 100 dimensional features, for example the dimension of the information is 100×1. For each preparation test, it will be given to the structure of the multi-grained scanning (Figure 3) to produce another consolidated feature vector. Like it can be seen in multi-grained scanning, sliding window innovation helps in building new occurrences from the genuine information. We Expect that the dimension of the sliding window is 10×1 and the progression size is 1, at that point for each example, 91 of 10-dimensional occasions would be delivered. Hence, every example is given to two diverse forests correspondingly to create a 2-dimensional vector as the output of class distribution vector. Subsequently, a novel91×2×2 D 364- size modified feature vector is delivered as the resultant yield of multi-grained examining.

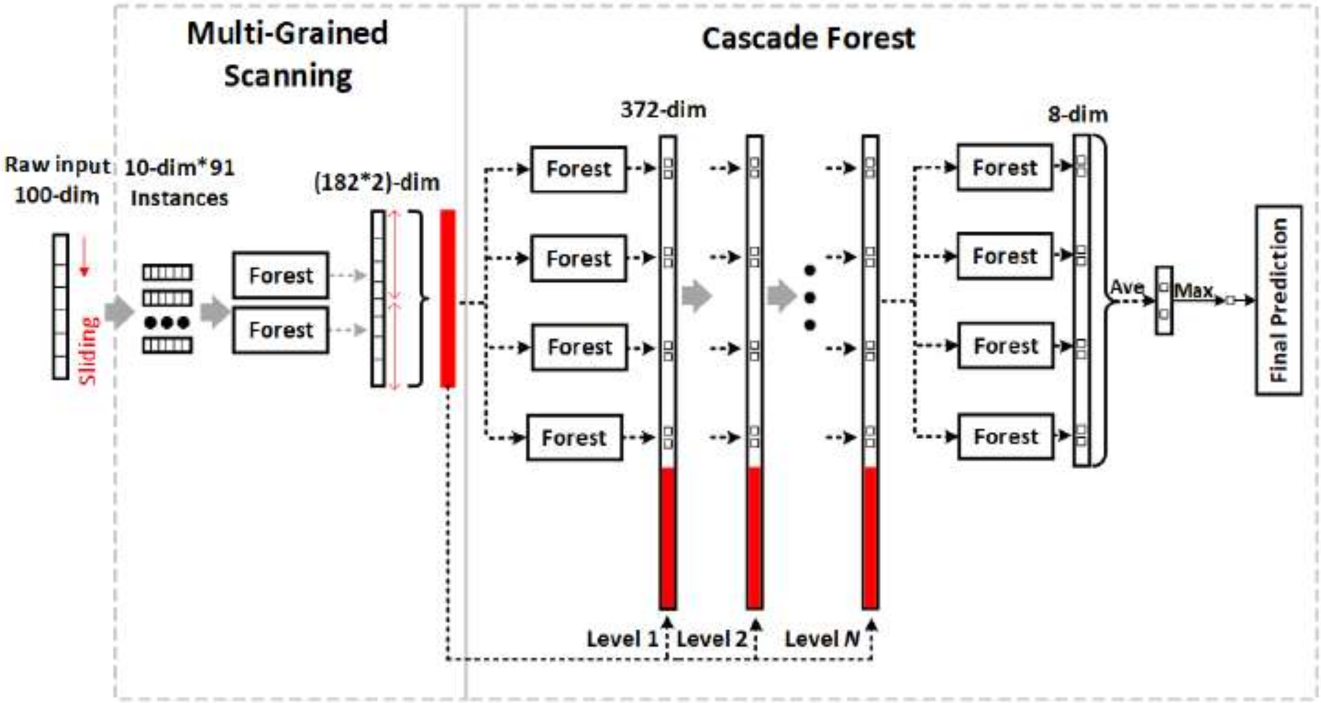

Figure 3. Structure of gcForest

As found in Figure 3 of cascade forest structure, gcForest utilizes multilevel blends of decision tree forests. Each degree of cascade has different forests, each forests will give a class distribution vector as the yield. At that point entire class vectors delivered by the forests are linked in a comparable level and the yield of multi-grained examining is given as input vector to the following ensuing level. For the binary characterization model, assume that there are four arbitrary forests in each level, it tends to be seen that each forests gives a 2-dimensional vector as the yield and each degree of cascade gives a 4×2 D 8-dimensional vector as the yield. Combining the 8-dimensional vector with the real 364-dimension farture, the contribution of the cascade level that follows is a 372-dimensional vector. A definitive expectation would deliver the maximum incentive in the vector of averaged class gained from the last degree of cascade. It must be seen that the forests in gcForest are not restricted to common irregular forests, whose spot considers the completely random forests or different classifiers, which gives the class distribution vectors as the yield.

Revised gcForest for fMRI Classification

Figure 4 shows the structure of the altered gcForest for fMRI data classification. For every last one of the fMRI tests, their FC and ALFF features are removed. After this, during the training procedure,

*e*ISSN 1303-5150

the training information will go through multi-grained scanning, SMOTE with ENN for data balancing and cascade forest for preparing the model parameters. When the preparation is done, the checking information will go through the multi-grained scanning and the prepared cascade forest is allowed for assessing the execution of classifier.

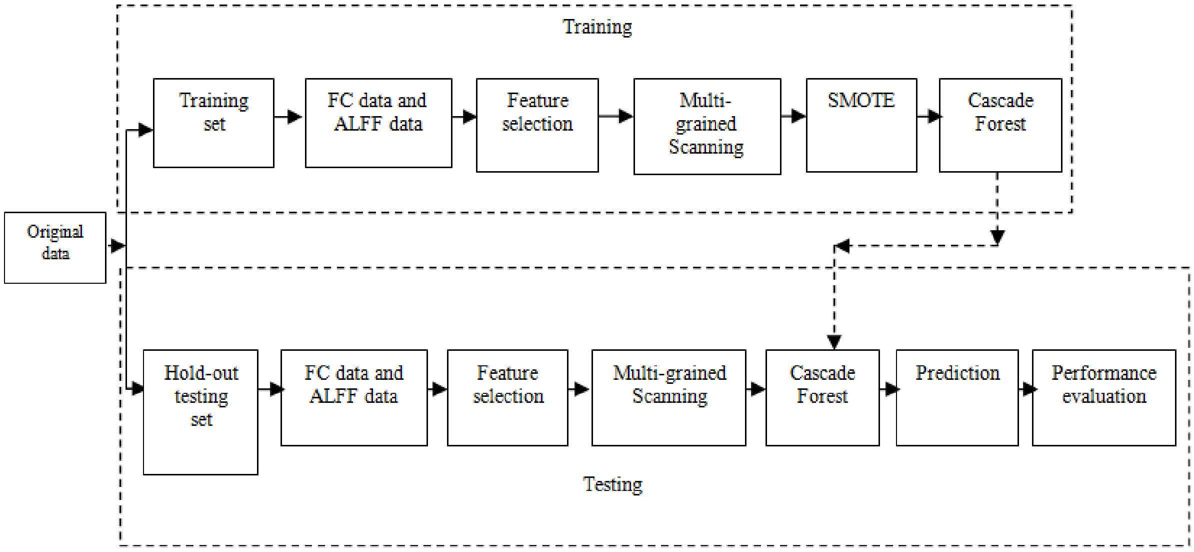

Figure 4. Flowchart for Model Training and Testing

Feature-Fused Multi-Grained Scanning Structure

For consolidating the 1-D FC feature with the 3-D ALFF ALFF, a ALFF combining structure containing two multi-grained scanning is utilized. For each example x_i, its FC feature is really a 1-D array. A 1-D sliding window having a consistent advance size and it will be used for the FC information scanning into a few cases. Every one of these examples will be sent into a consistent amount of random forests for delivering the class vectors as the yield. The connected type of these class vectors shapes the finished yield of 1-D scanning process. Along these lines, for the ALFF feature of sample x_i, as it is a 3-D array, a 3-D sliding window having steady advance dimension is used for the generation of numerous cases. Every one of these occasions will be given to a consistent amount of random forests for delivering the class vectors as the yield. Thus, a definitive yield of 3-D filtering is done for each example results because, these class vectors being connected. Lastly, the yield of the 1-D and 3-D multi-grained scanning are consolidated to make another vector, which has to be the altered the joined element of FC and ALFF.

Data Balancing

The ADHD set of test data utilized in these tests are extremely unnecessary, which infers that there is no balance in the positive and negative examples. As the standard learning algorithm may create imperfect classifiers, it is basic to handle with the issue of imbalance. SMOTE algorithm is a random over sampling algorithm that produces new manufactured examples by assessing the neighbors of the minority tests. Assume $S_A \in S$, where Srefers to the arrangement of the considerable number of tests and S_A indicates the arrangement of minority tests. SMOTE algorithm functions as beneath [24]. For each example $x_i \in S_A$, its knearest in S_Ais registered. After this, one of the $kx_i \in S_A$, its knearest in $S_A$is chose randomly, and the new manufactured minority test can be figured as:

$$x_s = x_i + (y_i - x_i) * r$$

where r indicates any number among [0,1] and x_sindicates the new synthetic example. Yet, SMOTE algorithm delivers new examples utilizing the real minority tests without contemplating its neighboring examples. It might prompt minority tests, which stay among the majority samples. This may prompt the expanding overlap amid different classes, subsequently bringing about characterization results that are not outstanding. With the point of improving the presentation, here SMOTE with ENN is utilized for adjusting the preparation sets. ENN is useful in dispensing with the new engineered tests, which are unique in relation to two of its three nearest neighbors [23]. As the size of the genuine fMRI information is excessively enormous, for having a few investment funds in the computational time, SMOTE with ENN is utilized right now the multi-grained scanning venture to straightforwardly create new synthetic altered affixed melded for minority feature tests.

Cascade Forest

Right now, two random forests as well as the two entirely-random forests are utilized at all degrees of the cascade forest. At the time of training process, the adjusted feature vector will be sent into the cascade forest for preparing the model parameters. During the testing strategy, the changed feature vectors will be sent into the prepared model of cascade forest without remembering the forecast outcomes as the yield.

Results and Discussion

For every information set, the experiments with/without data balancing were reproduced 10 times. The experiments are carried out on Peking, KKI, NYU, NI set of test datas. The results are calculated with respect to the average accuracy (ACC), sensitivity (SEN), specificity (SPE).The results of the novelMCS-CF are compared with other available techniques such as SVM, CF. A confusion matrix is frequently utilized for describing the classification model's execution (or "classifier") on a set of test data for which the accurate values are not unknown. It permits visualizing the algorithm performance.

- Accuracy (ACC) is one of the most popularly employed metrics for measuring the classification performance, and it is defined to be a proportion amid the rightly classified samples to the overall number of samples.

$$\text{Accuracy(ACC)}=\frac{(Tp+Tn)}{(Tp+Fp+Fn+Tn)}$$

- Sensitivity (SEN) (also identified as the true positive rate, the recall, or detection probability in few fields) provides a measure of the ratio of the original positives, which are rightly recognized as such (e.g., the % of sick people who are rightly found to have the condition). A classifier indicates the positive rightly classified samples to the overall positive sample's count.

$$\text{Sensitivity(SEN)}=\frac{Tp}{(Tp+Fn)}$$

- Specificity (SPE) (also known as the true negative rate) measures the ratio of the original negatives, which are rightly recognized as such (e.g., the % of healthy people who are rightly found to be not affected with the condition). The proportion of the rightly classified negative samples to the overall number of negative samples.

$$\text{Specificity (SPE)}=\frac{Tn}{(Tn+Fp)}$$

Where Tp, Fp, Tn and Fn are defined as:
True Positive (Tp): Observation is positive, and is also predicted positive.
False Negative (Fn): Observation is positive, but is predicted to be negative.
True Negative (Tn): Observation is negative, and is predicted as negative.
False Positive (Fp): Observation is negative, but is predicted to be positive.
Table 2 discusses the overall results computed of the techniques with metrics of performance evaluation is discussed.

Table 2. Performance Comparison Metrics vs. ADHD Classification Techniques

Dataset Name	Metrics	Techniques		
		SVM	CF	MCS-CF
Peking	ACC (%)	69.25	73.50	79.05
	SEN (%)	72.15	72.50	76.05
	SPE (%)	66.69	73.50	82.50
KKI	ACC (%)	65.87	72.50	85.91
	SEN (%)	77.67	79.50	82.25
	SPE (%)	83.80	86.20	91.20
NYU	ACC (%)	66.69	73.50	78.05
	SEN (%)	83.50	85.20	88.50
	SPE (%)	60.33	65.25	72.50
NI	ACC (%)	69.25	73.50	82.00
	SEN (%)	95.25	96.00	97.70
	SPE (%)	69.21	92.39	95.51

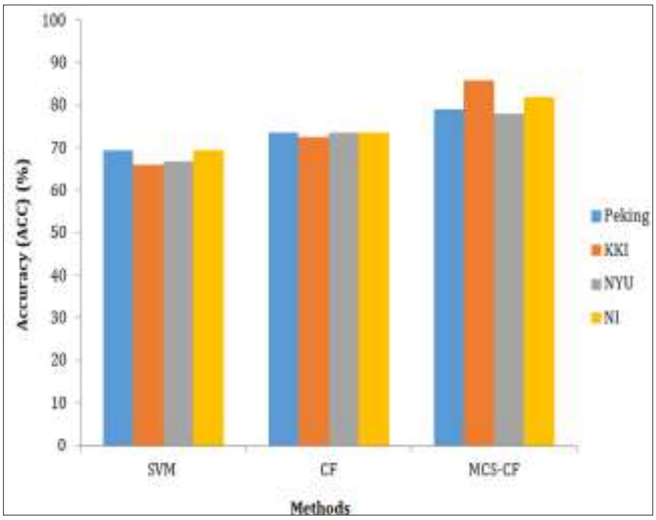

Figure 5. Accuracy (ACC) Results Evaluation of ADHD Classification Methods

Figure 5 shows the results of the accuracy results metrics with regard to three ADHD classification techniques such as the novel SVM, CF and MCS-CF. It has been proven from the figure 5, that the novel MCS-CF yields much better rates accuracy value for each dataset, and the existing techniques including SVM, CF yield slow rate of accuracy correspondingly.

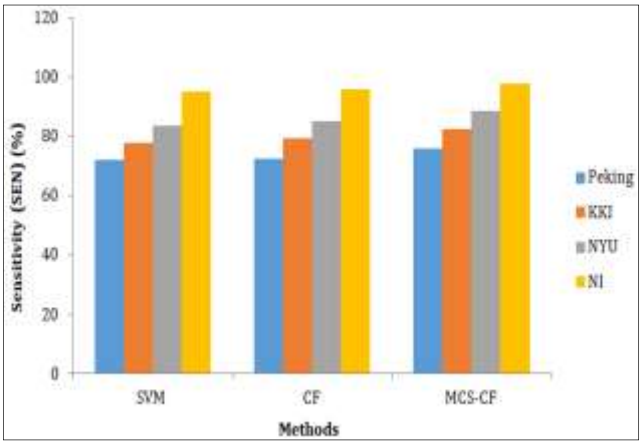

Figure 6. Sensitivity (SEN) Results Evaluation of ADHD Classification Methods

Figure 6 illustrates the results of the sensitivity comparison between the three ADHD classification techniques, which include SVM, CF and MCS-CF. The results show that the novel MCS-CF classification approach yields much better results of sensitivity for each of the datasets, while other existing techniques including CF, MCS-CF yields lower value correspondingly.

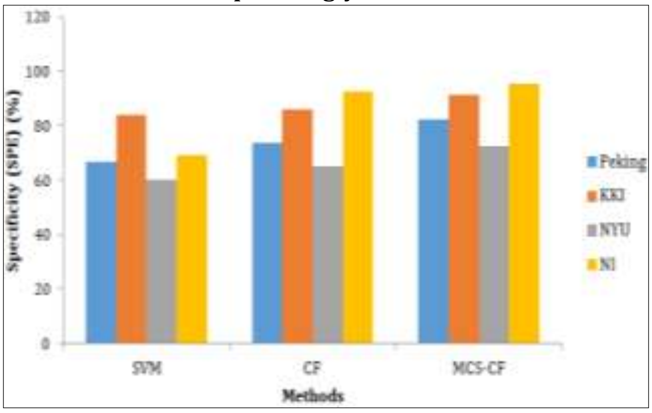

Figure 7. Specificity (SPE) Results Evaluation of ADHD Classification Methods

Figure 7 illustrates the results of the specificity comparisons of the three ADHD classification techniques. The results show that the novel MCS-CF classification technique yields much better rates of specificity for each of the datasets, while other existing techniques like SVM, CF yield low value correspondingly.

Conclusion and Future Work

Attention deficit hyperactivity disorder (ADHD) is a type of psychological syndrome, which is frequently observed amid younger children. In the form of a neuroimaging technology, functional Magnetic Resonance Imaging (fMRI) data is helpful in studying the ADHD's pathology and helps in its diagnosis. Automated ADHD diagnostic algorithms obtained from MRI/fMRI data is a huge challenge. Therefore, fMRI data is used indifferent machine learning approaches and Deep Neural Networks (DNNs) have been utilized for ADHD diagnosis. The deep forest makes use of a structure that is multiple layered where every layer has several random forests, and it offers competitive performance compared to DNNs. In comparison with DNNs which need training data in large-scale and massive efforts in tuning the hyper-parameter, the training of gcForest is easier and it offers good performance even when using the small-scale training data. The fMRI data values of ADHD were perfectly analyzed with the objective of enhancing the feature extraction, selection and classification procedure. In this work, a modified gcForest technique is proposed to find the ADHD and normal controls. This research work includes three important steps, which include (i) Preprocessing the Data and Extraction of Feature (ii) Selection of Feature, and (iii) Classification. The steps of preprocessing included correction of slice time, correction of motion, normalization, filtration of band pass and smoothing. After this, two types of features, which include 1-D Functional Connectivity (FC) and 3-D Amplitude of Low Frequency Fiuctuations (ALFF) are extracted from the fMRI data. Feature selection is an important step since often repetitive or corruptive data, which could result in degradation of the overall performance of the model developed. MCS algorithm utilized for feature selection aids in reducing the feature space which in turn,boosts the prediction accuracy and reduces the time taken for computation. At last, CF is introduced for classification and the selected feature vector samples act as the input. In addition, with the aim of handling the imbalanced data characteristic, the SMOTE is used with ENN for minority samples' generation. The experiments are carried out on ADHD-200 open sets of test data and its presentation is evaluated on the hold-out testing data. The ADHD fMRI data utilized are obtained from ADHD-200 Global competition. These experiments are carried out on four set of test datas, such as Peking, KKI, NYU and NI. The classification techniques are compared in terms of the metrics such as Accuracy (ACC), Sensitivity (SEN) and Specificity (SPE).The novel MCS-CF attained the highest degree of accuracy results in comparison with other available techniques such as SVM, CF.

The work intended for the future:

Proposed technique can also be used for the diagnosis of additional disease with fMRI data, like Alzheimer's disease and autism etc.

Functional Principal Component Analysis (PCA) kind of techniques for fMRI processing of information can be considered.

References

Lee H, Li Y, Yeh SC, Huang Y, Wu Z, Du Z. ADHD assessment and testing system design based on virtual reality. *In 2nd International Conference on Information Technology (INCIT)* 2017: 1-5.

Öztoprak H, Toycan M, Alp YK, Arıkan O, Doğutepe E, Karakaş S. Machine-based learning system: Classification of ADHD and non-ADHD participants. *In 25th Signal Processing and Communications Applications Conference (SIU)* 2017: 1-4.

Fahimi F, Guan C, Goh WB, Ang KK, Lim CG, Lee TS. Personalized features for attention detection in children with Attention Deficit Hyperactivity Disorder. *In 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2017: 414-417.

Zhou Z, Feng J. Deep forest: Towards an alternative to deep neural networks. *Proceedings of 26th International Joint Conference on Artificial Intelligence* 2017: 3553-3559.

Mwamba HM, Fourie PR, van den Heever D. PANDAS: Paediatric Attention-Deficit/Hyperactivity Disorder Application Software. *Applied Sciences* 2019; 9(8): 1645-1657.

Miao B, Zhang Y. A feature selection method for classification of ADHD. *4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS)* 2017: 21-25.

Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schweren LS, de Zeeuw P. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *The Lancet Psychiatry* 2017; 4(4): 310-319.

Sen B, Borle NC, Greiner R, Brown MR. A general prediction model for the detection of ADHD and Autism using structural and functional MRI. *PloS one* 2018; 13(4): 1-28.

Pereda E, García-Torres M, Melián-Batista B, Mañas S, Méndez L, González JJ. The blessing of Dimensionality: Feature Selection outperforms functional connectivity-based feature transformation to classify ADHD subjects from EEG patterns of phase synchronisation. *PloS one* 2018; 13(8): 1-24.

Chen M, Li H, Wang J, Dillman JR, Parikh NA, He L. A Multichannel Deep Neural Network Model Analyzing Multiscale Functional Brain Connectome Data for Attention Deficit Hyperactivity Disorder Detection. *Radiology: Artificial Intelligence* 2019; 2(1): e190012.

Riaz A, Asad M, Alonso E, Slabaugh G. Fusion of fMRI and non-imaging data for ADHD classification. *Computerized Medical Imaging and Graphics* 2018; 65: 115-128.

Shao L, Zhang D, Du H, Fu D. Deep forest in ADHD data classification. *IEEE Access* 2019; 7: 137913-137919.

Zou L, Zheng J, Miao C, Mckeown MJ, Wang ZJ. 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. *IEEE Access* 2017; 5: 23626-23636.

Wang T, Kamata SI. Classification of Structural MRI Images in Adhd Using 3D Fractal Dimension Complexity Map. *In IEEE International Conference on Image Processing (ICIP)* 2019: 215-219.

http://fcon_1000.projects.nitrc.org/indi/adhd200/

Chao-Gan Y. Data Processing Assistant for Resting-State fMRI (DPARSF) 2014. http://rfmri.org/DPARSF

Afyouni S, Smith SM, Nichols TE. Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. NeuroImage 2019; 199: 609-625.

Resting-State fMRI Data Analysis Toolkit plus V1.23 (RESTplus V1.23) 2016. http://restfmri.net

Cheng Z, Wang J, Zhang M, Song H, Chang T, Bi Y, Sun K. Improvement and application of adaptive hybrid cuckoo search algorithm. *IEEE Access* 2019; 7: 145489-145515.

Zhang M, He D, Zhu C. Cuckoo Search Algorithm Based on Hybrid-Mutation. *12th International Conference on Computational Intelligence and Security (CIS)* 2016: 538-542.

Majumdar D, Mallick S. Cuckoo search algorithm for constraint satisfaction and optimization. *Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)* 2016: 235-240.

Liu, L, Liu, X, Wang, N, Zou, P. Modified cuckoo search algorithm with variational parameters and logistic map. *Algorithms* 2018; 11(3): 30-40.

López V, Fernández A, García S, Palade V, Herrera F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information sciences* 2013; 250: 113-141.

Fernández, A, del Río, S, Chawla, N. V, Herrera, F. (2017). An insight into imbalanced big data classification: outcomes and challenges. *Complex & Intelligent Systems* 2017; 3(2): 105-120.

Abbas SR, Kadhem WJ, Abbas TM. Predicting of asymptotic properties of magnetic lens using analytical potential function. *NeuroQuantology* 2020; 18(2): 95-100. http://doi.org/10.14704/nq.2020.18.2.NQ20132

Aziz SA, Ali RS, Abd AN. Characterization studies of nickel oxide nanostructure films prepared by electrolysis method for photo detectors applications. *NeuroQuantology* 2020; 18(2): 45-49. http://doi.org/10.14704/nq.2020.18.2.NQ20123