

Old Dominion University  
**ODU Digital Commons**

---

Computer Science Faculty Publications

Computer Science

---

2019

## A Survey of Attention Deficit Hyperactivity Disorder Identification Using Psychophysiological Data

S. De Silva

S. Dayarathna

G. Ariyaratne

D. Meedeniya

Sampath Jayarathna  
*Old Dominion University*

Follow this and additional works at: [https://digitalcommons.odu.edu/computerscience\\_fac\\_pubs](https://digitalcommons.odu.edu/computerscience_fac_pubs)



Part of the [Biomedical Engineering and Bioengineering Commons](#), [Child Psychology Commons](#), and the [Computer Sciences Commons](#)

---

### Original Publication Citation

De Silva, S., Dayarathna, S., Ariyaratne, G., Meedeniya, D., & Jayarathna, S. (2019). A survey of Attention Deficit Hyperactivity Disorder identification using psychophysiological data. *International Journal of Online and Biomedical Engineering*, 15(13), 61-76. doi:10.3991/ijoe.v15i13.10744

This Article is brought to you for free and open access by the Computer Science at ODU Digital Commons. It has been accepted for inclusion in Computer Science Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact [digitalcommons@odu.edu](mailto:digitalcommons@odu.edu).

# A Survey of Attention Deficit Hyperactivity Disorder Identification Using Psychophysiological Data

<https://doi.org/10.3991/ijoe.v15i13.10744>

S. De Silva, S. Dayarathna, G. Ariyaratne, D. Meedeniya <sup>(✉)</sup>  
University of Moratuwa, Colombo, Sri Lanka  
dulanim@cse.mrt.ac.lk

S. Jayarathna  
Old Dominion University, Virginia, USA

**Abstract**—Attention Deficit Hyperactivity Disorder (ADHD) is one of the most common neurological disorders among children, that affects different areas in the brain that allows executing certain functionalities. This may lead to a variety of impairments such as difficulties in paying attention or focusing, controlling impulsive behaviours and overreacting. The continuous symptoms may have a severe impact in the long-term. This paper explores the ADHD identification studies using eye movement data and functional Magnetic Resonance Imaging (fMRI). This study discusses different machine learning techniques, existing models and analyses the existing literature. We have identified the current challenges and possible future directions to provide computational support for early identification of ADHD patients that enable early treatments.

**Keywords**—Attention Deficit Hyperactivity Disorder (ADHD), eye movements, fMRI, decision support system, comparative study

## 1 Introduction

Attention deficit Hyperactivity disorder (ADHD) is a common psychiatric disorder with a genetic component. This can be categorized with core symptoms of inattention, impulsivity and hyperactivity [1]. ADHD is known as under-diagnosed as there is no exact diagnosis method. Hence, a high ratio of children with ADHD is continued to develop the symptoms throughout their lives. Thus, early diagnosis minimizes the long-term impact and help to develop intellectual capabilities. At present, 5-11% of children affected from ADHD in the United States representing 6.4 million children in countrywide [2] and the patient count has increased by 43% from 2003 to 2016 [3].

Moreover, ADHD has the possibility of comorbid with other neuropsychiatric disorders. Therefore, early detection of ADHD is important to prevent relevant children from future symptoms and difficulties in executive functions such as planning, organizing, initiating activities, and monitoring.

Many studies have been conducted to diagnose ADHD using clinical data such as EEG and fMRI. Most of them have shown a significant improvement in classification

between ADHD and other control subjects [4]. The major aims of this study are to identify the possible psychophysiological measurements to recognize ADHD and explore different Machine Learning (ML) techniques to classify possible features and patterns in given measurements. This survey appraises the recent work between 2012 and 2017 which were able to obtain significantly better results using different techniques. Moreover, we identify the limitations of existing studies and suggest future research directions, enabling better decisions on ADHD diagnosis.

Section 2 gives an overview of ADHD and Section 3 states the clinical practices. Section 4 and Section 5 explore data pre-processing and classification techniques, respectively. Section 6 describes related models and Section 7 states the evaluation techniques. Section 8 presents the current limitation and Section 9 concludes the paper.

## **2 Overview of ADHD**

### **2.1 What is ADHD**

Attention Deficit Hyperactivity Disorder (ADHD) is defined as a consistent pattern of inattention, hyperactivity and impulsivity, which is at a higher rate compared to other control groups in the level of development [5]. ADHD is a controversial and contentious childhood disorder that can be prevailing into adulthood, which is perilous. Although, the common symptoms show learning difficulties and attention disorders, ADHD can be led to talented individuals, where the symptoms of hyperactivity lie in the fields of curiosity for knowledge and creativity.

Statistics on ADHD have shown that more males are affected than females with a ratio of 9:4. The gender-biased behaviour can be justified by nature, since boys tend to be hyperactive or inattentive; hence the brain areas are activated in developing extreme levels resulting in ADHD [6]. Studies have suggested that brain abnormalities can be found in the brain regions of frontal lobes, basal ganglia and cerebellum. The frontal lobes control the basic human functions, decision making and social behaviours. Thus, the diagnosis of ADHD indicates that these areas are mandatory, when fMRI is being analysed. The basal ganglia control the motor behaviour and the abnormalities in basal ganglia and cerebellum cause hyperactivity. In contrast, the inattention symptoms are associated with impairments in the prefrontal and frontal cortex [7].

Moreover, ADHD has a significant chance of comorbidity with other disorders such as depression, anxiety and learning disability and the associated impacts may worsen in the long-term. Thus, early diagnosis of ADHD early is important to start treatments to control the impact, so that adolescence will be able to handle the social interactions, control hyperactivity and become nearly a normal person. However, there is no single test in practice to diagnose ADHD, and many studies use clinical data such as EEG and fMRI. Most of them have shown an improvement in classification between ADHD and other control subjects [4]. However, the machine learning

approaches have failed to classify among the subtypes of ADHD, which is important as the treatments differ from subtype to subtype.

### **3 Current Practices of Clinical Diagnosis Measurements**

#### **3.1 Behaviour analysis**

ADHD identification in clinical practice is mainly focused on symptom monitoring and evaluation. The behavioural analysis often conducted as questionnaires containing 18-90 general questions [1]. The symptoms are categorized according to the disorder types of inattention, hyperactivity and impulsivity. Several diagnosis rating scales are used in practice [8]. Swanson, Nolan and Pelham Questionnaire 4th Edition (SNAP-IV) is a rating scale based on Diagnostic and Statistical Manual of Mental Disorders 4th Edition (DSM-IV), that has an internal consistency across the coverage of ADHD. However, it lacks the inter-rater agreement between the subjects that is the interviewer and the participant, sparsity, normality and the quality of the psychometric properties. Thus, the validity of this scale is doubtful and not widely used in clinical practice [8], as the age and gender-specific details are not considered as statistical methods.

Multi-informant-based scale for cognitive and clinical data is another rating scale. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V) diagnosis criteria has at least two multi-informants, including parents and teachers of the child. This is tested using structural equation modelling and logical algorithms [9], which is a well-defined method in clinical practice. However, this may bias on the participant, the participant count and lead to random measurement error, that causes inaccurate results. Another method of combining the different ratings of several participants is the use of probabilistic functions of severity, such as structural equation modelling, that calculates the average symptoms rate. Thus, it overcomes the partiality issues.

Moreover, objective scores are used to determine the severity of ADHD. Continuous Performance Test (CPT), and Stroop test are such objective measures. They conduct based on the visual performance on attention abilities of the participant to diagnose ADHD. Generally, a patient is asked to perform a task and measured the response time and correctness. Cognitive flexibility is measured using Shifting attention Test, where many faults and longer reaction time shown by ADHD subjects due to their inefficiency in complex tasks [10]. Many errors within shorter response time indicate impulsive responses and slower reaction time indicates the inefficiency in cognition [10].

#### **3.2 Psychophysiological measures**

**Eye movement Data:** The eye movement records in neurological disorders provide rich data set with simple vasomotor baseline and indicate a complex behavioural process. Since it does not need any advanced cognitive skills, this can be easily performed for children with ADHD disorder. Eye movements can be classified as hori-

zontal, vertical and torsional. The eye rotations are mainly achieved by contraction and relaxation of six extraocular muscles [11]. The gaze types including fixations and saccades are used to obtain the cognitive inferences. The fixations are the series of gaze points when the attention is fixed into one point. It is a processed output of a series of gaze points based on fixation duration, apart from the spatial and temporal components. Saccades are the rapid movements of fovea between two points of interest [12]. The results of eye movement data reveal that ADHD subjects take longer reaction time and identical variability compared to control subjects. Considering the ADHD subjects, there are different eye movement patterns among adults and fixations are difficult to maintain for children. Generally, the gaze points-based eye movement data are mapped to 2D coordinates.

**fMRI Data:** fMRI is an efficient technique to measure brain activity by detecting the changes of blood oxygen level in the brain which occurs as responses to neural activities. Structural measures such as MRI uses a magnetic field to capture the images of tissues, are not capable of revealing sensitive physiological changes occur in a short time with related to brain activations. Whereas, fMRI measures overcome this by measuring the changes in blood oxygen level and identifying different patterns of brain activation that occurs as a response to mental processes [13]. However, there can be a higher frequency of voxels in certain areas in fMRI, which are different from the actual scanning [14].

**EEG Data:** Electroencephalography (EEG) is a common physiological method of recording electrical activities generated in the brain by placing a scalp surface with mounted electrodes. EEG signals are nonlinear, nonstationary and noisy with a high sampling rate to detect brain wave patterns in cortical areas efficiently [15]. EEG is used to identify the cortex areas that are processing at a given time. Different areas are responsible for a specific task such as visual stimulus, motor functions and language processing. Moreover, these frequency patterns are used to identify the memory encoding, depth of sleep, relaxed states and motor regions. Hence, EEG can be used to identify abnormalities in the nervous system properties of neurological disorders.

## 4 Neuroimaging Data Processing and Learning Models

### 4.1 Neuroimaging data pre-processing

**Neuroimaging Correction:** Motion correction for fMRI data is performed by mapping image slices to anatomical volume derive from the same fMRI session to assist the inter-slice head motion. Head motion is an issue that affects fMRI data as a slight movement of the subject may cause a greater variation in the activation responses and missing edges of the recorded data. The reason is the fMRI images obtained by the scanner at spatial locations without keeping a link with the position of the brain [16]. Thus, extra slices of edges in a given area are taken to correct through-plane movements. Motion correction can adjust the brain position in the images by aligning image volumes spatially using a single volume as a reference, which is re-

ferred to as co-registration. Several methods are used to identify the parameters of the best reference volume [17].

Slice time correction for fMRI data is performed by shifting the signal phase to temporally aligned data. This is essential for fMRI data to have an exact timing in stimulus presentation to accurately analyses its time course. fMRI data acquired as a sequence of 2D imaging that causes to have an offset between each slice. The complete volume of data is acquired within repetition times, causing delays between each slice. Thus, time slice correction reduces the time differences between each slice during pre-processing. This realigns each slide based on a reference slice. Different interpolation methods are used in slice time correction. Linear interpolating is an efficient method that applies a phase shift in frequency using Fast Fourier transformation [16].

Geometric or intensive distortion corrects the inhomogeneity field distortion of fMRI data by adjusting different order magnetic field gradients, which are created by shimming coils. This mainly caused by spatial warping with field inhomogeneity which prevents fMRI data from matching with structural images accurately, due to the non-uniformity of static fields. This may cause signal losses or variations in intensities of images. These inhomogeneity field distortions can be corrected using distortion correction methods. The noise and smoothness of signal properties can also be estimated to improve the image uniformity [16].

**Noise Removal:** The accuracy and spatial resolution of fMRI data can be improved by removing the associated cardiac and respiration-induced physiological noise. Generally, noise removal is done during acquisition or synchronization in the post-processing. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are two different approaches to remove noise in fMRI data. Both approaches decompose fMRI data into multiple components and each component shares the same frequency bands in feature projection. The frequency overlapping signal and noise can be distinguished as the feature projections are performed in different domain of frequency. Kernel Principal Component Analysis (KPCA) is another nonlinear extension of PCA for noise removal that categorizes high-order dependence among voxels by providing more features than linear PCA. KPCA removes the Gaussian by removing the least significant component in reconstruction [18]. Moreover, canonical correlation is used to reduce the fMRI noise, by classifying the noise into structured and unstructured data. Then the structured noise is removed from time series of all voxels and the unstructured noise and temporal correlations were reduced by randomization techniques.

**Normalization:** Normalization performs the inter subject averaging by increasing the possible degree of freedom allowed for a statistical model and activation signal above the obtained subject. The size and the shape of the brain in the fMRI experiment may differ from 30% for each subject. In spatial normalization, the image volumes are compared to identify differences between them. Then the identified differences in shapes are reduced by stretching, warping and squeezing images using mathematical approaches. fMRI data normalization compares the activity areas of different subjects by representing data in a common scheme or mapping system that uses 3D coordinates. The techniques such as Talairach space, probabilistic space are used for the normalization [16].

**Smoothing:** Smoothing averages the signals from adjacent voxels which improves the signal to noise ratio (SNR) and cause to decrease the spatial resolution. Raw SNR that depends on voxel sizes of images and the time spent on data collection from the voxels, is high in fMRI data. It is vital to apply smoothing methods and identify slight changes within the signals to improve the SNR of each voxel. This causes resolution distance to be increased in proportion to the area reduction with the windowing function and reducing the spatial resolution. The correct activations can detect by random noise removal [16].

**Image Registration:** Image registration transforms an image such that its structure is well-aligned to the homologous features in a second image [19]. This can be applied to both 2D and 3D image sequences by enabling the fusion of imaging techniques, which are acquired in different angles at different times. These methods are also used to segment bio structures, extract information and build valid models. This includes both image matching and image interpolation to detect the feature similarities among the related images and to apply transformations [19]. The main types are point method, edge method, moment method and similarity criterion optimization method [20]. Geometric-based registration is related to a feature or intensity-based methods. Feature-based methods align the similarity between images by computing the geometrical transformation. Intensity-based methods consider the image pixel intensities and repeatedly adjust the transformation by minimizing the cost function. Landmark-based registration identifies the similarities between image landmarks. For optimal results, pre-processing methods are required to perform prior to image registration.

## 4.2 Feature extraction

Feature extraction produces new features from the existing features as a method of dimensionality reduction, that transforms complex, high dimensional data into a lower dimension. The common python libraries Numpy[21] and Pandas [22] are used for dimension reduction. Numpy provides multidimensional array objects with high-level mathematical functions for their operation. Pandas provide data structures and data manipulation functions of time series and numeric tables.

**Principal Component Analysis (PCA):** is an unsupervised linear algorithm for feature extraction, thermal noise and feature dimensionality reduction and classification of fMRI data by identifying the Eigenvalues and Eigenvectors of covariance data matrices [23]. fMRI data involved in time sampling functions with noise, where smooth functions are used to replace voxels. It performs dimension reduction to summarize the features, as high dimensional data are complex to analysis, requires high computational cost and may increase the error rate. This can be applied for the linear structure of data to define outliers. Thus, PCA does not apply for highly correlated data, as the principal components are uncorrelated with the goal of increasing the variance.

**Independent Component Analysis (ICA):** is a computational and statistical feature extraction technique to expose hidden factors in signals and random variables. ICA decomposes complex datasets into independent subparts, using the intrinsic spatio temporal structure of the data without having prior assumptions. ICA identifies the

connectivity of the brain as a data-driven, multivariate method. Mutually independent and non-Gaussian latent variables are the independent components of the data [24].

**Fast Fourier transform (FFT):** is a feature selection and re-sampling technique for raw data. This identifies the similarity and patterns among two electrical load datasets. It is used to convert a sequence from its original domain to the frequency domain. This algorithm computes the Discrete Fourier transform (DFT) of the signal. The algorithm decomposes the original signal of  $N$  number of points, into  $N$  number signals with a single point. Then the spectrum analysis of each of the demonised signals is calculated and ultimately the  $N$  spectra are synthesised into a single frequency signal.

Several studies have used FFT for the feature extraction of EEG signals and fMRI data. Kuang et.al [14] have used FFT to eliminate the error by converting the BOLD signals obtained from fMRI data from the time domain to the frequency domain. This has enabled complex features extraction such as the max amplitude of the frequency domain for each voxel as the frequency is higher when the voxel is active. Further, FFT with max pooling have supported with only one feature per pixel and obtained a cluster of properties by combining a set of voxels [4][25].

**Symlets 5 Wavelet:** Wavelets techniques are used to split data into subcomponents of frequencies, which are then processed by matching resolutions to its scale. These short waves with a limited duration are scaled based on the frequency and there are many types of sub techniques. Symlet 5 wavelets (Sym5) are a modified version of Daubechies wavelets (dbN), which defines a discrete wavelet transform and considers the maximal number of disappearing instances, with higher symmetry. It is widely applied function in the orthogonal wavelet families. In contrast to Fourier analysis, in wavelets user can select the transformation points. Sym5 family has the features of arbitrary granularity, orthogonal with compact support, near symmetry, an arbitrary number of zero moments. It supports orthogonal analysis, biorthogonal analysis, exact reconstruction, FIR filters, continuous transformation, discrete transformation while being a fast algorithm [26]. Apart from FFT, Sym5 is a major feature extraction technique in signal processing. Kuang [25] et.al have used the Sym5 which is the Symlet of order 5 as the feature extraction method apart from FFT, Coif3 and was able to an equivalent accuracy with FFT.

### 4.3 Neuroimaging data machine learning models

**Support Vector Machines (SVM):** is a classifier for both linear and nonlinear data, that uses nonlinear mapping for the transformation of input data into another higher dimension. The data are separated into classes by finding an optimal decision boundary among the new dimension space. The appropriate hyperplane to distinguish two classes are found by using support vectors and their margins defined. When there are many sets of hyperplanes for a given dataset, the maximum distance between two data points is considered [25]. Since, SVM supports two-class classification, a binary classification can be multiplied to address multi-classification for the classification of ADHD subtypes of hyperactivity, inattentive, combined and control groups.



**Convolutional Neural Networks (CNN):** is a deep learning model to classify data types such as audio, video and images, even with small training sets. In image processing, CNN's controlled input of two-dimensional neuron array is similar to pixels of the image. The output layer is one dimensional neuron set. Also, the CNN contain sparsely connected convolutional layers, pooling layers and fully connected layers. Convolution layers process the image and the pooling layers reduce the number of neurons. Then the fully connected layers connect pooling layers to output layers [27].

**Fully connected deep networks (FNN):** is a deep learning technique consists of a series of fully connected layers, where each neuron in a layer is connected to every other neuron in the previous layer and each with its own weight [28]. The output from the neural network is a combination of the previous inputs and their weights in each layer. A nonlinear function combines the weights and the inputs, that enables to stack the fully connected network. FNN tends to memorize the training data, hence the training loss converges to zero. However, when new data are added, the model tries to formulate the differences and gets out of the memorized data.

**Extreme Learning Machine (ELM):** Several ADHD classifications studies have used the Extreme Learning Machine (ELM) algorithm on structural MRI data to provide an objective clinical diagnosis. Peng et al., [29] have shown ELM learning algorithm classifies data with high accuracy and performance, compared to Support Vector Machine (SVM). ELM is used in both classification and regression problems, due to the fewer optimization constraints and simpler implementation.

**Deep Belief Network:** Deep Belief Network (DBN) is a deep learning method with a stack of Restricted Boltzmann Machines (RBM) that uses for pre-training and a feed-forward network for fine-tuning [30]. The DBN consists of hidden layers and a visible layer at the start of the network with the observation vectors as the raw data. The hidden layers relate to the weights generated on the raw data [25]. Although, the hidden layers are linked, there is no interconnectivity within the layer where the pre-training models are present [30]. The feature extraction with RBM layer enhances the higher-level features learning in an unsupervised manner [14].

#### 4.4 Comparison of related techniques

Table 1 states a comparison of related studies in terms of the considered techniques.

**Table 1.** Comparison of related techniques

Related study	Pre-processing								Classification							
	<i>Normalization</i>	<i>Motion Correction</i>	<i>Slice time correction</i>	<i>Noise Removal</i>	<i>Smoothing</i>	<i>Dimensionality Correction</i>	<i>Co-registration</i>	<i>Max-pooling</i>	<i>Forward selection</i>	<i>SVM</i>	<i>CNN</i>	<i>FNN</i>	<i>DBN</i>	<i>DB</i>	<i>ELM</i>	<i>NN</i>
[4]						X	X			X			X	X		
[14]						X							X			X
[24]					X				X	X						
[25]	X		X		X		X						X			
[27]	X	X	X		X		X				X					
[28]	X	X			X					X		X				
[29]	X						X			X					X	
[31]				X						X						

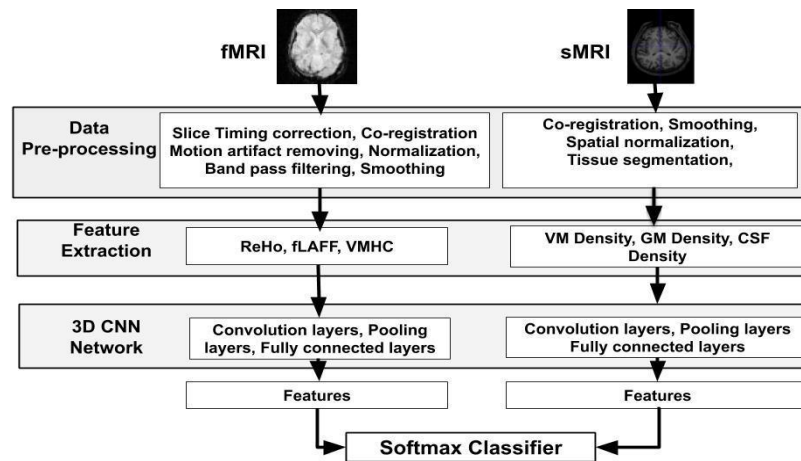
Several pre-processing techniques have used to process psychophysiological data. For instance, the main neuroimaging corrections methods for fMRI data are slice time correction, motion correction, normalization and co-registration. PCA and ICA have been used for the feature extraction process of fMRI data. The pre-processing of EEG data includes artefact-based neuroimaging correction and uses high-pass filters for noise removal in ASDGenus approach [32]. Among the widely used techniques, forwards selection is used to reduce dimensionality and FFT, Sym5 have used to extract features. Considering the classification learning models, SVM has been widely used for ADHD classification. However, as an emerging technology, DM techniques can be used as a future research direction, to increase the classification accuracy and result optimization.

## 5 Related Neuroimaging Classification Studies

Neuroscience Decision Support Systems are computerized systems that are used to solve complex problems and decision making based on neuroscience measurements. These decision support systems are created on top of the existing feature selection, extraction and classifiers implementations. The appraised models were selected based on the latest studies and their accuracy levels, usability, scalability etc.

An ADHD classification model based on structural and functional MRI using the 3D Convolutional Neural Networks (CNN) is presented by Zou [27]. As shown in

Figure 1, they have used separate data pre-processing and feature extraction techniques for fMRI and sMRI data. The CNNs and softmax classifier have trained to classify ADHD subjects. They have shown the usage of 3D CNN with SVM has performed well in identifying 3D brain images. However, this approach has low-level features as vector and potentially neglects the local patterns in the 3D images. Moreover, this model outperforms over other approaches such as deep belief networks and multi-kernel learning, with fewer training samples [27].



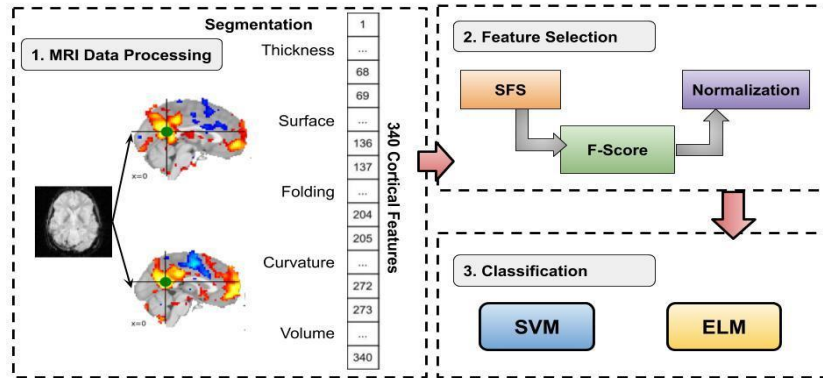
**Fig. 1.** ADHD classification flow chart based on fMRI and sMRI using 3D CNN [27]

Extreme Learning Machine (ELM) based classification of ADHD using brain structural MRI data, is addressed by Peng [29]. As the pre-processing technique, motion correction has applied to MRI data and 340 features have divided into sets of 68 components based on brain segments as shown in Figure 2.

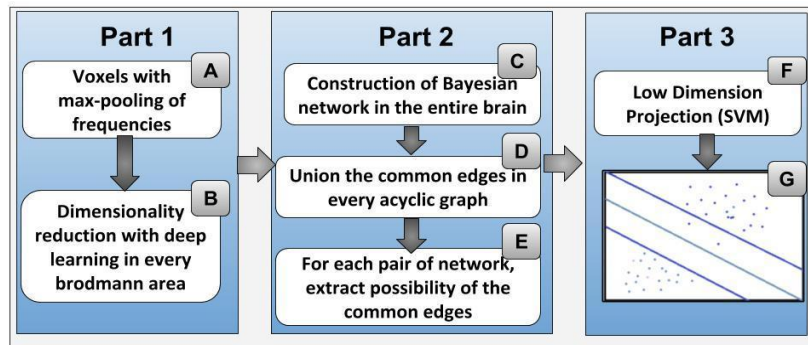
Sequential Forward Selection (SFS) and F-Score were applied for feature selection to achieve high classification accuracy, after the normalization. Then the ELM and SVM were used to perform the leave one out cross validation for the dataset. However, they have considered only the structural MRI data from the ADHD-200 Global Competition.

Figure 3 illustrates an ADHD classification framework based on Deep Belief Network (DBN) [4], for three datasets from ADHD 200 global competition. This approach is used to determine the cognitive aspect of the brain by analysing fMRI data to examine the relationship between different areas of the brain. Initially, they have pre-processed the voxels with max pooling frequencies.

Then DBN is used to reduce dimensionality in every Brodmann area and Bayesian network extracts the features with normalized fMRI data. Finally, SVM is used to classify ADHD subjects. However, this method is only applicable to specific brain areas.



**Fig. 2.** ADHD classification workflow with MRI human cortical feature measures [29]



**Fig. 3.** The procedure of Deep Bayesian Network [4]

Bag of Words (BoW) approach [31], is another ADHD classification method based on SVM, that uses fMRI data of resting state brain to compute the correlation between voxel pairs for a given area. It is a Natural Language Processing dictionary-based model for biomedical imaging. The dataset is from ADHD-200 competition. The network of voxels has developed by connecting each voxel by an edge, that denotes a high correlation value between two voxels. Features have extracted based on the network features such as degree of voxels and intensity values of the time series in each voxel. The classification accuracy has improved by using a combination of network features and raw intensity time series, without considering separate features. This approach can be combined with different feature types to analyse brain disorders. As shown in Figure 4,  $N \times N$  correlation metrics were computed using 4D fMRI data and then the adjacency matrix was computed by thresholding high correlation values, which were taken between each couple of voxel time series. BoW codebook generation was done using the features like node degree and cycle. Finally, the SVM has used for the classification.

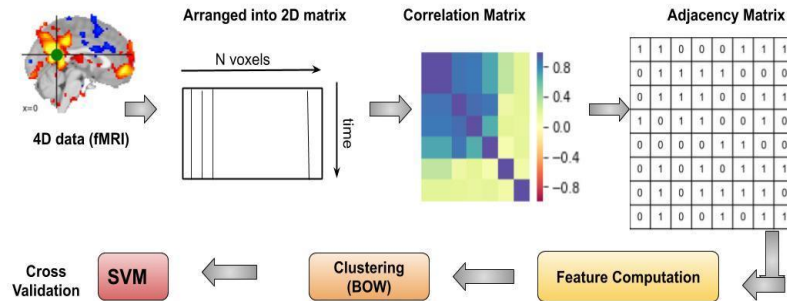


Fig. 4. Overview of the Bag of Word approach [31]

Several classification techniques such as SVM, decision tree, ANN, Bayesian networks, ensemble classifiers have reviewed for pigmented skin lesion classification with dermoscopic images. For example, an SVM based classification has been used to classify whether a skin lesion is a melanoma or not, by verifying the influence of using various features to improve the classification performance. Further, Decision Tree classification has been used in skin lesion classification because of its' ease of rule generation and simplicity of visualization and understanding. SVM has also been applied due to its capability of clarifying nonlinear data in a simple manner and ANN for handling the complexity of pattern recognition [33].

## 6 Evaluation Techniques

Several evaluation methods have been used to evaluate the classification accuracy. The Holdout method divides the original data set into two sets as testing and training. Then the classifier builds on training data set and evaluates it on testing data set which was held out. This method can be improved by repeating the same process by selecting a random set of data for training and testing in every turn [14][27]. K-fold cross-validation is another evaluation method. It partitions the data set into equal sizes of k subsets and each subset is used as a test set and the remaining set is used as a training data. The accuracy of each turn with different testing sets, folds, are averaged to prevent the overlapping of test sets and to increase efficiency [29]. Leave one out cross-validation is a special case of the k-fold validation method. It uses a test set of size one and every subject is used as a test subject in turn. This also increases the efficiency of the dataset by avoiding the random selection issues in the datasets. Further, generalized error estimation is considered to average the results for the available training points.

## 7 Discussion

Table 2 summarises some of the related work with their pre-processing, classification techniques, accuracy levels and limitations. The classification accuracy of the

model is mainly based on the extracted features from input data for the discrimination between different classes. Although, several methods were used for both feature extraction and classification, it is challengeable to extract common features, which is the best fit for the classification process.

**Table 2.** Comparison of the existing literature

Related work	Performance status	Limitations
[4]	Improved classification performance using DBN, Accurate relationships of Brodmann brain areas, Use feature extraction of the links between well-performed brain areas.	Applicable only for specific brain areas.
[25]	High prediction accuracy for ADHD subtypes, Avoid the impact of imbalance dataset. Accuracy of 44.63% in Coif3, that is 2.43% higher than the balanced data set.	Perform well with only large datasets.
[27]	Accuracy of 69.15%. High performance, applicable to small training samples.	An average accuracy due to the large sample size and heterogeneity.
[28]	The overall accuracy of 90% and 95% for classifying subtypes. FCC, ANN has well performed compared to SVM.	Applicable for adult data.
[29]	Prediction accuracy of 90.18% for ELM combining 11 features, 84.73% for SVM-Linear and 86.55% for SVM-RBF. High classification performance with robustness with a change in the small training set and fast changing speed.	Applicable only for structural MRI

The main limitation in related studies is the use of one data type for the classification. A combination of more than one measurement can improve ADHD identification accuracy as the feature extraction has a direct impact on the classification process. Feature selection methods also affect the classification accuracy. However, many studies have used only one key algorithm for the feature selection which reduces the performance. Research can be conducted with a combination of several techniques to improve classification accuracy. Most of the fMRI-based studies were implemented only for a given area of the brain and extensions are required to apply the same method to other brain areas. Thus, the use of single dataset and single classifier have limited the identification process into a one subject type. This study has considered ADHD diagnosis using EEG, resting state fMRI and eye movement data. Neuropsychological measures such as MRI, MEG, PET can be further used as imaging techniques.

Moreover, there is no single method to identify subtypes of ADHD including impulsivity, inattention, and hyperactivity. The rating scales of behavioural analysis such as manual rating scales, objective measures and, anatomical rating scales have their own measurements of providing a numeric value for ADHD diagnosis. However, the current practice lacks a well-defined objective measure or a mathematical score to identify ADHD. Thus, an automated neuroscience decision support system, that addresses the existing limitations is a possible future direction.

## 8 Conclusion

Attention-deficit Hyperactivity disorder is a common psychophysiological disorder with a genetic component, which can be categorized with core symptoms of inattention, impulsivity, and hyperactivity. ADHD is known as under-diagnosed as there is no exact mechanism to diagnose accurately. The early detection of ADHD helps to treat the affected subjects, avoiding long-term impacts. Data types such as fMRI, EEG, eye movement data are considered for the identification process. It is important to select the appropriate feature extraction technique and the learning algorithm together with the feature combination approach for an accurate ADHD identification process.

This paper has explored the related literature with current challenges and provided possible suggestions as a future research direction. The considered pre-processing, classification and evaluation methods were selected based on the latest studies. The related studies were compared using the given prediction accuracy, performance metrics, scalability, robustness, the applicability of different data scales. The findings and suggestions of this survey will be useful for researchers and practitioners to identify the possible techniques and approaches in the psychophysiological data processing.

## 9 References

- [1] American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). 5th ed., Washington, DC.
- [2] Hamed, A. M., Kauer, A. J. and Stevens, H. E. (2015). Why the diagnosis of attention deficit hyperactivity disorder matters, *Frontiers in Psychiatry*, 1-10. <https://doi.org/10.3389/fpsyt.2015.00168>
- [3] Centers for Disease Control and Prevention. (2019). Data and Statistics About ADHD. [Online]. Available: <https://www.cdc.gov/ncbddd/adhd/data.html>.
- [4] Hao, J., He L. and Yin, H. (2015). Discrimination of ADHD children based on Deep Bayesian Network, *IET International Conference on Biomedical Image and Signal Processing*, 1-6. <https://doi.org/10.1049/cp.2015.0764>
- [5] Centers for Disease Control and Prevention. (2019). Attention-Deficit/ Hyperactivity Disorder. [Online]. Available: <https://www.cdc.gov/ncbddd/adhd/>
- [6] Karyotaki, M. and Drigas, A. (2015). Online and other ICT Applications for Cognitive Training and Assessment, *Int. J. Online and Biomedical Engineering*, 11(2): 36-42. <https://doi.org/10.3991/ijoe.v11i2.4360>
- [7] Gargiulo, R. M. (2010). *Special education in contemporary society*, 4th ed. Sage Publications, Inc.
- [8] Bussing, R., et al. (2008). Parent and Teacher SNAP-IV Ratings of Attention Deficit Hyperactivity Disorder Symptoms, *Assessment*, 15(3): 317-328. <https://doi.org/10.1177/1073191107313888>
- [9] Gizer, I. et al. (2008). Relations between multi-informant assessments of ADHD symptoms, DAT1, and DRD4, *Journal of Abnormal Psychology*, 117(4): 869-880. <https://doi.org/10.1037/a0013297>
- [10] Gualtieri, C. T. (2005). A Practical Approach to Objective Attention Deficit/Hyperactivity Disorder Diagnosis and Management, *Psychiatry*, 2(7): 16-25.

- [11] De Silva, S. Dayarathna, S. Ariyaratne, G., Meedeniya, D., Jayarathna, S., Michalek, A. M. P., & Jayawardena, G. (2019). A Rule-Based System for ADHD Identification using Eye Movement Data. In Moratuwa Engineering Research Conference (MERCon), IEEE explorer, Sri Lanka, 2019. pp. 538-543 <http://dx.doi.org/10.1109/MERCon.2019.8818865>
- [12] Tobii. (2019). Eye tracking for research, Tobiiipro.com. [Online]. Available: <https://www.tobiiipro.com/>
- [13] Glover, G. (2011). Overview of Functional Magnetic Resonance Imaging, *Neurosurgery Clinics of North America*, 22(2): 133-139. <https://doi.org/10.1016/j.nec.2010.11.001>
- [14] Kuang, D., Guo, X., An, X., Zhao, Y. and He, L. (2014). Discrimination of ADHD Based on fMRI Data with Deep Belief Network. *Intelligent Computing in Bioinformatics*, 8590: 225-232. [https://doi.org/10.1007/978-3-319-09330-7\\_27](https://doi.org/10.1007/978-3-319-09330-7_27)
- [15] Brihadiswaran, G., Haputhanthri, D., Gunathilaka, S., Meedeniya, D., & Jayarathna, S. (2019). A Review of EEG-based Classification for Autism Spectrum Disorder, *Journal of Computer Science (JCS)*, vol.15, no.8, pp. 1161.1183. <http://dx.doi.org/10.3844/jcssp.2019.1161.1183>
- [16] Huettel, S. A., Song, A. W. and McCarthy, G. (2008). *Functional magnetic Resonance Imaging*, 3rd ed. Oxford University Press, 267-287.
- [17] Patel, A. X., Kundu, P., Rubinov, M., et al. (2014). A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series, *Neuroimage*, 95: 287–304. <https://doi.org/10.1016/j.neuroimage.2014.03.012>
- [18] Liu, Y., Hu, D., Member, S., Zhou, Z., Shen, H. and Wang, X. (2008). fMRI Noise Reduction Based on Canonical Correlation Analysis and Surrogate Test, 2(6): 870–878. <https://doi.org/10.1109/JSTSP.2008.2008495>
- [19] Tavares, J. M. R. S. (2014). Analysis of Biomedical Images based on Automated Methods of Image Registration, *Advances in Visual Computing*, In: G. Bebis et al. (eds), Springer, 8887: 21-30. [https://doi.org/10.1007/978-3-319-14249-4\\_3](https://doi.org/10.1007/978-3-319-14249-4_3)
- [20] Oliveira, F. P. M. and Tavares, J. M. R. S. (2012). Medical image registration: a review, *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2): 73-93. <https://doi.org/10.1080/10255842.2012.670855>
- [21] NumPy. (2019). About NumPy. [Online]. Available: <https://www.numpy.org/>.
- [22] Pandas. (2019). Python Data Analysis Library. [Online]. Available: <https://pandas.pydata.org/>
- [23] Lever, J., Krzywinski, M. and Altman, N. S. (2017). Points of Significance: Principal component analysis, *Nature Methods*, 14(7): 641-642. <https://doi.org/10.1038/nmeth.4346>
- [24] Hyvarinen, A., Karhunen, J. and Oja, E. (2001). *Independent component analysis*, John Wiley & Sons.
- [25] Kuang, D. and He, L. (2014). Classification on ADHD with Deep Learning, in *Int. Conf. on Cloud Computing and Big Data*, 27-32. <https://doi.org/10.1109/CCBD.2014.42>
- [26] Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J. (2007). *Wavelets and their Applications*. Wiltshire: Antony Rowe Ltd, pp 89-96. <https://doi.org/10.1002/9780470612491>
- [27] Zou, L., Zheng, J., Miao, C., McKeown, M. J. and Wang, Z. J. (2017). 3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI, *IEEE Access*, 5:23626–23636. <https://doi.org/10.1109/ACCESS.2017.2762703>
- [28] Deshpande, G., Wang, P., Rangaprakash, D. and Wilamowski, B. (2015). Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification from Functional Magnetic Resonance Imaging Data, *IEEE Transactions on Cybernetics*, 45(12): 2668-2679. <https://doi.org/10.1109/tcyb.2014.2379621>



- [29] Peng, X., Lin, P., Zhang, T. and Wang, J. (2013). Extreme Learning Machine-Based Classification of ADHD Using Brain Structural MRI Data, PLoS ONE, 8(11): e79476. <https://doi.org/10.1371/journal.pone.0079476>
- [30] Patterson, J. and Gibson, A. (2017). Deep learning: A Practitioner Approach, O'Reilly Media, Inc.
- [31] Solmaz, B., Dey, S., Rao, R. and Shah, M. (2012). ADHD Classification Using Bag of Words Approach on Network Features, The International Society for Optical Engineering, 83144T. <https://doi.org/10.1117/12.911598>
- [32] Haputhanthri, D., Brihadiswaran, G., Gunathilaka, S., Meedeniya, D., Jayarathna, S., Jaime, M., & Jayawardena, Y. (2019). An EEG based Channel Optimized Classification Approach for Autism Spectrum Disorder, in Moratuwa Engineering Research Conference (MERCon), IEEE explorer, Sri Lanka, 2019. pp. 123-128. <http://dx.doi.org/10.1109/MERCon.2019.8818814>
- [33] Ma, Z. and Tavares, J. (2017). Effective features to classify skin lesions in dermoscopic images, Expert Systems with Applications, 84: 92-101. <https://doi.org/10.1016/j.eswa.2017.05.003>

## 10 Authors

**Ms. S. De Silva** is an undergraduate student at the, University of Moratuwa, Sri Lanka. She has research experience in Biomedical and Machine learning. Orcid Id is <https://orcid.org/0000-0002-4526-9586>. Email: [ktsdesilva.15@cse.mrt.ac.lk](mailto:ktsdesilva.15@cse.mrt.ac.lk)

**Ms. S. Dayarathna** is an undergraduate student at the University of Moratuwa, Sri Lanka. She has research experience in Biomedical, Machine Learning and Data Mining. Orcid Id is: <https://orcid.org/0000-0001-8669-2893>. Email: [sanu-waniudara.15@cse.mrt.ac.lk](mailto:sanu-waniudara.15@cse.mrt.ac.lk).

**Ms. G. Ariyaratne** is an undergraduate student at the University of Moratuwa, Sri Lanka. She has research experience in Machine Learning, Data Mining, Biomedical and Computer Security. Orcid Id is <https://orcid.org/0000-0003-4205-6574>. Email: [gcariyaratne.15@cse.mrt.ac.lk](mailto:gcariyaratne.15@cse.mrt.ac.lk)

**Dr. D. Meedeniya** is a Senior Lecturer at the University of Moratuwa, Sri Lanka. She holds a PhD from the University of St Andrews, UK. Her main research interests are Software engineering and Data analytics. She is a Fellow of HEA(UK), MIET, MIEEE and CEng at EC (UK). <https://orcid.org/0000-0002-4520-3819>. Email: [dulanim@cse.mrt.ac.lk](mailto:dulanim@cse.mrt.ac.lk)

**Dr. S. Jayarathna** is an Assistant Professor at the Department of Computer Science, College of Science, Old Dominion University, Norfolk, VA. He received his Ph.D. from Texas A&M University. His research interests include Data science, Information retrieval and Machine learning. <https://orcid.org/0000-0002-4879-7309>. Email: [sampath@cs.odu.edu](mailto:sampath@cs.odu.edu).

Article submitted 2019-04-27. Resubmitted 2019-07-18. Final acceptance 2019-07-18. Final version published as submitted by the authors.