

Article

A Novel Knowledge Distillation-Based Feature Selection for the Classification of ADHD

Naseer Ahmed Khan ^{1,*} , Samer Abdulateef Waheeb ¹ , Atif Riaz ² and Xuequn Shang ¹

¹ School of Computer Science and Technology, Changan Campus, Northwestern Polytechnical University, Xi'an 710072, China; samirabdulateef@mail.nwpu.edu.cn (S.A.W.); shang@nwpu.edu.cn (X.S.)

² Department of Computer Science, University of London, London WC1E 7HU, UK; atif.riaz@city.ac.uk

* Correspondence: naseerkhan@mail.nwpu.edu.cn; Tel.: +86-133-1927-3686

Abstract: Attention Deficit Hyperactivity Disorder (ADHD) is a brain disorder with characteristics such as lack of concentration, excessive fidgeting, outbursts of emotions, lack of patience, difficulty in organizing tasks, increased forgetfulness, and interrupting conversation, and it is affecting millions of people worldwide. There is, until now, not a gold standard test using which an ADHD expert can differentiate between an individual with ADHD and a healthy subject, making accurate diagnosis of ADHD a challenging task. We are proposing a Knowledge Distillation-based approach to search for discriminating features between the ADHD and healthy subjects. Learned embeddings from a large neural network, trained on the functional connectivity features, were fed to one hidden layer Autoencoder for reproduction of the embeddings using the same connectivity features. Finally, a forward feature selection algorithm was used to select a combination of most discriminating features between the ADHD and the Healthy Controls. We achieved promising classification results for each of the five individual sites. A combined accuracy of 81% in KKI, 60% Peking, 56% in NYU, 64% NI, and 56% OHSU and individual site wise accuracy of 72% in KKI, 60% Peking, 73% in NYU, 70% NI, and 71% OHSU were obtained using our extracted features. Our results also outperformed state-of-the-art methods in literature which validates the efficacy of our proposed approach.

Keywords: ADHD; autoencoder; classification; connectivity; features selection; neural networks; fMRI; rs-fMRI



Citation: Khan, N.A.; Waheeb, S.A.; Riaz, A.; Shang, X. A Novel Knowledge Distillation-Based Feature Selection for the Classification of ADHD. *Biomolecules* **2021**, *11*, 1093. <https://doi.org/10.3390/biom11081093>

Academic Editor: Ryuji Hamamoto

Received: 17 June 2021

Accepted: 21 July 2021

Published: 23 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brain is considered the most intricate and mysterious organ in the human body with complexity in networks in the spatial as well the temporal domain. Brain's functional and physical levels consists of five major regions: Frontal, Occipital, Parietal, Subcortical, and Temporal regions [1]. The complexity of the human brain is related to both the increasing age and difficulty level of the computational task. Increasing age of the person and the complicated nature of the task make it difficult for the brain to make decisions [2,3]. The volume of data generated by the human brain is huge and this amount of data in just half a minute is equivalent to the data generated by the Hubble telescope in its entire life [4], which makes analysis based on human brain data a challenging task.

Attention Deficit Hyperactivity Disorder (ADHD) is a brain disorder that is characterized with persistent lack of attention, high impulsiveness, restlessness, and hyperactivity with numerous environmental, neurological, and genetic factors [5–7]. The rate of ADHD diagnosis is increasing in children and it affects 8% to 12% of the world's child population as indicated in the studies in [8,9]. A benchmark for the prevalence of ADHD among children using meta-analysis based on 179 estimates of the prevalence in 175 studies is proposed in [10]. There are both genetic- [11] and neurological-related [12] interpretations of the cause of ADHD, specifically, genes LPHN3 and CDH13 and damage to the frontal lobe, respectively.

ADHD on the subjects is measured using a variety of modalities with each modality has peculiar characteristics. These modalities include DTI (Diffusion Tensor Imaging), EEG (Electroencephalography), fMRI (Functional Magnetic Resonance Imaging), PET (Positron Emission Tomography), and SPECT (Single Photon Emission Computed Tomography). Brain structural alterations were observed in proband, alterations in brain functional connectivity and influence of the drug in the treatment of ADHD using the DTI modality were discussed in the studies in [12–14]. Diagnostic psychiatry tests for ADHD were based on four steps using the the 17 meta-studies and a meta-analysis based on randomized control trials on the ADHD were discussed in [15,16], respectively. A triple-blinded studies on the 275 children and adolescents to integrate the biomarkers for the diagnosis of ADHD and a deep learning-based framework for the diagnosis of ADHD were discussed in [17,18]. A convolution neural network and a comparisons of alpha powers between the 25 patients and 22 healthy controls for the diagnosis of ADHD using the EEG modality were discussed in [19–21], respectively. Effect of psychostimulants on the 16 youths, with modeling of the brain using the resting state of the brain with the help of Independent Component Analysis (ICA) and A meta-analysis of 55 studies involving 55 children and adults were discussed in [22–24], respectively. The inter-connections and intra-connections in the brain functional regions, effect of 40 mg methylphenidate on the 37 individuals, and effect of L-theanine (2.5 mg) and caffeine (2.0 mg) on the patients with ADHD using fMRI modality were studied in [25–27], respectively. PET-related studies have also shown promising results on the ADHD subjects and the biology of this disorder [28]. A significant increase in Dopamine Transporter (DAT) binding was observed in [29] conducted on 47 subjects with matched control. In [30], alterations in the cortical thickness were found between the ADHD and the healthy controls. One multimodal study [31] using PET and genetic data on the 20 ADHD and matched healthy controls and another machine learning-based study [32] on 16 ADHD subjects and 22 healthy controls found promising results in the diagnosis of this disorder. SPECT-based studies were conducted in [33] to distinguish sub-types of ADHD. A meta-analysis [34] involving 51 studies on 53 ADHD subjects found 13 promising genes for the diagnosis of ADHD using the SPECT-based image scans on the patients. A study on ADHD and other disorders using the SPECT modality showed alterations in some of the brain regions [35]. A genetic SPECT study [36] found a decrease in DAT on the 20 adolescents. A use of SPECT as to how it is aiding the medical treatment of the ADHD subject is discussed in the study [37].

fMRI studies based on functional connectivity on the ADHD are becoming, of late, very popular due the the noninvasive nature of fMRI and interpretable regions found from the extraction of functional connectivity matrix. In this regard, ADHD diagnosis using personal characteristics (age, IQ, and handedness) [38] showed promising results. Independent Component Analysis (ICA) with the combination of functional connectivity matrix found neural network dysregulation in ADHD [39]. Fusion [40] of non-imaging data with the imaging data also showed promising results in the study. fMRI study also found functional connectivity alterations [41] in the right inferior frontal cortex of the adolescents. A study using Convolution Neural Network (CNN) [42] on the multi-site resting state fMRI data showed promising results in the classification of the ADHD .

In this study, we propose a Deep Learning-based approach for the classification of ADHD subjects and healthy controls. Our approach is inspired from the work in [43], as a similar distillation model is conceptualized in that problem, but the novelty of our work is that we have used this model as a part of the bigger pipeline of feature selection. We used knowledge distillation as a first step to find the indices of the most discriminating features from the final feature index vector and after that to selection of the most discriminating features subset using Sequential Forward Feature Selection (SFFS) Approach. Details of our knowledge distillation process is as follows. First a connectivity matrix between the each pair of region on a subject is computed using a “community matrix” which is described in Section 3.3.2 and the lower triangular unique elements excluding the diagonal are extracted from this matrix to be used as an “Input” in our study. In phase one, we trained a large

neural network on all the input corresponding to all the subjects and extracted the trained codes of the hidden layer, the idea was that the large neural network should be able to understand the complex structure that existed between the Input and the output condition of ADHD and healthy controls. In the second phase, we trained a one hidden layer-based autoencoder on the input connectivity features and the extracted trained codes on the subjects from the large neural work. Here, the purpose of autoencoder was to be able to learn the underlying structure between the input connectivity features and the trained codes. After that we extracted the weights from the input to hidden layer matrix from the autoencoder and sorted them in descending order of magnitude. The higher the value of the weight, the most likely that feature index correspond to most discriminating feature vector. In the final phase, as we needed a set of most discriminating features, we used a Sequential Forward Feature Selection (SFFS) approach to select the subset of features that were most discriminant using various classifiers and compared our results using the experiments that we described in the experimental subsection. In the sections to follow, we will proceed as follows. First, we will discuss about the related research. Second, we will discuss about the dataset and our proposed methodology. Third, we will discuss in details about the feature selection for the classifiers. After that we will discuss the and interpret the usefulness of the related features, selected using our approach. Last, we will conclude our study, will discuss its limitations and future improvement to solve these issues.

2. Related Research

In this section, we have categorized and explained various studies that were conducted on ADHD using the fMRI modality.

2.1. Structural Information Based Approaches

In [44], the authors used morphological information to classify 210 ADHD subjects from the 226 healthy controls. They used isotropic local binary patterns on three orthogonal planes to extract features from the high-resolution MRI scan data on the subjects resulting in 69% accuracy. In the study [45], high-resolution 3-D scans of 55 ADHD subjects and matched healthy controls were acquired using MRI machine. After processing them with the FreeSurfer [46] software, 340 features such as cortical thickness, curvature, volume, etc. were measured for each type of subjects resulting in maximum accuracy of 90.18% when given to Extreme Learning Machine classifier. Gaussian Process Classification was applied in [47] to the brain gray matter volumetric data including 29 ADHD and matched control subjects resulting in overall accuracy of 79.3%. Structural as well as functional features were used in [48,49] to classify the ADHD subjects resulting in 76% accuracy in the multi-class setting and 92.8% accuracy in the binary class setting in the first study and 67% accuracy in the second study. A study [50] on 508 individuals containing ADHD subjects and healthy control using the source based morphometry of the brain scans showed alterations in bilateral CrusI and bilateral insula between the two conditions among subjects.

2.2. Functional Connectivity Based Approaches

In [51], decreased functional connectivity was observed in dorsal anterior cingulate cortex and regions of default mode network between the 21 ADHD patients and 21 matched healthy controls. In study [52], 20 medication-naive ADHD children with 20 age- and gender-matched healthy controls were investigated for the alterations in functional connectivity and found delayed maturation in two functional networks. In [53], functional connectivity alterations in the brain areas related to motor circuitry which contribute to the functioning of motor and attention were exhibited in children. A Fully Connected Cascade (FCC) neural network was proposed in [54] to discriminate ADHD from the healthy controls, and directional and non-directional based connectivity features were given to the classifier resulting in 90% accuracy. In [55], involving 20 ADHD patients and 27 healthy controls, increased connectivity in the brain Default Model Network (DMN) was found both between and among the functional connectivity networks. In [56], from the

data on 95 ADHD subjects and 90 healthy controls, the authors selected five subcortical regions. Their analysis showed significant difference in resting state functional connectivity in caudate nucleus. In [57], the authors formed two cohorts: a child cohort consisting of 34 ADHD and 28 health controls and an adult cohort consisting of 112 ADHD and 77 healthy controls. Functional connectivity alterations were found both in the children cohort and in the adult cohort. A multi-objective scheme using Support Vector Machine (SVM) was used in [58] to first tackle the task of imbalanced dataset and then classifying the ADHD subjects from the healthy controls with promising results. The dual subspace method was observed in [59] by first making two subspaces corresponding to ADHD and healthy control and then using them based on the energy principle to classify ADHD from the healthy controls.

2.3. Deep Learning-Based Approaches

Deep learning is a computational model using which we learn multilevel abstraction of the input data and learn the intricate pattern from the data by training layer wise feed forward neural network with back-propagation algorithm [60]. Deep learning is closely associated to machine learning which has applications in various practical domains such as IoT, renewable energy, medicine, and agriculture [61–64]. Of late, deep learning is being used increasingly more in the medical image analysis domain to replace handcrafted features with automatic extracted features [65,66]. In [67], the authors discussed scenarios where the subdomains of deep learning including computer vision, natural language processing and reinforcement learning can be applied in the healthcare setups. A 4D-CNN-based algorithm was proposed in [68] with data augmentation for balancing to extract both spatial and temporal features from the ADHD subjects and healthy controls resulting in 71.3% accuracy. DeepfMRI was proposed in [69], three networks—a feature extractor network, a functional connectivity network, and a classification network—were all combined into one big network to form an End-To-End approach with promising results across three ADHD sites. In [70], a Convolution Denoising Autoencoder (CDAE) was used to extract the discriminating features between the ADHD subjects and healthy control and then the Adaptive boosting Decision Trees (AdaDt) was used for classification using the extracted Features.

3. Materials and Methods

3.1. Dataset

The dataset used in this study is ADHD-200 [71], which is compiled by the International NeuroImaging Datasharing Initiative (INDI) consortium, consisting of eight international imaging sites around the world. This dataset consists of 776 training and 197 testing subjects with information on gender, age, IQ, and handedness. The dataset that we used consists of five different sites out of total 8 sites which are Kennedy Krieger Institute (KKI), NEUROIMAGE Sample (NI), New York University Child Study Center (NYU), Peking University (PEK), and Oregon Health and Science University (OHSU). The INDI consortium has also provided the testing dataset for the six sites, for comparison with state-of-the-art methods we have only used dataset from the five sites. Details of training and testing dataset along with phenotypic condition is presented in Table 1.

3.2. Preprocessing of the Dataset

The ADHD-200 dataset was preprocessed with the Neuroimaging Analysis Kit (NIAK). The steps of the preprocessing include, slice time correction, motion correction, coregistration, normalization, quality control, spatial smoothing, and many others as described in [72–74]. Moreover, each of the sites has different scanning parameters that they used for the participants and varying inclusion, exclusion, and update criteria. For example, OHSU site has problem in four subjects in T1-fMRI coregistration step they made adjustment with alignment and after that updated those four subjects.

Table 1. ADHD-200 Preprocessed Dataset.

Sr	Imaging Site	Training			Testing		
		ADHD	HC	Total	ADHD	HC	Total
1	KKI	18	51	69	3	8	11
2	NI	25	23	48	11	14	25
3	NYU	118	98	216	29	12	41
4	OHSU	37	42	79	6	28	34
5	Peking_1	24	61	85	24	27	51
Total		497			162		

3.3. Methodology

3.3.1. Affinity Propagation

Clustering is a crucial step in forming the community matrix that we used as the input in our proposed approach. Clustering approaches suffer from the drawback of pre-specifying the number of clusters to be formed out of the dataset. We have used Affinity Propagation [75] clustering message passing approach in our algorithm that does not need pre-specification of number of clusters. An important consideration in using Affinity Propagation is the value of preference as this value affects the number of clusters. In our proposed approach, we set preference value from negative $-10,000$ to $+10,000$ in increments of 100 so that the complete distribution of the preference value is taken into consideration when calculating the community connectivity matrix.

3.3.2. Architecture of the Proposed Approach

The detailed architecture of the proposed approach is displayed in Figure 1. Our proposed approach works by training of a large neural network to learn the representation of the features, an autoencoder for features reconstruction and finally a feature selection method that we explain in the following three sections.

3.3.3. Latent Representation of Dataset

We converted the subject matrix corresponding to the ADHD and Control subjects into the feature vector that were extracted using the community matrix and trained a large neural network based on these features on the whole dataset as displayed under the title of “Training Neural Network” in Figure 1. The core purpose of training a relatively large neural network was to learn the trained embeddings corresponding to the ADHD and Control subjects that could be used in the later stage of knowledge distillation.

3.3.4. Knowledge Distillation

After learning the trained embeddings on the dataset, we used a one-layer Autoencoder with the output layer set to the trained embeddings and the input layer set to the community based features. This is called the knowledge distillation phase as we want our features get reconstructed based on the bottleneck layer so that discriminating features could be sorted as displayed under the title of “Knowledge Distillation” in the Figure 1.

3.3.5. Sequential Forward Feature Selection

Finally, a sequential features selection algorithm, which will select the features in stepwise increments, is used. It selected the most discriminating features corresponding to ADHD and Control subjects as displayed under the title of “Features Selection” in the Figure 1. The reason we want to use the sequential feature selection approach is that we do not want to miss the discriminating features.

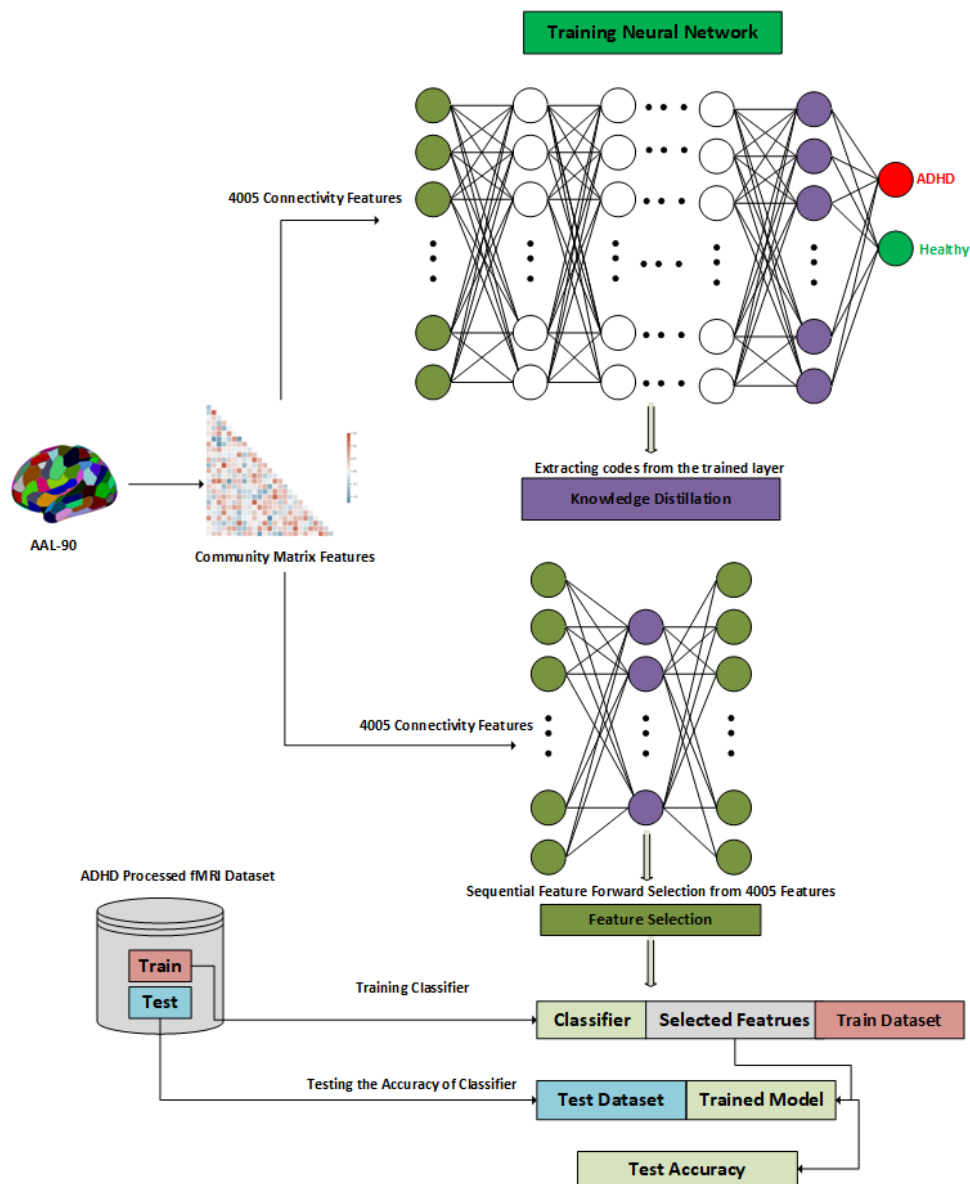


Figure 1. Proposed architecture of the feature selection approach.

3.4. Algorithms

We describe now the algorithms that are used in our proposed approach.

3.4.1. Connectivity Matrix

We have used a connectivity matrix based features as the input to our model, that is described in detail in the Algorithm 1. Here, first a clustering of the subject region wise time series is done using the Affinity Propagation [75] which will assign the clusters to each of the 90 regions of the subjects which we can name as a region clustered labels. Then, a matrix of 90×90 regions is constructed based on the similarity of the labels of the region clustered labels. The reason of using Affinity Propagation or message passing algorithm is that it does not require predefined number of clusters. In this algorithm, “X” is the input connectivity matrix, “P” is the preference value and “AffProp” is the Affinity Propagation algorithm.

3.4.2. Community Matrix

A correlation matrix based on the Pearson correlation [76] is constructed for the connectivity matrices; however, here we have used the community matrix as described

in [77] instead. The details of the community matrix are also described in the Algorithm 2 where different connectivity matrix based on the change in the “preference” parameter in the Affinity propagation is used to construct a number of connectivity matrix which are then averaged together. The core reason for using the community matrix instead of the correlation matrix as described in the study [77] is the sparseness of the community matrix. Moreover, the value of the community matrix corresponding to the two brain regions measures the probability of those two regions to be in the same cluster. Connectivity matrices are usually constructed using the Person’s correlation coefficient and they are easy to calculate. However, two regions which are not connected still have a value corresponding to their correlation which makes it difficult for the classifier to ignore this value. On the other hand, community-based connectivity matrices are calculated using the clustering method. Therefore, two regions which are in the same cluster are more probable to be connected. Due to clustering the community matrix based connectivity matrix is more sparse and make it easy for classifier in learning the pattern of the latent dimension from the data. In this algorithm “X” is the input connectivity matrix and “P” is the preference value of the Affinity Propagation algorithm.

Algorithm 1 Connectivity Matrix

Input: $X_{M \times N}$, P
Output: $C_{R \times R}$

- 1: $C_{R \times R} \leftarrow 0$
- 2: $Y_R \leftarrow \text{AffProp}(X_{m \times n}, P)$
- 3: **for** $i = 0$ **to** R **do**
- 4: **for** $j = 0$ **to** R **do**
- 5: **if** $Y_i = Y_j$ **then**
- 6: $C_{i,j} = 1$
- 7: **else**
- 8: $C_{i,j} = 0$
- 9: **end if**
- 10: **end for**
- 11: **end for**

Algorithm 2 Community Matrix Features

Input: $X_{M \times N}$, $PArray$
Output: $M_{R \times R}$

- 1: $M_{R \times R} \leftarrow 0$
- 2: **for** $i = 0$ **to** $\text{Len}(PArray)$ **do**
- 3: $P \leftarrow PArray_i$
- 4: $C \leftarrow \text{ConnectivityMatrix}(X_{m \times n}, P)$
- 5: $M \leftarrow M + C$
- 6: **end for**
- 7: $M \leftarrow M / \text{Len}(PArray)$
- 8: $M \leftarrow \text{LOWERTRIA}(M)$

3.4.3. Ranking of Features

The details of features ranking are explained in Algorithm 3, where the first community matrix features are fed to a large neural network and the learned codes extracted from the large neural network are fed to a one layer Autoencoder, and finally a ranking of the features is obtained using a simple Argsort method. In this algorithm ‘X’ is input connectivity matrix, ‘PArray’ is the list of preference values, and ‘LOWERTRIA’ is the lower triangular elements of the square matrix.

Algorithm 3 Features Ranking Algorithm

Input: $F_{M \times N}$
Output: $I_{N \times 1}$
1: $CODES_{M \times N} \leftarrow TrainingNeuralNetwork(F_{M \times N})$
2: $W_{M \times D} \leftarrow KnowledgeDistill(F_{M \times N}, CODES_{M \times N})$
3: $SCORES \leftarrow Diag(W.W^T)$
4: $I \leftarrow ARGSORT(SCORES)$

3.4.4. Features Subset Selection

ADHD is a disease in which not just one region but a subset of regions are involved, as described in the “Introduction” section. Therefore, after ranking the features we need to find a subset of the most discriminating features between ADHD and Controls which is explained in Algorithm 4, where a stepwise forward feature ranking algorithm is used to extract the most discriminating subset of features which could apart Healthy controls from the ADHD. In this Algorithm, ‘X’ is input connectivity matrix, ‘I’ is the indexed of the features, CLF’ is a classifier, and ‘ACCURACY’ is 10-fold accuracy.

Algorithm 4 Sequential Feature Selection

Input: $X_{M \times N}$, $I_{N \times 1}$, CLF
Output: F
1: $featuresBox \leftarrow Empty$
2: $accList \leftarrow Empty$
3: **for** $i = 0$ to $Len(I)$ **do**
4: $features \leftarrow I[0$ to $i]$
5: $DataMatrix \leftarrow X[features]$
6: $ACC \leftarrow ACCURACY(CLF[DataMatrix])$
 $featuresBox.append(features)$
 $accList.append(ACC)$
7: **end for**
8: $maxIndex \leftarrow ArgMax(accList)$
9: $F \leftarrow featureBox[maxIndex]$

3.5. Experimental Settings

A challenging area in training a Neural Network is the setting the values of various hyperparameters of the model. In this study, we used two types of neural network for training and Knowledge Distillation. For learning a latent representation of all the data we used a 13 layer Neural Network model with the input layer set to 4005 connectivity features units and the output layer set to 1 ‘sigmoid’ units corresponding to the ADHD disease presence or absence. The hidden layers of this model were set to 3000, 2500, 2000, 1500, 1000, 500, 100, 50, 10, 5, and 2 units, respectively. A dropout of 0.3 in the initial 3 hidden layers and then 0.2 and 0.1 were used in the later layers. We used ‘tanh’ activation function so that negative values are not clipped away, a batch size of 16, an epoch size of 1000, and optimizer ‘adadelta’ with a learning rate of 0.01 was used for training this large neural network. The combinations of these hyperparameters were used after observing too many oscillations in loss and slow convergence, which are typical in training neural networks. After training, these neural networks codes were extracted from layer 10 of this network and then an Autoencoder with 4005 connectivity units, a hidden layer of 100 units, and the output layer of the trained 10 codes were used for the training with the same hyperparameters settings as in the Neural Network with the exception of optimizers which was set to ‘mse’ as this time task was not classifications but of reconstruction. Note that extraction of codes from a specific layer is a challenging task, as layers with too many parameters or with too few parameters do not produce good reconstruction results. Another justification for selecting the layer from the later layers of the trained neural

network is that the final layers of the neural network capture a more abstract representation of the input data in the latent dimension.

4. Features Selection

The proposed feature selection philosophy using various classifiers is explained in the Figure 2 where the x-axis represents the number of features(4005) and the y-axis represents the 10-fold accuracy. We have employed various classifiers to check the validity of our feature selection approach so that the robustness of our feature selection approach be emphasized. It can be observed using Figure 2. Initially, the accuracy is low because of the lesser number of features then the accuracy increases for the classifiers until it reached the maximum level and from there it again starts to decrease. In the following three sections we will explain the reasons for having such a phenomena in the case for all the classifiers.

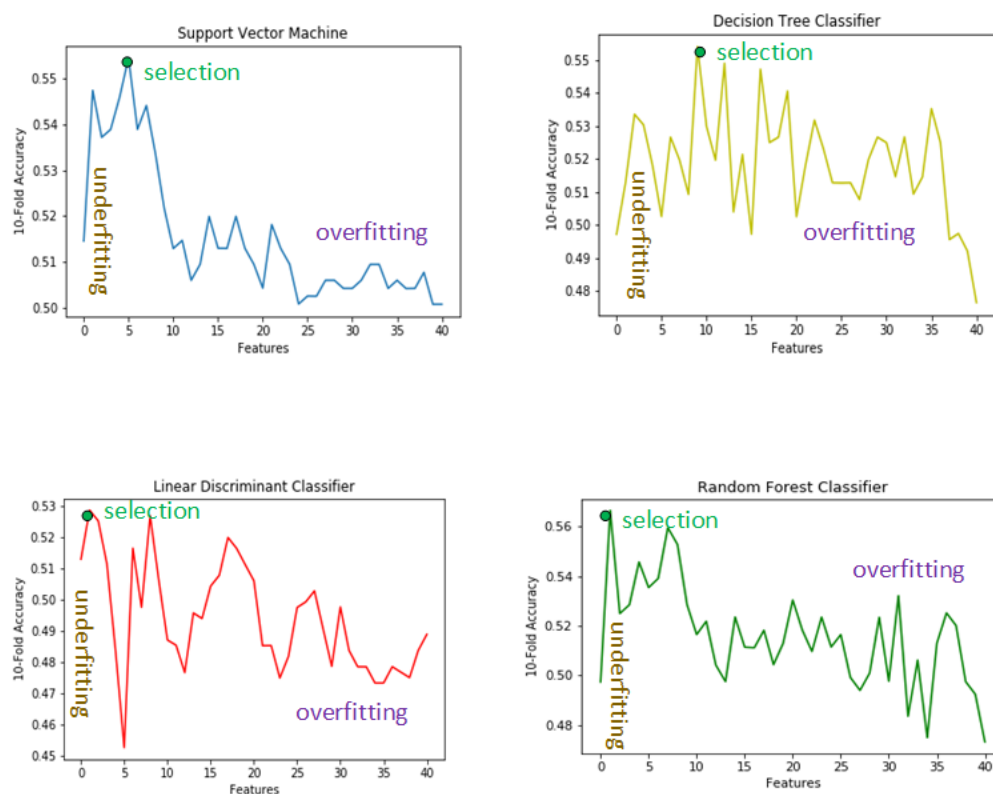


Figure 2. Feature selection approach.

4.1. Low Number of Features

A low features set represents the case of underfitting which is a common problem in all those statistical models where the complex phenomena is to be modeled on the given data containing large number of features. In the under-fitting scenario the model is too simple to represent the complex relationship that exists in the problem under hand. As can be seen in Table 2, as the number of features or features set size increases, so does the accuracy of the classifiers; this validates the phenomena of underfitting or bias of the machine learning models. This increasing accuracy corresponding to features set size in all the classifiers thereby ascertaining our assertion of the presence in bias in the model.

4.2. Large Number of Features

Similarly, having a higher number of features set represents the typical overfitting scenario, that is, when the model becomes too complex that it cannot correctly predict the result for unseen data. It can be seen from Table 3 that there exists an inverse relation between the number of features and the accuracy of the classifier. Therefore, for all the classifiers, as the set of top features set increases, the accuracy of the classifiers decreases.

This phenomena validates our assertion of overfitting or high variance that becomes prevalent in case of selection of high number of features.

Table 2. Accuracy comparison of low number of features.

Top Features	SVM	LDC	DTC	DTC
1	0.49	0.5	0.49	0.49
5	0.53	0.55	0.51	0.53
10	0.54	0.56	0.52	0.54
15	0.55	0.56	0.53	0.56
20	0.55	0.57	0.54	0.56

Table 3. Accuracy Comparison of High Number of Features.

Top Features	SVM	LDC	DTC	DTC
100	0.55	0.53	0.54	0.53
500	0.55	0.52	0.55	0.53
1000	0.54	0.52	0.52	0.52
1500	0.52	0.51	0.48	0.45
3000	0.3	0.2	0.4	0.3

4.3. Selected Features

After adjusting the cases of underfitting and overfitting as described in the previous two subsections we selected the features that resulted in the highest accuracy for the classifiers. The selected features corresponding to the classifiers are tabulated in Table 4.

Table 4. Selection of the Features.

Classifier	Selected Features
SVM	22
LDC	43
DTC	589
RFC	64

4.4. Support Vector Machine Features

As can be observed in Figure 3, those features corresponding to the SVM classifiers are the smallest and those corresponding to the decision tree classifiers are the largest.

Our dataset, as described in Table 1, is imbalanced, that is, there exists unequal cases of ADHD class as compared to HC subjects. SVM has been extensively studied in the literature under the scenario of imbalanced data and there exist optimized decision threshold [78] methods and kernel scaling [79] techniques that can make SVM perform well in the imbalanced cases. Therefore, based on these observations we selected SVM features, as these features are common in all the classifiers and they validate the robustness of this small number of features. Henceforth, in the following sections we will compare results based on the selected SVM features and discuss the significance of these selected features from the anatomical point of view also in the later section.

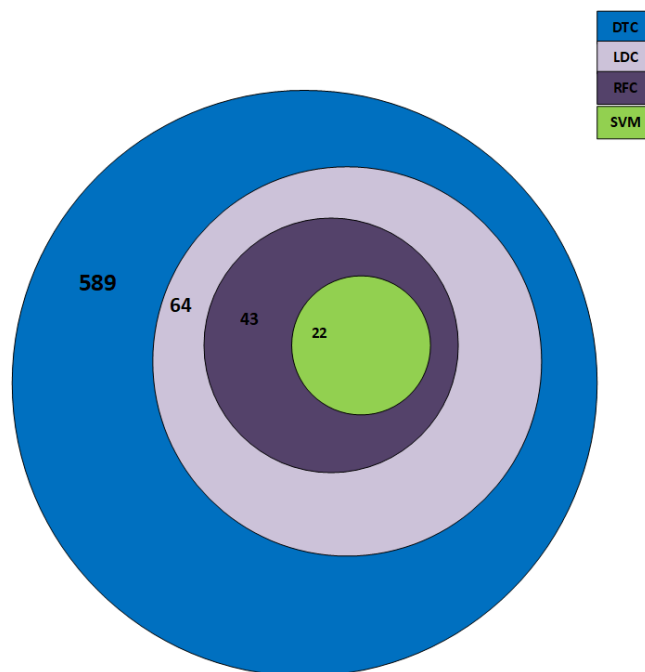


Figure 3. Intersection of features corresponding to classifiers.

5. Results

Our proposed feature selection approach has produced promising results when compared with the state-of-the-art methods available in the literature. In the following two experiments, we will describe our results based on the two scenarios that are used in literature for the ADHD classification problem.

5.1. Training Using Combining Sites

In the first experiment, we trained our model using all the training dataset available in the ADHD-200. The training dataset corresponding to all five sites in Table 1 was combined, and then for testing, the trained model the testing dataset corresponding to each site in Table 1 was used . It can be seen in the Table 5 that our model performed well on all the sites and it was able to perform well when compared to the state-of-the-art methods. Moreover, the accuracy on the remaining two sites for which the comparison accuracy was not available was also very promising.

Table 5. Accuracy(%) comparison of trained model using combined and individual sites.

Sr	Site	Trained on Each Site		Trained on Combined Sites	
		Proposed	DeepFMRI (2020) [69]	Proposed	DeepFMRI (2020) [69]
1	NYU	73	73.1	56	65.8
2	NI	70	67.9	64	60
3	Peking	60	62.7	60	43.1
4	OHSU	71	-	56	-
5	KKI	72	-	81	-

5.2. Testing on Benchmark Datasets

In the second experiment, we trained the model on the training dataset for each site presented in Table 1 and tested the accuracy of our model on the benchmark testing dataset that is presented in the Table 1. Our accuracy results are promising in all the sites as can be seen in Table 6 and robustness of our features can be validated based on the observations that for other model where the accuracy was highest in one site and lower

in other, our model still consistently performed better in all scenarios. The accuracy of SC-CNN-ATT [42] is high in the site NYU but low in the sites NI and Peking. Similarly for the work in [80], the accuracy is very high in the site NYU but low in the other two sites. A repeatable observation can be seen in all other methods presented in Table 6.

Table 6. Comparison of proposed method with state-of-the-art techniques.

Method	NYU	NI	Peking	OHSU	KKI
Proposed	73.3	70	73.3	71	60
DeepFMRI (2020) [69]	73.1	67.9	62.7	-	-
SC-CNN-ATT (2020) [42]	77.7	75.3	60.4	-	65.2
FCNet (2018) [40]	67.4	72.9	25.4	-	85.3
Nunez et al. (2015) [81]	-	-	56	-	58
AJHao et al. (2015) [82]	64.7	-	66.3	-	59
Dey et al. (2014) [80]	81	-	56	-	58
ADHD-200 (2012) [38]	-	56.9	35.1	-	51

6. Discussion

Graph theory [83] is often used in the analysis of various brain disorders using connectivity features. BrainNetViewer [84] is a tool that is employed to visualize the connectivity between the hemispheres as well as the intra-hemispheric regions of the brain. For the BrainNet tool, we need nodes and edges so that they can be visualized. As we used the brain AAL 90 regions atlas, we know out of 4005 features or connectivity regions which pair of regions are connected based on the selected features. Next, the regions that are connected are mapped to the nodes and their existence in the 4005 look-up corresponds to the edge between those two edges. There exists an altered connectivity pattern between the inter-hemispherical and intra-hemispherical connectivity in the two hemispherical regions of the ADHD disorder as can be seen in Figure 4. It can be seen that there is an uneven connectivity pattern in the inter-hemispherical and intra-hemispherical connectivity patterns. Our results are in agreement with the altered connectivity patterns found in [85–88].

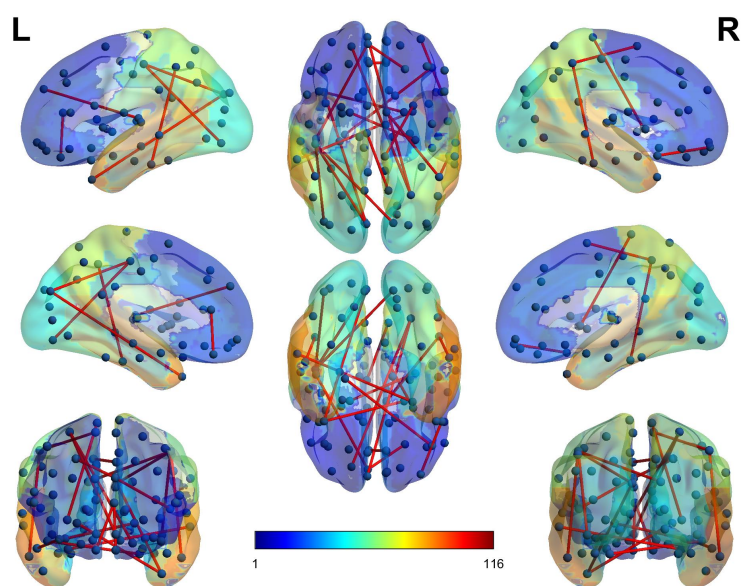


Figure 4. BrainNet visualizations of hemispherical connectivity.

7.2. Future Directions

Of late, End-to-End pipelines are getting more and more attention from the research community in the deep learning paradigm. Therefore, keeping this in view, we believe that an End-To-End deep learning model on ADHD disease based on our proposed approach is an interesting direction to consider. It is no doubt challenging to combine such a heterogeneous pipeline to a one unified model, but we believe that it will open up more opportunities for researchers in knowledge distillation- based approaches working in other domains.

7.3. Limitations

We have achieved good accuracy based on our feature selection approaches and our results outperform the best methods available in literature, but still there are two key areas that we need to consider and made progress before this technique could be adapted in studies. First, we believe that the size of the ADHD-200 Dataset is a big issue in training such classification models on ADHD disease. Particularly the availability of data on the sites KKI and NI is far too sparse when compared with the NYU site. Second, the ADHD-200 Dataset is also imbalanced and this can be seen from the data available on the sites KKI and Peking. Based on the above two key issues on the dataset, we are well aware that machine learning models are not well suited in those two cases and lead to biased results in favor of the high number sites which distorts the applicability of the machine learning models in cases where the Dataset is low in examples.

Author Contributions: N.A.K. conceptualized the idea of the Knowledge Distillation to be used in ADHD, wrote algorithms, code and the draft of the manuscript. S.A.W. did experimentation with the code and reported results. A.R. suggested the idea of community matrix to be used instead of correlation matrix for better results. X.S. administered the overall project, suggested valuable edits and arrange the administrative aspects of the Project. All authors have read and agreed to the published version of the manuscript

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 61772426).

Data Availability Statement: ADHD-200 Dataset is publicly available for download at <http://preprocessed-connectomes-project.org/adhd200/>.

Acknowledgments: We are extremely thankful to China Scholarship Council (CSC) and National Natural Science Foundation of China for giving us both the administrative and financial support to complete this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADHD	Attention Deficit Hyperactivity Disorder
DTC	Decision Tree Classifier
LDC	Linear Discriminant Classifier
RFC	Random Forest Classifier
SVM	Support Vector Machine
rs-fMRI	Resting State Functional Magnetic Resonance Imaging

References

1. Bassett, D.S.; Gazzaniga, M.S. Understanding complexity in the human brain. *Trends Cogn. Sci.* **2011**, *15*, 200–209. [[CrossRef](#)] [[PubMed](#)]
2. Bossaerts, P.; Murawski, C. Computational complexity and human decision-making. *Trends Cogn. Sci.* **2017**, *21*, 917–929. [[CrossRef](#)] [[PubMed](#)]
3. Anokhin, A.P.; Birbaumer, N.; Lutzenberger, W.; Nikolaev, A.; Vogel, F. Age increases brain complexity. *Electroencephalogr. Clin. Neurophysiol.* **1996**, *99*, 63–68. [[CrossRef](#)]

4. Abbott, A. Solving the brain. *Nature* **2013**, *499*, 272. [[CrossRef](#)] [[PubMed](#)]
5. Wender, P.H. Attention-deficit hyperactivity disorder in adults. *Psychiatr. Clin. N. Am.* **1998**, *21*, 761–774. [[CrossRef](#)]
6. Swanson, J.M.; Volkow, N.D.; Newcorn, J.; Casey, B.; Moyzis, R.; Grandy, D.; Posner, M. Attention deficit hyperactivity disorder. *Encycl. Cogn. Sci.* **2006**, *45*, 642–657.
7. Faraone, S.; Asherson, P.; Banaschewski, T.; Biederman, J.; Buitelaar, J.; Ramos-Quiroga, J.; Franke, B. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Prim.* **2015**, *1*, 15020. [[CrossRef](#)]
8. Doyle, A.E.; Willcutt, E.G.; Seidman, L.J.; Biederman, J.; Chouinard, V.A.; Silva, J.; Faraone, S.V. Attention-deficit/hyperactivity disorder endophenotypes. *Biol. Psychiatry* **2005**, *57*, 1324–1335. [[CrossRef](#)]
9. Bush, G. Attention-deficit/hyperactivity disorder and attention networks. *Neuropsychopharmacology* **2010**, *35*, 278–300. [[CrossRef](#)]
10. Thomas, R.; Sanders, S.; Doust, J.; Beller, E.; Glasziou, P. Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Pediatrics* **2015**, *135*, e994–e1001. [[CrossRef](#)]
11. Franke, B.; Faraone, S.; Asherson, P.; Buitelaar, J.; Bau, C.; Ramos-Quiroga, J.A.; Mick, E.; Grevet, E.; Johansson, S.; Haavik, J.; et al. The genetics of attention deficit/hyperactivity disorder in adults, a review. *Mol. Psychiatry* **2012**, *17*, 960–987. [[CrossRef](#)] [[PubMed](#)]
12. Chiang, H.L.; Yung-Chin, H.; Chi-Yuan, S.; Tseng, W.Y.I.; Gau, S.S.F. White matter endophenotype candidates for ADHD: A diffusion imaging tractography study with sibling design. *Psychol. Med.* **2020**, *50*, 1203–1213. [[CrossRef](#)] [[PubMed](#)]
13. Qiu, M.G.; Ye, Z.; Li, Q.Y.; Liu, G.J.; Xie, B.; Wang, J. Changes of brain structure and function in ADHD children. *Brain Topogr.* **2011**, *24*, 243–252. [[CrossRef](#)] [[PubMed](#)]
14. de Luis-García, R.; Cabús-Piñol, G.; Imaz-Roncero, C.; Argibay-Quiñones, D.; Barrio-Arranz, G.; Aja-Fernández, S.; Alberola-López, C. Attention deficit/hyperactivity disorder and medication with stimulants in young children: A DTI study. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2015**, *57*, 176–184. [[CrossRef](#)]
15. Boutros, N.; Fraenkel, L.; Feingold, A. A four-step approach for developing diagnostic tests in psychiatry: EEG in ADHD as a test case. *J. Neuropsychiatry Clin. Neurosci.* **2005**, *17*, 455–464. [[CrossRef](#)]
16. Micoulaud-Franchi, J.A.; Geoffroy, P.A.; Fond, G.; Lopez, R.; Bioulac, S.; Philip, P. EEG neurofeedback treatments in children with ADHD: An updated meta-analysis of randomized controlled trials. *Front. Hum. Neurosci.* **2014**, *8*, 906. [[CrossRef](#)]
17. Heinrich, H.; Busch, K.; Studer, P.; Erbe, K.; Moll, G.H.; Kratz, O. EEG spectral analysis of attention in ADHD: Implications for neurofeedback training? *Front. Hum. Neurosci.* **2014**, *8*, 611. [[CrossRef](#)]
18. Snyder, S.M.; Rugino, T.A.; Hornig, M.; Stein, M.A. Integration of an EEG biomarker with a clinician’s ADHD evaluation. *Brain Behav.* **2015**, *5*, e00330. [[CrossRef](#)] [[PubMed](#)]
19. Chen, H.; Song, Y.; Li, X. A deep learning framework for identifying children with ADHD using an EEG-based brain network. *Neurocomputing* **2019**, *356*, 83–96. [[CrossRef](#)]
20. Dubreuil-Vall, L.; Ruffini, G.; Camprodon, J.A. Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. *Front. Neurosci.* **2020**, *14*, 251. [[CrossRef](#)] [[PubMed](#)]
21. Deiber, M.P.; Hasler, R.; Colin, J.; Dayer, A.; Aubry, J.M.; Baggio, S.; Perroud, N.; Ros, T. Linking alpha oscillations, attention and inhibitory control in adult ADHD with EEG neurofeedback. *NeuroImage Clin.* **2020**, *25*, 102145. [[CrossRef](#)] [[PubMed](#)]
22. Peterson, B.S.; Potenza, M.N.; Wang, Z.; Zhu, H.; Martin, A.; Marsh, R.; Plessen, K.J.; Yu, S. An fMRI study of the effects of psychostimulants on default-mode processing during Stroop task performance in youths with ADHD. *Am. J. Psychiatry* **2009**, *166*, 1286–1294. [[CrossRef](#)]
23. Zuo, X.N. Modeling Resting-State Brain Functional Integration. *Nat. Preced.* **2010**. [[CrossRef](#)]
24. Cortese, S.; Kelly, C.; Chabernaud, C.; Proal, E.; Di Martino, A.; Milham, M.P.; Castellanos, F.X. Toward systems neuroscience of ADHD: A meta-analysis of 55 fMRI studies. *Am. J. Psychiatry* **2012**, *169*, 1038–1055. [[CrossRef](#)]
25. Kaboodvand, N.; Irvani, B.; Fransson, P. Dynamic synergetic configurations of resting-state networks in ADHD. *NeuroImage* **2020**, *207*, 116347. [[CrossRef](#)] [[PubMed](#)]
26. Sweitzer, M.M.; Kollins, S.H.; Kozink, R.V.; Hallyburton, M.; English, J.; Addicott, M.A.; Oliver, J.A.; McClernon, F.J. ADHD, smoking withdrawal, and inhibitory control: Results of a Neuroimaging Study with methylphenidate challenge. *Neuropsychopharmacology* **2018**, *43*, 851–858. [[CrossRef](#)] [[PubMed](#)]
27. Kahathuduwa, C.N.; Wakefield, S.; West, B.D.; Blume, J.; Dassanayake, T.L.; Weerasinghe, V.S.; Mastergeorge, A. Effects of l-theanine–caffeine combination on sustained attention and inhibitory control among children with ADHD: A proof-of-concept neuroimaging RCT. *Sci. Rep.* **2020**, *10*, 13072. [[CrossRef](#)]
28. Zimmer, L. Positron emission tomography neuroimaging for a better understanding of the biology of ADHD. *Neuropharmacology* **2009**, *57*, 601–607. [[CrossRef](#)]
29. Spencer, T.J.; Biederman, J.; Madras, B.K.; Dougherty, D.D.; Bonab, A.A.; Livni, E.; Meltzer, P.C.; Martin, J.; Rauch, S.; Fischman, A.J. Further evidence of dopamine transporter dysregulation in ADHD: A controlled PET imaging study using altoprane. *Biol. Psychiatry* **2007**, *62*, 1059–1061. [[CrossRef](#)]
30. Cherkasova, M.V.; Faridi, N.; Casey, K.F.; Larcher, K.; O’Driscoll, G.A.; Hechtman, L.; Joobar, R.; Baker, G.B.; Palmer, J.; Evans, A.C.; et al. Differential associations between cortical thickness and striatal dopamine in treatment-naïve adults with ADHD vs. healthy controls. *Front. Hum. Neurosci.* **2017**, *11*, 421. [[CrossRef](#)] [[PubMed](#)]
31. Sigurdardottir, H.L.; Kranz, G.S.; Rami-Mark, C.; James, G.M.; Vanicek, T.; Gryglewski, G.; Kautzky, A.; Hienert, M.; Traub-Weidinger, T.; Mitterhauser, M.; et al. Effects of norepinephrine transporter gene variants on NET binding in ADHD and healthy controls investigated by PET. *Hum. Brain Mapp.* **2016**, *37*, 884–895. [[CrossRef](#)]

32. Kautzky, A.; Vanicek, T.; Philippe, C.; Kranz, G.; Wadsak, W.; Mitterhauser, M.; Hartmann, A.; Hahn, A.; Hacker, M.; Rujescu, D.; et al. Machine learning classification of ADHD and HC by multimodal serotonergic data. *Transl. Psychiatry* **2020**, *10*, 104. [[CrossRef](#)]
33. Grodner, K.; Harcourt, S.; Sattuar, Z.; Strong, A.; Golden, C.; Amen, D.; Willeumier, K.; Taylor, D. B-01 Differentiating Attention Deficit/Hyperactivity Disorder (ADHD) Combined Type and ADHD Inattentive Type Using SPECT Imaging. *Arch. Clin. Neuropsychol.* **2016**, *31*, 613–613. [[CrossRef](#)]
34. Klein, M.; Onnink, M.; van Donkelaar, M.; Wolfers, T.; Harich, B.; Shi, Y.; Dammers, J.; Arias-Vásquez, A.; Hoogman, M.; Franke, B. Brain imaging genetics in ADHD and beyond—mapping pathways from gene to disorder at different levels of complexity. *Neurosci. Biobehav. Rev.* **2017**, *80*, 115–155. [[CrossRef](#)]
35. Karande, S.; Deshmukh, N.; Rangarajan, V.; Agrawal, A.; Sholapurwala, R. Brain SPECT scans in students with specific learning disability: Preliminary results. *J. Postgrad. Med.* **2019**, *65*, 33. [[CrossRef](#)] [[PubMed](#)]
36. Akay, A.P.; Kaya, G.C.; Kose, S.; Yazıcıoğlu, Ç.E.; Erkan, H.Ö.; Güney, S.A.; Oğuz, K.; Keskin, D.; Baykara, B.; Emiroğlu, N.İ.; et al. Genetic imaging study with [Tc-99m] TRODAT-1 SPECT in adolescents with ADHD using OROS-methylphenidate. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2018**, *86*, 294–300. [[CrossRef](#)]
37. Henderson, T.A.; van Lierop, M.J.; McLean, M.; Uszler, J.M.; Thornton, J.F.; Siow, Y.H.; Pavel, D.G.; Cardaci, J.; Cohen, P. Functional Neuroimaging in Psychiatry—Aiding in Diagnosis and Guiding Treatment. What the American Psychiatric Association Does Not Know. *Front. Psychiatry* **2020**, *11*, 276. [[CrossRef](#)]
38. Brown, M.R.; Sidhu, G.S.; Greiner, R.; Asgarian, N.; Bastani, M.; Silverstone, P.H.; Greenshaw, A.J.; Dursun, S.M. ADHD-200 Global Competition: Diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. *Front. Syst. Neurosci.* **2012**, *6*, 69. [[CrossRef](#)] [[PubMed](#)]
39. Hoekzema, E.; Carmona, S.; Ramos-Quiroga, J.A.; Richarte Fernandez, V.; Bosch, R.; Soliva, J.C.; Rovira, M.; Bulbena, A.; Tobeña, A.; Casas, M.; et al. An independent components and functional connectivity analysis of resting state fMRI data points to neural network dysregulation in adult ADHD. *Hum. Brain Mapp.* **2014**, *35*, 1261–1272. [[CrossRef](#)]
40. Riaz, A.; Asad, M.; Alonso, E.; Slabaugh, G. Fusion of fMRI and non-imaging data for ADHD classification. *Comput. Med. Imaging Graph.* **2018**, *65*, 115–128. [[CrossRef](#)] [[PubMed](#)]
41. Rubia, K.; Criaud, M.; Wulff, M.; Alegria, A.; Brinson, H.; Barker, G.; Stahl, D.; Giampietro, V. Functional connectivity changes associated with fMRI neurofeedback of right inferior frontal cortex in adolescents with ADHD. *NeuroImage* **2019**, *188*, 43–58. [[CrossRef](#)] [[PubMed](#)]
42. Zhang, T.; Li, C.; Li, P.; Peng, Y.; Kang, X.; Jiang, C.; Li, F.; Zhu, X.; Yao, D.; Biswal, B.; et al. Separated Channel Attention Convolutional Neural Network (SC-CNN-Attention) to Identify ADHD in Multi-Site Rs-fMRI Dataset. *Entropy* **2020**, *22*, 893. [[CrossRef](#)] [[PubMed](#)]
43. Mirzaei, A.; Pourahmadi, V.; Soltani, M.; Sheikhzadeh, H. Deep feature selection using a teacher-student network. *Neurocomputing* **2020**, *383*, 396–408. [[CrossRef](#)]
44. Chang, C.W.; Ho, C.C.; Chen, J.H. ADHD classification by a texture analysis of anatomical brain MRI data. *Front. Syst. Neurosci.* **2012**, *6*, 66. [[CrossRef](#)]
45. Peng, X.; Lin, P.; Zhang, T.; Wang, J. Extreme learning machine-based classification of ADHD using brain structural MRI data. *PLoS ONE* **2013**, *8*, e79476. [[CrossRef](#)]
46. Fischl, B. FreeSurfer. *NeuroImage* **2012**, *62*, 774–781. [[CrossRef](#)]
47. Lim, L.; Marquand, A.; Cubillo, A.A.; Smith, A.B.; Chantiluke, K.; Simmons, A.; Mehta, M.; Rubia, K. Disorder-specific predictive classification of adolescents with attention deficit hyperactivity disorder (ADHD) relative to autism using structural magnetic resonance imaging. *PLoS ONE* **2013**, *8*, e63660. [[CrossRef](#)]
48. Qureshi, M.N.I.; Oh, J.; Min, B.; Jo, H.J.; Lee, B. Multi-modal, multi-measure, and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain MRI. *Front. Hum. Neurosci.* **2017**, *11*, 157. [[PubMed](#)]
49. Sen, B.; Borle, N.C.; Greiner, R.; Brown, M.R. A general prediction model for the detection of ADHD and Autism using structural and functional MRI. *PLoS ONE* **2018**, *13*, e0194856. [[CrossRef](#)]
50. Jiang, W.; Duan, K.; Rootes-Murdy, K.; Hoekstra, P.J.; Hartman, C.A.; Oosterlaan, J.; Heslenfeld, D.; Franke, B.; Buitelaar, J.; Arias-Vasquez, A.; et al. Structural brain alterations and their association with cognitive function and symptoms in Attention-deficit/Hyperactivity Disorder families. *NeuroImage Clin.* **2020**, *27*, 102273. [[CrossRef](#)]
51. Sato, J.R.; Hoexter, M.Q.; Castellanos, X.F.; Rohde, L.A. Abnormal brain connectivity patterns in adults with ADHD: A coherence study. *PLoS ONE* **2012**, *7*, e45671.
52. Choi, J.; Jeong, B.; Lee, S.W.; Go, H.J. Aberrant development of functional connectivity among resting state-related functional networks in medication-naïve ADHD children. *PLoS ONE* **2013**, *8*, e83516. [[CrossRef](#)] [[PubMed](#)]
53. McLeod, K.R.; Langevin, L.M.; Goodyear, B.G.; Dewey, D. Functional connectivity of neural motor networks is disrupted in children with developmental coordination disorder and attention-deficit/hyperactivity disorder. *NeuroImage Clin.* **2014**, *4*, 566–575. [[CrossRef](#)]
54. Deshpande, G.; Wang, P.; Rangaprakash, D.; Wilamowski, B. Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE Trans. Cybern.* **2015**, *45*, 2668–2679. [[CrossRef](#)] [[PubMed](#)]

55. Mowinckel, A.M.; Alnæs, D.; Pedersen, M.L.; Ziegler, S.; Fredriksen, M.; Kaufmann, T.; Sonuga-Barke, E.; Endestad, T.; Westlye, L.T.; Biele, G. Increased default-mode variability is related to reduced task-performance and is evident in adults with ADHD. *NeuroImage Clin.* **2017**, *16*, 369–382. [[CrossRef](#)] [[PubMed](#)]
56. Damiani, S.; Tarchi, L.; Scalabrini, A.; Marini, S.; Provenzani, U.; Rocchetti, M.; Oliva, F.; Politi, P. Beneath the surface: Hyperconnectivity between caudate and salience regions in ADHD fMRI at rest. *Eur. Child Adolesc. Psychiatry* **2020**, *4*, 619–631. [[CrossRef](#)] [[PubMed](#)]
57. Guo, X.; Yao, D.; Cao, Q.; Liu, L.; Zhao, Q.; Li, H.; Huang, F.; Wang, Y.; Qian, Q.; Wang, Y.; et al. Shared and distinct resting functional connectivity in children and adults with attention-deficit/hyperactivity disorder. *Transl. Psychiatry* **2020**, *10*, 65. [[CrossRef](#)] [[PubMed](#)]
58. Shao, L.; You, Y.; Du, H.; Fu, D. Classification of ADHD with fMRI data and multi-objective optimization. *Comput. Methods Prog. Biomed.* **2020**, 105676. [[CrossRef](#)]
59. Chen, Y.; Tang, Y.; Wang, C.; Liu, X.; Zhao, L.; Wang, Z. ADHD classification by dual subspace learning using resting-state functional connectivity. *Artif. Intell. Med.* **2020**, *103*, 101786. [[CrossRef](#)]
60. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
61. Gheisari, M.; Najafabadi, H.E.; Alzubi, J.A.; Gao, J.; Wang, G.; Abbasi, A.A.; Castiglione, A. OBPP: An ontology-based framework for privacy-preserving in IoT-based smart city. *Future Gener. Comput. Syst.* **2021**, *123*, 1–13. [[CrossRef](#)]
62. Gao, J.; Wang, H.; Shen, H. Machine learning based workload prediction in cloud computing. In Proceedings of the 2020 29th International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 3–6 August 2020; pp. 1–9.
63. Gao, J.; Wang, H.; Shen, H. Smartly handling renewable energy instability in supporting a cloud datacenter. In Proceedings of the 2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS), New Orleans, LA, USA, 18–22 May 2020; pp. 769–778.
64. Gao, J.; Wang, H.; Shen, H. Task failure prediction in cloud data centers using deep learning. *IEEE Trans. Serv. Comput.* **2020**. [[CrossRef](#)]
65. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)] [[PubMed](#)]
66. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
67. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [[CrossRef](#)]
68. Mao, Z.; Su, Y.; Xu, G.; Wang, X.; Huang, Y.; Yue, W.; Sun, L.; Xiong, N. Spatio-temporal deep learning method for ADHD fMRI classification. *Inf. Sci.* **2019**, *499*, 1–11. [[CrossRef](#)]
69. Riaz, A.; Asad, M.; Alonso, E.; Slabaugh, G. DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI. *J. Neurosci. Methods* **2020**, *335*, 108506. [[CrossRef](#)] [[PubMed](#)]
70. Liu, S.; Zhao, L.; Wang, X.; Xin, Q.; Zhao, J.; Guttery, D.S.; Zhang, Y.D. Deep Spatio-Temporal Representation and Ensemble Classification for Attention deficit/Hyperactivity disorder. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *29*, 1–10. [[CrossRef](#)] [[PubMed](#)]
71. Bellec, P.; Chu, C.; Chouinard-Decorte, F.; Benhajali, Y.; Margulies, D.S.; Craddock, R.C. The neuro bureau ADHD-200 preprocessed repository. *NeuroImage* **2017**, *144*, 275–286. [[CrossRef](#)]
72. Lavoie-Courchesne, S.; Rioux, P.; Chouinard-Decorte, F.; Sherif, T.; Rousseau, M.E.; Das, S.; Adalat, R.; Doyon, J.; Craddock, C.; Margulies, D.; et al. Integration of a neuroimaging processing pipeline into a pan-canadian computing grid. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2012; Volume 341, p. 012032.
73. Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; Joliot, M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **2002**, *15*, 273–289. [[CrossRef](#)]
74. Bellec, P.; Perlberg, V.; Jbabdi, S.; Péligrini-Issac, M.; Anton, J.L.; Doyon, J.; Benali, H. Identification of large-scale networks in the brain using fMRI. *NeuroImage* **2006**, *29*, 1231–1243. [[CrossRef](#)] [[PubMed](#)]
75. Frey, B.J.; Dueck, D. Clustering by passing messages between data points. *Science* **2007**, *315*, 972–976. [[CrossRef](#)] [[PubMed](#)]
76. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
77. Rajpoot, K.; Riaz, A.; Majeed, W.; Rajpoot, N. Functional connectivity alterations in epilepsy from resting-state functional MRI. *PLoS ONE* **2015**, *10*, e0134944. [[CrossRef](#)] [[PubMed](#)]
78. Yu, H.; Mu, C.; Sun, C.; Yang, W.; Yang, X.; Zuo, X. Support vector machine-based optimized decision threshold adjustment strategy for classifying imbalanced data. *Knowl. Based Syst.* **2015**, *76*, 67–78. [[CrossRef](#)]
79. Zhang, Y.; Fu, P.; Liu, W.; Chen, G. Imbalanced data classification based on scaling kernel-based support vector machine. *Neural Comput. Appl.* **2014**, *25*, 927–935. [[CrossRef](#)]
80. Dey, S.; Rao, A.R.; Shah, M. Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects. *Front. Neural Circ.* **2014**, *8*, 64. [[CrossRef](#)]

81. Nunez-Garcia, M.; Simpraga, S.; Jurado, M.A.; Garolera, M.; Pueyo, R.; Igual, L. FADR: Functional-anatomical discriminative regions for rest fMRI characterization. In *International Workshop on Machine Learning in Medical Imaging*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 61–68.
82. Hao, A.J.; He, B.L.; Yin, C.H. Discrimination of ADHD children based on Deep Bayesian Network. In Proceedings of the 2015 IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015), Beijing, China, 19 November 2015; pp. 1–6.
83. Bondy, J.A.; Murty, U.S.R. *Graph Theory with Applications*; Macmillan: London, UK, 1976; Volume 290.
84. Xia, M.; Wang, J.; He, Y. BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS ONE* **2013**, *8*, e68910. [[CrossRef](#)] [[PubMed](#)]
85. Rolls, E.T.; Cheng, W.; Feng, J. Brain dynamics: The temporal variability of connectivity, and differences in schizophrenia and ADHD. *Transl. Psychiatry* **2021**, *11*, 70. [[CrossRef](#)]
86. Jie, B.; Wee, C.Y.; Shen, D.; Zhang, D. Hyper-connectivity of functional networks for brain disease diagnosis. *Med. Image Anal.* **2016**, *32*, 84–100. [[CrossRef](#)] [[PubMed](#)]
87. Kyeong, S.; Park, S.; Cheon, K.A.; Kim, J.J.; Song, D.H.; Kim, E. A new approach to investigate the association between brain functional connectivity and disease characteristics of attention-deficit/hyperactivity disorder: Topological neuroimaging data analysis. *PLoS ONE* **2015**, *10*, e0137296. [[CrossRef](#)]
88. Liston, C.; Cohen, M.M.; Teslovich, T.; Levenson, D.; Casey, B. Atypical prefrontal connectivity in attention-deficit/hyperactivity disorder: Pathway to disease or pathological end point? *Biol. Psychiatry* **2011**, *69*, 1168–1177. [[CrossRef](#)] [[PubMed](#)]
89. Kumar, U.; Arya, A.; Agarwal, V. Neural network connectivity in ADHD children: An independent component and functional connectivity analysis of resting state fMRI data. *Brain Imaging Behav.* **2021**, *15*, 157–165. [[CrossRef](#)] [[PubMed](#)]
90. Aradhya, A.M.; Subbaraju, V.; Sundaram, S.; Sundararajan, N. Discriminant Spatial Filtering Method (DSFM) for the identification and analysis of abnormal resting state brain activities. *Expert Syst. Appl.* **2021**, *181*, 115074. [[CrossRef](#)]