International Journal of Automation and Computing
www.ijac.net

18(2), April 2021, 233-255
DOI: 10.1007/s11633-020-1252-1

Computational Decision Support System for ADHD Identification

Senuri De Silva[1] Sanuwani Dayarathna[1] Gangani Ariyarathne[1] Dulani Meedeniya[1]
Sampath Jayarathna[2] Anne M. P. Michalek[3]

[1] Department of Computer Science and Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka

[2] Department of Computer Science, College of Science, Old Dominion University, Norfolk 23529, USA

[3] Department of Communication Disorders and Special Education, Old Dominion University, Norfolk 23529, USA

Abstract: Attention deficit/hyperactivity disorder (ADHD) is a common disorder among children. ADHD often prevails into adulthood, unless proper treatments are facilitated to engage self-regulatory systems. Thus, there is a need for effective and reliable mechanisms for the early identification of ADHD. This paper presents a decision support system for the ADHD identification process. The proposed system uses both functional magnetic resonance imaging (fMRI) data and eye movement data. The classification processes contain enhanced pipelines, and consist of pre-processing, feature extraction, and feature selection mechanisms. fMRI data are processed by extracting seed-based correlation features in default mode network (DMN) and eye movement data using aggregated features of fixations and saccades. For the classification using eye movement data, an ensemble model is obtained with 81% overall accuracy. For the fMRI classification, a convolutional neural network (CNN) is used with 82% accuracy for the ADHD identification. Both ensemble models are proved for overfitting avoidance.

Keywords: Attention deficit/hyperactivity disorder (ADHD), functional magnetic resonance imaging (fMRI), eye movement data, seed-based correlation, ensembled model, convolutional neural network (CNN), default mode network (DMN), saccades, fixations, ADHD-Care decision support system (DDS).

Citation: S. De Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna, A. M. P. Michalek. Computational decision support system for adhd identification. *International Journal of Automation and Computing*, vol.18, no.2, pp.233–255, 2021. http://doi.org/10.1007/s11633-020-1252-1

1 Introduction

Biomedical data processing has been recognized as a prominent method of classifying fatal diseases for early detection and treatment to prevent severe impacts. As medical data provide relevant information to understand the problems of patients, medical data is also crucial to the classification process of various diseases which is the basis of healthcare informatics[1]. This includes clinical data, which refers to informative health care data with many effective details for the classification process. Many scientific types of research have been implemented to improve the treatment of various medical diseases using biomedical data for experimental purposes. Certain clinical data measurements such as magnetic resonance imaging (MRI), electrocardiography (ECG), electroencephalography (EEG), eye tracking, heart rate, x-ray, etc. can be processed using various novel computational techniques to extract relationships between biomedical data, making the classification process, and precise decision making possible[2]. The process of biomedical data extraction and classification has become an emerging technological need for the identification and classification of medical problems providing rich and valid information for analysis. This paper focuses on using various types of biomedical data for the development of a decision support system for the identification of attention deficit/hyperactivity disorder (ADHD).

ADHD is a common neurological disorder among children, which can be identified by persisting behavioral symptoms of inattention, hyperactivity, and impulsivity. At present, there is an increasing growth in psychophysiological disorders. World statistics have shown that there is an increase of 8.5%–9.5% in the number of children who were diagnosed with ADHD from 2011 to 2017[3].

ADHD has a high prevalence of continuing into adulthood. Also, being a neuro-developmental disorder, there is a potential of having comorbid disorders by nature. Therefore, early identification and treatments must be fa-

Research Article
Manuscript received May 14, 2020; accepted September 8, 2020; published online December 8, 2020
Recommended by Associate Editor Yi Cao
Colored figures are available in the online version at https://link.springer.com/journal/11633

cilitated to avoid long-term negative academic and occupational outcomes[1]. The current clinical identification process of behavioral rating scales is based on the symptomology reported by guardians, hence making this manual process subjective and unreliable[2,4]. This motivates the research into clinical decision support systems that uses healthcare informatics.

Currently, there is a growth in developing computational learning models for neural image analysis to support bio-health informatics[5,6]. The objective of this study is to address the need for systematic measures as an alternative to the use of ADHD classification using subjective psychophysiological measures. We propose an integrated system for both the child and adult ADHD identification process, which resolves the limitation of bounding the classification process into one specific age group as in related studies. As a novel contribution, the proposed system uses functional magnetic resonance imaging (fMRI) data in the resting-state brain for the children, and eye movement data gathered during a working memory capacity (WMC) task for adults. The methodology uses appropriate data pre-processing and classification learning models that are selected using an extensive literature review. Additionally, our work presents a novel methodology to obtain an objective rating scale to indicate the severity of the disorder for the subject who has been identified with the possibility of having ADHD. Thus, the proposed approach will assist psychiatrists with diagnosing ADHD and aligning suitable treatments accordingly for both children and adults.

As the use of eye gaze data is emerging in the field of cognitive disorders, the feasibility of using psychophysiological measures in ADHD classification is possible[7]. Hence, we believe using both eye movement and fMRI data would increase the possibility of using these physiological measures for accurate and valid ADHD classification. Considering eye movements, WMC tasks have generated identifiable abnormalities or differences in fixations and saccades in ADHD subjects that are sufficient for the classification process.

This study contributes to the design and development of an approach to extract features from eye movement and fMRI data, and classify them while maintaining high accuracy, specificity and sensitivity by avoiding overfitting. As far as we know, this is the first study that has focused on both child and adult ADHD classification using two different datasets. Thus, having an extendable framework is one of the novelties of this study. This resolves the constraint of bounding the data classification into one specific age group. The development is based on a serialized composition of a set of derived models. Thus, the pipeline task implementation is another notable feature. In addition, we propose a mechanism to measure a similarity score for the prediction of the disorder. Moreover, we present an online ADHD-Care application as two separate models in a single framework decision support system (DSS) that can be easily used by medical practitioners and technicians over different platforms and locations. Since this research mainly targets children and adolescents, early detection and classification will be helpful to improve their lives as other control groups.

This paper is structured as follows. Section 2 describes the eye movement related studies and the techniques used for pre-processing, feature selection, and classification of fMRI data. Section 3 discusses the datasets, methodology in terms of the techniques used along with the architectural aspects of the overall system. The key features of the overall system are discussed in Section 4, and Section 5 states the discussion of obtained results and comparison with the existing work.

2 Background

2.1 Overview of eye movement data

Eye movement measurements provide relevant information for the neuropathology of various disorders which involve neurological substrates[8]. Therefore, analysis of eye movements has become a prominent technique in neuroscience to understand the complex neurological underpinning of psychiatric disorders. Eye movement tracking provides informative data on visual and attention abilities[9]. Eye movement measurements provide an understanding of different psychiatric disorders based on their underlying cognitive neuroscience[10]. For instance, eye movement-related data are mainly used for the classification process of ADHD, reading disorder (RD), and Autism, which provide a rich dataset for metrics of oculomotor control[11].

The oculomotor controls for performed tasks, revealing deficits that are related to each underlying process as a result of a composite relationship between eye movements and cognitive tasks[8]. As gaze related data or the eye movements interact with congruent cognitive processes, the derived oculomotor metrics can be used to detect deficits in the above process for a variety of subjects. Thus, it provides indications of different complex behaviors of people who are suffering from psychiatric disorders. Different types of gaze related movements have been studied for research purposes to gain quantitative information on the classification process. Eye movement data include a variety of oculomotor controls such as duration, velocity, trajectories of saccadic movements, and fixations[12]. Saccades are rapid eye movements that redirect gaze. It moves the gaze centre from one fixation point to another during the scanning process, thus, suppressing the visual observations[9]. Fixation is known as the movement of gaze which is fixed at a point typically lasts for 250 milliseconds on average. During fixation, the gaze point is on a target stimulus and needs a series of saccades. ADHD subjects are found to perform many intrusive saccades during tasks where they need to maintain fixation[7].

2.2 Eye movement data related studies

Among related studies, Fried et al.[13] have explored oculomotor markers of ADHD that can be used to differentiate ADHD subjects with normal groups. The testing process based on eye movements were able to reveal higher number of average micro-saccades, and blink rates of ADHD groups in the time interval around the onset of a stimulus. These rates were monotonically increased for both groups. However, in ADHD group these rates were increased rapidly, with a decrease of the pupil diameter over the session. In another study, Michalek et al.[7] analyzed eye movements and their patterns to detect ADHD confirming that complex span tasks used to measure WMC generated fixation features which could be used to classify ADHD. They have found that adults with ADHD have fixations primarily below the stimulus with less direct fixations toward components within sentences[7].

Moreover, Ross et al.[14] have evaluated deficits in working memory for both ADHD and schizophrenia patients. Eye movements of both subjects in oculomotor response tasks have used a large percentage of premature saccades, while schizophrenia subjects have shown impaired working memory. A similar study by Ross et al.[15] has used smooth pursuit eye movement to compare the abnormalities in ADHD and schizophrenia. They have shown that many ADHD subjects have no smooth pursuit abnormalities with worse smooth pursuit gain, which allow the eyes to closely follow a moving object. However, leading saccades are present in schizophrenia subjects.

Eye movement tracking in the identification of ADHD and reading disorders has significantly contributed to identifying atypical gaze related parameters. Munoz et al.[16] have shown the difficulties of ADHD subjects in maintaining fixation for an extended time duration with reduced saccadic eye movements. These findings were also supported by Deans et al.[9], where subjects diagnosed with the reading disorder had a longer duration of fixations, and ADHD subjects tended to have atypical eye movements compared to healthy groups. Saccades and fixations related gaze parametric data were also used to develop a rule-based system for the classification of ADHD[11]. This study considered eye moment and eye positions of different gaze event type data extracted from a working memory task. They have considered the gaze points, saccades, fixation and pupil diameter to extract the features. Their methodology based on decision tree classification has shown high accuracy. Hyun et al.[17] have assessed the visuospatial working memory impairments of ADHD adolescents compared to normal subjects by conducting a rey-osterrieth complex figure test, that is based on copying and recalling capabilities.

A summary of eye-movement classification related studies is stated in Table 1, where N/A refers to the studies which tested the different eye movement pursuits in ADHD. These studies have not been specifically used for ADHD classification. Hence, accuracy values are not included. The statistical approaches include techniques such as analysis of variance (ANOVA) test, two sampled, and paired t-tests. The decision tree algorithms consist of techniques such as linear model tree (LMT), random forest (RF), J48, and decision stump. The classification rule-based algorithms include Part, and JRip[18,19].

2.3 Overview of fMRI data

Functional magnetic resonance imaging (fMRI) is a brain imaging method that was developed to denote time-varying, regional variations in brain metabolism.

fMRI can detect abnormalities in the brain by examining the brain′s functional anatomy. fMRI uses magnetic resonance imaging (MRI) to evaluate the small changes of blood flow in the active part of the brain[20]. The variations happen as a result of either the task producing cognitive state variations or the effect of resting-state brain′s unregulated processes. Fig. 1. illustrates the sample 4D fMRI data in 2D format from the ADHD-200 global competition New York University (NYU) dataset[21], portraying the first five slices of the volume.

The basic form of fMRI uses the contrast of blood oxygen level dependent (BOLD), which indicates the input, integrative processing, and the output firing of neurons. fMRI is mainly used to determine the brain's functional anatomy, and to evaluate the diseases or as a guide for brain treatments. Since fMRI has the qualities of high availability, non-invasive, relative cost-effectiveness, and better spatial resolution, it is widely used for clinical psychology, cognitive neuroscience and presurgical planni-

Table 1 Summary of related studies on eye movement data

Related study	Features	Used technique and obtained accuracy		
		Statistical approach	Decision tree algorithm	Rule-based algorithm
[7]	Fixations	–	91.11%	–
[13]	Blinks, micro-saccades, pupil diameter	70.00% (ANOVA)	–	–
[11]	Fixations, saccades, pupil diameter	–	84.48%	82%
[14]	Saccades	(Post-hoc)	–	–
[9]	Fixations, saccades, reading time	60.00%	–	–
[16]	Saccades	(P-test)	–	–

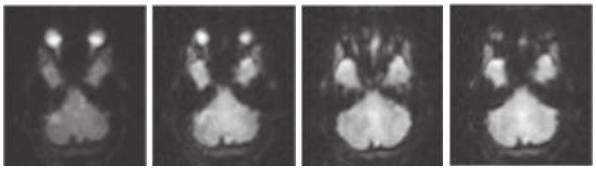

Fig. 1 Sample 4D fMRI data in 2D format[21]

ng[2, 22, 23].

2.4 fMRI related studies

Several fMRI related studies[24] have revealed various brain patterns that are a result of different mental processes, which connect brain activation with mental functionalities. There are significant approaches available in related studies[25–27] to handle fMRI for neurological disorders. The study by Subbaraju et al.[25], have used resting-state fMRI data to identify various brain activities due to autism spectrum disorder. Another study by Dhayne et al.[27], has further proved the propensity of handling fMRI data to identify neurological disorders. ADHD is highly related to the default model network (DMN), which is a broad brain network in neuroscience. Table 2 summarizes some related studies. On attention-demanding tasks, ADHD subjects display complications in suppressing default mode activities. Therefore, studies have addressed the effect of brain regions in the DMN when classifying ADHD data[28].

Table 2 Summary of related studies for fMRI data

Relatedness	Limitation
ADHD classification framework with functional connectivity measures and different classifiers[6].	Limited with customizable classifier implementations, thus, accuracy values may not strong.
A classification with saccades and fixa-tions in working memory to test the feasibility metrics for ADHD identification[7].	Only the tree-based classification algorithms are applied.
Clinical usefulness of eye movement data to classify ADHD and reading disability (RD) using the reading time and left to right saccades[9].	Less valid data and brief reading task. Results provide less support to distinguish ADHD and RD based on eye movements.
Eye movements related to oculomotor response tasks consideration for both ADHD and schizophrenia subjects to show dis-inhibition[14].	The task cannot differentiate primary from secondary inhibitory working memory dysfunction.
Seed-based correlation (SBC) feature extraction for fMRI classification for ADHD subjects[23].	Mainly based on DMN brain region. Thus, the classification can be improved considering different voxels.
Examine a 3D CNN model to classify ADHD by utilizing fMRI and structural magnetic resonance imaging (sMRI). Encoded prior knowledge on 6 types of 3D low-level features (ReHo, fALFF, voxel-mirrored homotopic connectivity (VMHC))[24].	FMRI and sMRI data are typically analyzed separately and the joint information is not fully explored.
SBC approach to identify the main DMN regions for ADHD classification[28].	The classification is based on a single fMRI measure and 4 DMN regions. Results may be subject to overfitting.
Establish an automatic and efficient ADHD classification using ELM learning algorithm on structural MRI data[29].	Only considers structural MRI data from the subjects in the ADHD-200 Global Competition.
Shows the utility of FCC ANN archi- tecture for ADHD classification. Used connectivity-based features[30].	The comparison was only done with the deep belief network (DBN) classifier and only applicable for adult data.

The work by Ariyarathne et al.[28], used pre-processed fMRI data using configurable pipeline for the analysis of connectomes (C-PAC) application for the resting-state of the brain to classify ADHD subjects. This study used a seed-based correlation to obtain the functional connectivity among the seeds and identified brain voxels. The extracted seed correlations from DMN have been used for the classification process based on convolution neural networks (CNN). They have identified the most correlated brain areas of ADHD subjects, assuring high accuracy, sensitivity, and specificity values, and showed the possibility of identifying the correlated brain regions of ADHD subjects. An extension of this work is presented in [23], that uses un-processed fMRI images for ADHD classification using CNN including optimal feature selection methods before applying the learning model. They have shown that seed-based correlation gives high classification accuracy values compared to other feature extraction methods such as the fractional amplitude of low-frequency fluctuations (fALFF) and regional homogeneity (ReHo)[23].

A similar study is presented with different classifiers[6] including a generic decision support system (DSS) for neuroimaging data to facilitate a single platform for the computational solutions of psychophysiological chronic disorders. Their methodology is based on different preprocessing and learning models applicable to diverse datasets for better decision making because many neurological disorders share commonalities. The proposed methodology is tested for ADHD with fMRI data using selected classifiers over three types of functional connectivity. The proposed generic platform can be extended as a support tool for practitioners to make confident decisions.

Among the characteristics of ADHD that include executive functions, working memory, and behavioral functions, many studies have addressed the modalities in the working memory. The abnormalities of ventrolateral prefrontal and cerebellar regions have been reported and verified using fMRI data[31].

A study obtained acceptable accuracy in identifying ADHD on frequency domain with respect to DMN areas on fMRI features in the cingulate cortex, visual cortex and prefrontal cortex that ADHD[5]. They analyzed the executive functional defects in ADHD subjects and the resulting attention lapses were a result of the reduction of network activity[5]. The recent studies on neurological behavior of ADHD can be categorized as the connectivity of the different regions of the brain such as medial frontal cortex[32], posterior cingulate gyrus, and default model network component b (DMNb)[33, 34]. Also, many studies[33, 35] have been focused on identifying the anatomical and functional differences between ADHD and con-

trol subjects.

Feature extraction and selection methods are an important aspect of the learning models to reduce the high dimensionality of data and to discard the noise in fMRI data[2] independent component analysis (ICA), and principal component analysis (PCA) are some of the common feature extraction methods used for modelling learning[36]. fALFF, ReHo, amplitude of low-frequency fluctuations (ALFF), and resting state network (RSN) are some of the most considered features[24] of ADHD classification using the fMRI data. ReHo features are used to relate voxel′s functional activities and its neighbors. Here, the extracted ReHo maps are combined with PCA to classify ADHD.

For the ADHD classification, several studies[6,24,30,37] have used various machine learning approaches, i.e., support vector machine (SVM), extreme learning machine (ELM), convolutional neural networks (CNN), deep Bayesian network (Deep-BN). There are many related studies[30,38] that followed the SVM to classify ADHD, where the study by Tenev et al.[38], achieved 85% accuracy using the SVM model. Peng et al.[29], built a strong comparison between ELM and SVM classifiers in terms of accuracy and performance. In this work, ELM is identified as the most accurate classifier compared to SVM as 90.18% accuracy has been shown using the ELM learning algorithm. Further, the Stockwell transform of the time-series based approach has shown an accuracy of 96.68%[39].

The deep belief network (DBN) based approach was effectively used by Kuang and He[40] for ADHD classification as it analyses fMRI brain regions. The deep Bayesian network (Deep-BN) that merges Bayesian network (BN) with DBN by combining the dimension reduced features and global feature extraction has obtained 64.70% accuracy by applying the SVM classifier[37]. Zou et al.[24], have presented a 3D CNN based model with multi-modality CNN architecture by combining structural and functional fMRI. The study of the fully connected cascade neural network architecture (FCC-ANN) has shown its adequacy for classifying ADHD subjects by relating its high discriminatory power[34]. Further, Deshpande et al.[41], have shown that FCC-ANN outperforms ADHD classification accuracy with fMRI data compared to other classifiers, irrespective of the considered features.

2.5 Clinical decision support system

A clinical decision support system (CDSS) is a knowledge-based health information system which supports decision making by compiling information from the specific data to facilitate valid and reliable decision making by medical practitioners. CDSS combines patient specific details with clinical knowledge thereby providing early disease detection and resulting improvements in patient care. Accordingly, CDSS provides patient data analytics capable of improving the quality of outcomes. Excessive testing can be reduced, and patient safety can be enhanced while providing psychiatrists with reliable and consistent information. However, existing clinical practices lack CDSS due to technical constraints in research and development of such DSS[2,42]. Using an ensemble model in building a CDSS is efficacious in identifying ADHD accurately, since the ADHD has significant variability across the patients[43].

Generally, CDSS helps to diagnosis disorder conditions during clinical practices. Delavarian et al.[44] have presented a CDSS that acts as a preliminary assistant to distinguish ADHD subjects among the children. This study has shown promising classification accuracies. Further, Chu et al.[45] have used the basis of test of variables of attention (TOVA) detection and screening tool to build a DSS for ADHD which has proved a reliable system to screen ADHD.

3 System design

3.1 Overview of the decision support system

In this study, ADHD DSS architecture is based on eye movement and fMRI features of ADHD and non-ADHD subjects. Before the extraction of relevant features, the raw fMRI, and eye movement data are pre-processed using the pipeline specified in [2,11]. After the feature extraction, fMRI data are fed into the deep learning model of CNN, and various machine learning models are evaluated for classification based on eye movement data. Two ensemble models were built for each of the data types based on their classifiers. Finally, the decision support system is built using the classifier models to infer an objective score for a patient based on the severity of the ADHD. The pre-processing, feature extraction, classification, and evaluation technique pipeline of our system is summarized in Fig. 2 and described in detail in Section 4. Based on the proposed methodology and the instructions

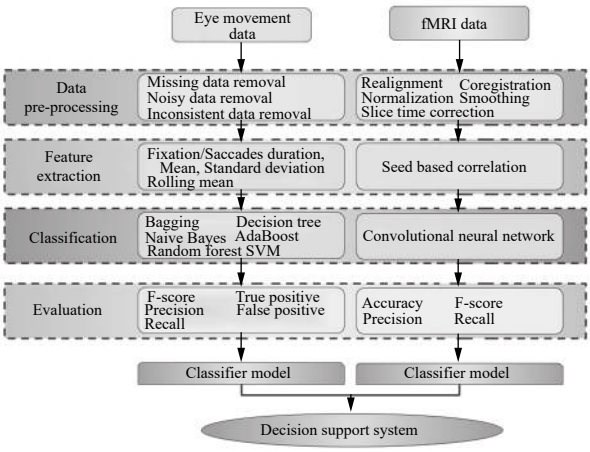

Fig. 2 System overview of the ADHD-Care DSS

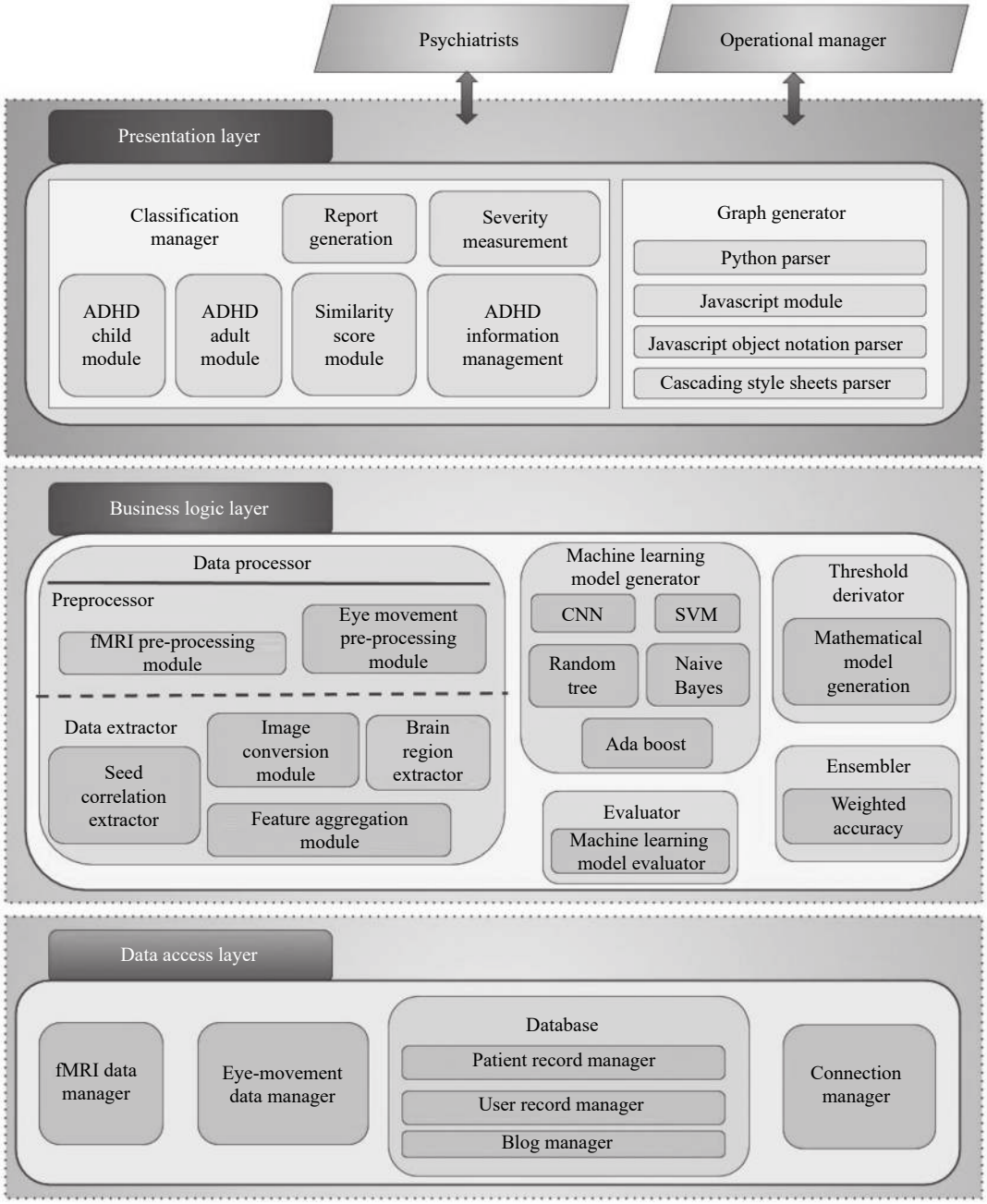

Fig. 3 System architecture

from the professional psychiatrists, ADHD-Care web application was designed[46].

As shown in Fig. 3, the application was developed according to the layered architecture including the presentation layer, business logic layer and data link layer. The internal models are described in Section 4 with the implementation details. The layered architecture of the ADHD-Care web application is designed to facilitate the maintainability, flexibility, and extendibility of the system. As the main users of the system, psychiatrists and operational managers are interactively involved. The main components in the presentation layer are classification manager and graph generator. They consist of submodules to manage the classification process and the graphical representation of the system, respectively.

The business logic layer comprises a data pre-processor, machine learning model generator, evaluator, threshold derivator and the ensemble model to comply with the system overview as shown in Fig. 2. The modules of fMRI and eye movement managers and the connection manager belong to the data link layer along with the system database. The backend database stores patient details, user details, patient's history and updated details on ADHD. The layered architecture of the ADHD-Care web application is designed to facilitate the main-

tainability, flexibility and extendibility of the system.

4 Methodology

4.1 Experimental setup for eye movement data

The processed eye movement dataset is provided by Old Dominion University, Virginia. A Tobii Pro X2-60 eye tracker was used with Tobii studio analysis software to detect eye movements of participants[7]. The accuracy of the eye tracker is 0.40 and precision is 0.340 (degree of visual angle). A total of 14 participants between the ages of 18–35 were engaged in the experimental process. Seven adults from the total participants were diagnosed to have ADHD and their classification was confirmed by the physicians′ relevant documentation. All the participants were asked to avoid medication for 12 hours before the experiment.

As shown in Fig. 4, during the experimental process, the participants were asked to read the sentences that were displayed on the computer screen along with the letter at the end of each sentence. While participants are performing the given tasks, their detailed fixations, saccadic eye movements were recorded by the eye tracker. Apart from that, fixation duration in milliseconds, pupil diameters of left and right eyes and class label, gender were acquired.

4.2 Eye movement data processing

The obtained eye movement data were pre-processed by applying the data cleaning techniques such as noise removal and removing duplicate values and inconsistent data[11]. Then, the missing data were handled by replacing them by the mean value of the feature if the record does not contain more than 10% missing values. When deriving the correlation with the class label, it was visible that gender has a strong influence on it. This situation is known as a presence of bias because common phenotypic data such as gender, the race would introduce bias to the data. Also, it becomes an inherent bias to the developed machine learning model, which was removed under bias correction in data pre-processing by adding a fairness regularizer and calibrating the prediction probability threshold to maintain fair outcomes. The correlation of the entire feature set and the correlation with the class variable are shown in Fig. 5. The variation of the data was handled by normalizing the feature values to a common range.

Aggregated data were generated by applying statistical functions for an interval of adjacent fixations or saccades that gives the mean, standard deviations for both fixations and saccade durations, and performed correlation analysis among all other features. We also removed the irrelevant features with empty values and the features that have the same values for all the subjects. The selected set of features are shown in Fig. 5. Moreover, we have measured the accuracies for different classifier models with the selected features of the pre-processed data. Initially, classifier models were obtained by using the classification ruling and decision tree algorithms. Rule-based algorithms were used in the classification process[11].

Rule-based algorithms which have given higher accuracy values were used for the final ensemble model along with other machine learning algorithms. These models were evaluated using the 10-fold cross-validation and the 66% split as discussed in Section 6. During the process, the models with high classification accuracy values are selected. The selected classifiers of bagging, SVM, random forest, decision tree, AdaBoost, and Naive Bayes were

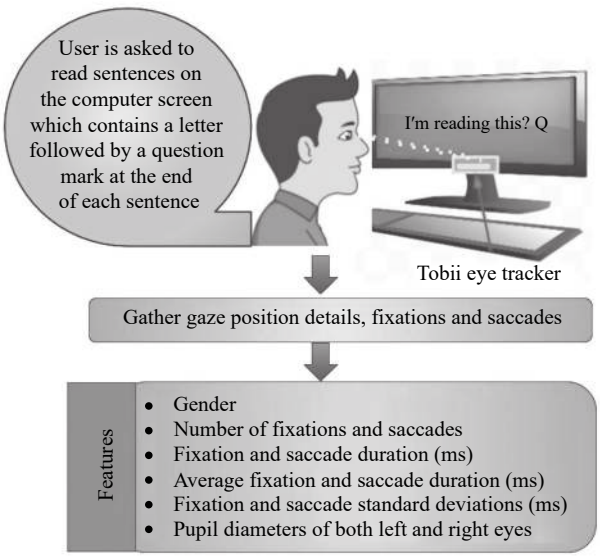

Fig. 4 Experimental setup for eye movement

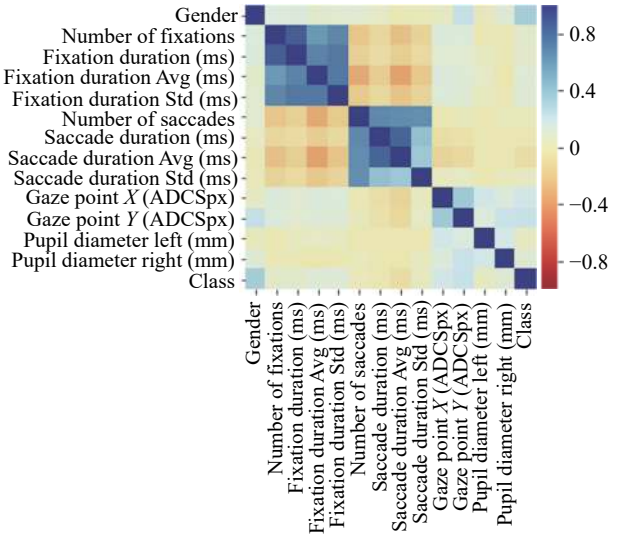

Fig. 5 Correlation matrix of all eye movement features. Color figures are available in the online version.

again applied to ADHD classification and obtained the accuracy values. According to the given accuracy values, weights were applied to all the six classifiers and designed an ensemble model as the final classifier model of the ADHD identification using the eye movement data.

4.3 Experimental setup for fMRI dataset

The fMRI dataset was obtained from the New York University (NYU) Child Study Center of ADHD-200 global competition data repository[21]. The dataset consists of 266 participants of children aged between 7 to 18 years. The demographic details are presented in Fig. 6, as box plots of verbal IQ and age in the form of min, quantile 1, quantile 2, quantile 3 and max values. Among the 266 subjects, 133 were labelled as ADHD positive and 133 were labelled as negative. Few original records were discarded because they contained more than 10% missing values. The subjects selected were fairly representing the population considered and the IQ measures indicate that the children have enough cognitive ability for the experimentation. The verbal IQ was measured by the Wechsler abbreviated scale of intelligence (WASI) scale and the fMRI dataset includes brain scan images of subjects with the verbal IQ from 82 to134.

4.4 fMRI data processing

The fMRI data processing is based on 266 fMRI 4D images and the process view is shown in Fig. 7. Pre-processing of fMRI data was categorized into the main stages of realignment, slice-time correction, normalization, co-registration, and smoothing[2,4]. The pre-processing was done using statistical parametric mapping (SPM12), which is a free-open source software for pre-processing fMRI data.

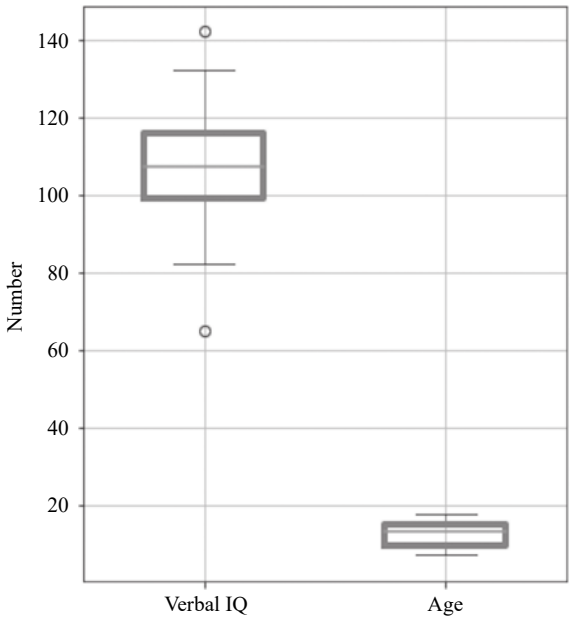

Fig. 6 fMRI data verbal IQ in WASI scale and age groups

First, the neuroimages of 4D fMRI data were realigned using SPM12 to remove movement artefacts by realigning the time series values. This process has obtained the optimal values by calculating the sum of squared differences between images. Then the realigned fMRI data was used for correction of slice time. Slice time correction was performed to adjust the image values by making all the image voxels appear as they were taken at the same time. The task was performed by making the data on each slice, which corresponds to the same point in time. Here, the data was temporally aligned by shifting the signal phase. The slice acquired in the middle of the sequence was used as the reference slice. Next, the fMRI data were normalized to reduce the effect of brain size, as the size of the human brain varies from 30% from each other[47]. SPM12 is used to identify differences of image volumes and used different stretching, squeezing, and warping techniques to normalize the brain images. Under the normalization, noise reduction and bias correction were also performed.

Coregistration was performed to realign different images from the same participant. The feature similarities were identified using image matching and interpolation. The inter-subject averaging as spatial normalization does not well-align all structures. Hence, the fMRI data was smoothed to increase the signal to noise ratio (SNR) by averaging the voxels with a weighted sum of its neighbors. The weight is defined using the Gaussian kernel. Smoothing was used to detect true and task-related changes in the signal.

The feature extraction process is based on 266 4D fMRI pre-processed images. The seed-based correlation approach is used in this study to extract the features in the model generation[23,28]. A seed-based correlation has been used to define the functional connectivity of the brain because the connectivity metrics in the brain dynamics are predominant in creating the connectivity maps. The selection of seeds directly affects the generated network accuracy. Similar to the accuracy, the factors such as the simplicity and unsophisticatedness in implementation make this a better approach compared to the other techniques like PCA, clustering, graph-based methods, etc. To compute the correlation, resting state-fMRI data were used by considering the DMN regions as the seeds[28].

Generally, a default model network shows the active brain regions when the brain is in resting-state. Various DMN regions provide different information for ADHD identification[5]. The seeds used for the correlation calculation are posterior cingulate cortex (PCC), medial prefrontal cortex (MPC), left temporoparietal junction (LTJ), right temporoparietal junction (RTJ), hippocampal formation (HF), inferior parietal cortex (IPC) and medial temporal lobe (MTL) in medial view[48]. These re-

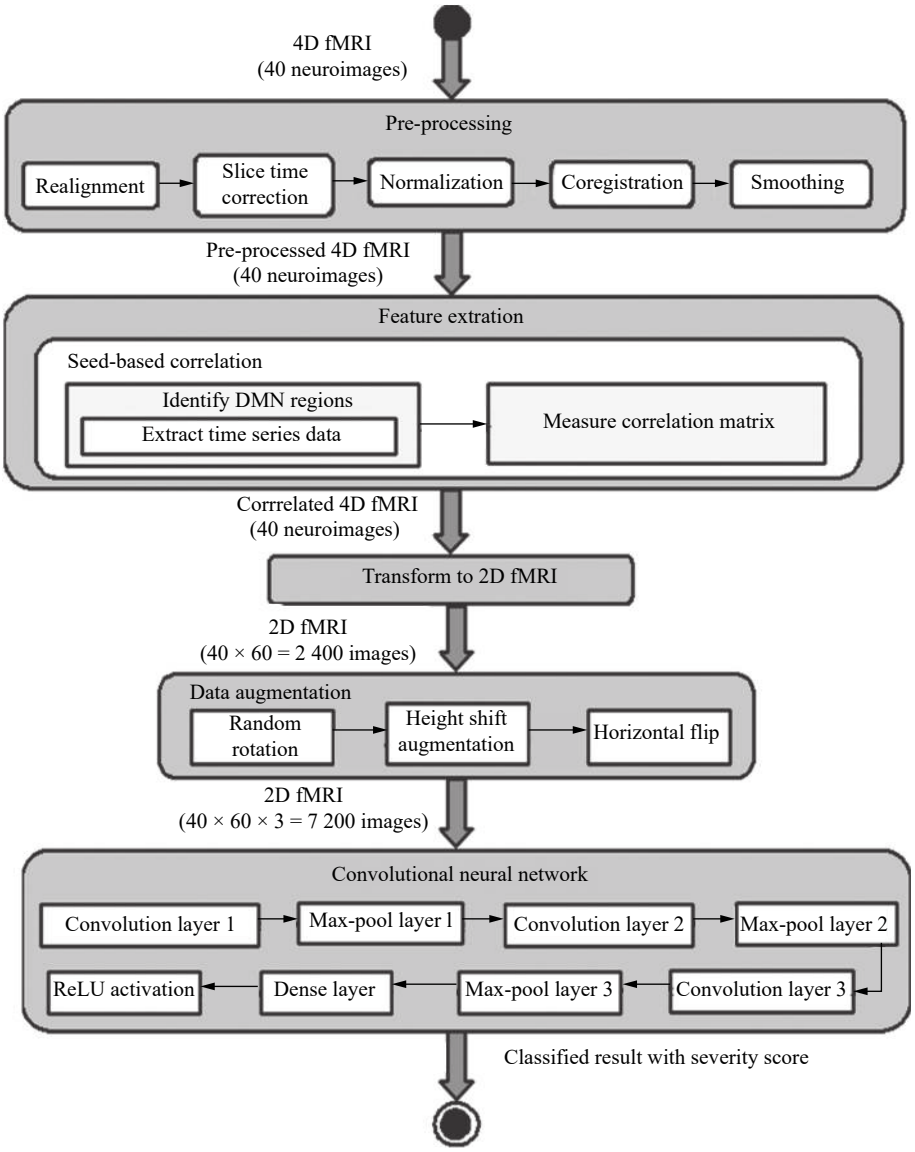

Fig. 7 Flow of the fMRI data processing

gions were selected by considering the effective neural activity of sub-regions in the DMN and to test which of these regions are affecting ADHD the most. Fig. 8 shows an overview of different DMN regions.

We extracted the DMN regions of the raw data using the Nilearn python library. The time series of the DMN region of interests (ROI) was used to calculate the correlation with the other voxels of a subject. The graphical representation of the anatomical parcellation is portrayed in Fig. 9 for the specified regions. These images are generated in such a way that a sample neuroimage is applied to the Nilearn module with the desired seed region coordinates via labels. Then the ROIs on the surface are plotted by the library. In Fig. 9, the left and right temporoparietal junctions are defined as the supramarginal and superior temporal gyrus in lateral view, where the intersection between these regions are known as the temporoparietal junction.

We then transformed the volumes of 4D fMRI images, which contain the extracted seed correlation features of different DMN regions to 2D images of 64 × 64 pixels, where each 4D image is represented by 60 sliced 2D images. The implementation was done using med2image, an open-source Python library. The 266 × 60, 2D fMRI images were fed to the data augmentation process. This was used to increase the dataset using three augmentation techniques, as shown in the process view in Fig. 7. The resultant 15 960 × 3 2D fMRI images were fed for the learning model.

The implemented CNN consists of 7 layers, where the dataset is 47 480 2D fMRI images. The classification process was repeated for all the DMN regions by generating CNN models with extracted correlations from pre-processed fMRI images. Since this study has used the seed-based correlation of DMN regions as the main features for CNN, the result is used to evaluate the effect of different

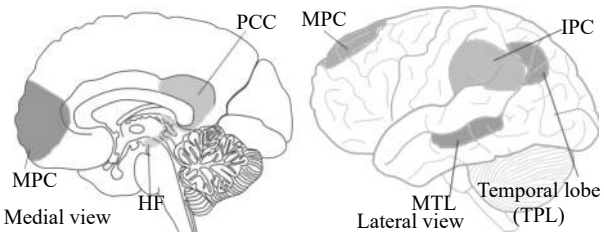

Fig. 8 DMN regions and cortexes considered for the study

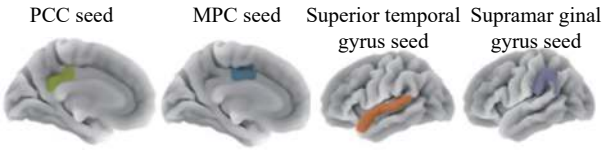

Fig. 9 Seeds used for correlation generation in fMRI data

DMN brain regions for ADHD identification. This process extracts the connectivity between seed voxels and other remaining brain voxels, instead of feeding the whole brain images directly to the neural network.

The learning model of the proposed solution is based on the binary classification to categorize the data into two classes by assigning each data point to one of the categories. In binary classification, a learning function is learned to minimize the probability of misclassification of data into two classes[42]. The actual prediction value of most of the binary classifiers is a score that gives the certainty of an observation fit into the positive class.

This prediction score can further be interpreted by comparing it against a threshold value to efficiently categorize observations into one of the two classes. The observation with a score value less than the threshold is predicted as a negative class[49].

The CNN architecture is shown in Fig. 10, altering convolutional and max-pooling layers of 64 and 32 channels, respectively. Initially, in the convolution and max-pooling layers, a smaller channel size of 32 is chosen, to identify the features such as edges. As the neural net goes deeper, complex features are extracted, hence dense filters of convolutional and max-pooling are required. A channel size of 64 is used for the latter part of the neural net. The dataset was split in the ratio of 80% and 20% for the training and testing sets, respectively, based on the Pareto principle[50,51].

Next, the complexity of the model is reduced avoiding overfitting of the CNN model by adjusting the number of activation functions and the number of layers. Since there is a high prospect of the model overfitting, the peculiarities of the training data, regularization methods were applied. The general regression equation in condensed form is given in (1), where b is the influence and W is the weight.

$$Y = Wx + b. \tag{1}$$

In CNN, the bias and weights adjust continuously to train the network. When the bias (b) is low, the model fits well on the training set and resulting in low training error. The kernel regularization modifies the error function, such that it decreases the weights and bias. Thus, smooths the output and minimizes over-fitting. The technique RMSprop[52] was used as the model optimizer for the generalization, where it divides the learning rate by a rolling mean of the recent magnitude found.

4.5 Severity score threshold derivation

We defined a threshold to obtain the correct classified value, as the learning model is based on binary classification. Since the machine learning model involves randomness by its nature, a pre-defined threshold for the classification process is beneficial to identify the severity of the disorder. Since there are two models derived for adult data and child data, the thresholds are defined separately.

Generally, the receiver operating characteristic (ROC) curve shows the relationship between sensitivity and a function of (1 − specificity), for different cut-off points. Thus, a given point on the curve shows the pair (sensitivity/(1 − specificity)), for a given threshold[52]. For the learning model of the eye movement data, the threshold was defined using the model's receiver operating characteristics (ROC) graph as shown in Fig. 11. It contains the false-positive rates and the true-positive rates gained for different thresholds considered. Since there is a trade-off between these rates, it is vital to keep a balance and find the optimal cut-off point defined as the threshold value.

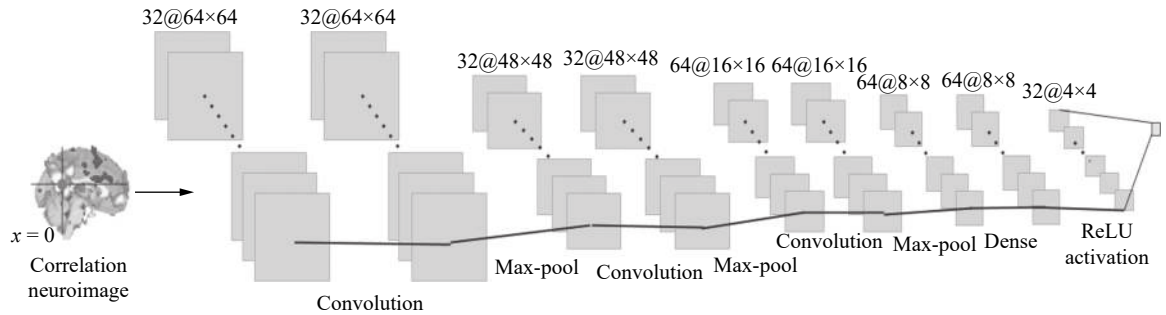

Fig. 10 Convolutional neural network architecture

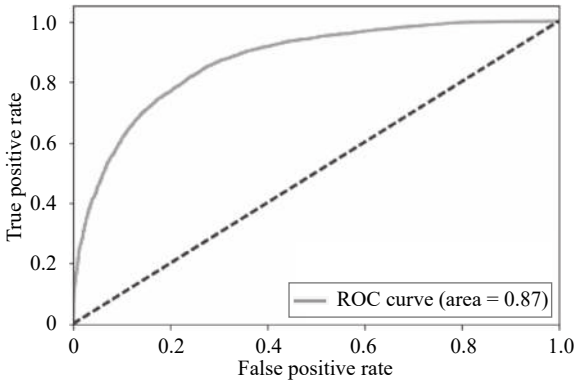

Fig. 11 ROC graph for classification using eye movement data

Several methods were used to derive the threshold value, using sensitivity (Se), specificity (Sp) and area under the curve (AUC)[53,54]. Sensitivity is the probability of the positive test result when the condition is present. Specificity is the probability of the negative test result, given that the condition is absent. The area under the curve refers to the intrinsic accuracy apart from the classification thresholds, which is plotted as sensitivity plotted versus (1− specificity)[53].

Several studies have used the Youden index, concordance probability, and index-of-union to derive the optimal cut-point value in ROC analysis and assess the effectiveness of bio-markers[55–57]. The Youden index[55] is defined in (2), where the index as the maximum is identified as the threshold, where Se refers to the sensitivity of the model and Sp refers to the specificity of the model.

$$\mathrm{J}(c) = \{\mathrm{Se}(c) + \mathrm{Sp}(c) - 1\}. \tag{2}$$

We also used the concordance probability method to calculate the optimal cut-off point[55] as defined in (3). The cut-off point is selected as the point where the CZ value is maximized.

$$\mathrm{CZ}(x) = \mathrm{Se}(c) \times \mathrm{Sp}(c). \tag{3}$$

Index-of-Union (IU)[55] is a recently developed threshold defining a method in medical research, where the cut-off point is defined when the IU values are minimized, where AUC is the area under the curve.

$$\mathrm{IU}(c) = \{|\mathrm{Se}(c) - \mathrm{AUC}| + |\mathrm{Sp}(c) - \mathrm{AUC}|\}. \tag{4}$$

The optimal cut-off point is defined as the point which classifies most of the individuals correctly and thus least of them incorrectly. That is, the optimal cut-off points refer to the point where sensitivity and specificity are maximized[58]. According to the obtained cut-off points as given in Table 3, the index-of-union method has shown the minimum mean squared error compared to the other two methods.

The mean of the results of these three methods has produced the lowest mean squared error values. The optimal cut-off points for the eye movement and fMRI data are obtained at 0.556 1 and 0.555 4, respectively by calculating the mean of the values obtained for the considered methods. Since we have considered a threshold value instead of a given range of values, it does not affect the confidence level.

Table 3 Derived classification thresholds

Technique	fMRI data	Eye movement data
J	0.564 4	0.558 9
CZ	0.533 1	0.559 6
IU	0.568 9	0.549 8
Optimal point	0.555 4	0.556 1

4.6 Ensemble model derivation

The proposed solution has used different classifiers and two datasets to predict the accuracy results to support ADHD classification. In this study, we have derived a separate ensemble model for each dataset classification process, to aggregate the result of each model avoiding overfitting to predict the accuracy of the test data. Different DMN regions were used for the fMRI data-based ADHD classification, and obtained the corresponding accuracy, specificity and sensitivity values. It was found that some of the specific regions with high accuracy values are highly active in ADHD subjects. However, we have used all the brain-regions in fMRI to build the ensembled model, as all the DMN regions contribute to the final classification model to reduce the bias and variance in the model that are caused by the training data. The ensembled model was generated using the active brain regions by assigning weights for each DMN based on the obtained classification accuracy value.

For eye movement data classification, the highest accuracy is obtained by the random forest classifier compared to the other decision tree algorithms shown in the tables of Section 6.1. Generally, the rule-based algorithms have shown overfitting issues on building the final model[11]. Along with random forest classifier, bagging, SVM, decision tree, AdaBoost, and Naive Bayes were evaluated using 10-fold cross-validation and the 66% split to select the best classifiers to generate the ensemble model. Finally, the ensembled model was implemented using the selected learning models by assigning weights for each classifier proportional to the acquired accuracies.

The weight assignment of the ensemble model mainly considers the classification accuracy values obtained in individual model generation. Different weights were assigned based on the value of the accuracies, as it is not applicable to consider all the classifier results equally. Thus, the weights are calculated such that the assignment of weights minimizes the mean squared error of the summation of weighted models, for each iteration. Here,

the weight is proportional to the accuracy obtained for the model, as shown in (5), where α is the proportional constant. This mathematically equals to $w_i = ka_i$, where k is a constant. In each iteration of the generation of the ensemble model, the weight is recalculated proportionally to the accuracy, such that the MSE is minimized for all the considered models. In this process, the initial weights (w_i) are assigned randomly for each model.

$$w_i = ka_i. \tag{5}$$

The ensemble model is generated based on (6), where w_i refers to the calculated weight and y_i denotes the accuracy prediction for a given model i for the testing data.

$$EM = w_1.y_1 + w_2.y_2 + w_3.y_3 + \cdots + w_n.y_n = \sum_{i=1}^{N} w_i y_i. \tag{6}$$

The prediction accuracy is calculated for the model generated by ensembling all the models as given in (6). Then this ensemble model is tuned up by changing w_j values such that the total MSE for the ensemble model $(w_1y_1 + w_2y_2 + \cdots + w_jy_j)$ is minimized. Consequently, the ensembled model is selected based on the minimum mean squared error (Min MSE), which is defined as the summation of bias and the variance of the models.

The prediction of a base learner m(x) is represented in m'(x). The expected prediction error (E) of model m is represented by $E[m'(x)]$ and the actual class values are denoted by $m(x)$. The bias of the model m is given by the difference between the model′s expected prediction error and its actual class values as given in (7). The variance of the model ($Var[m'(x)]$), is defined in (8). The total expected prediction error or MSE is given in (9), where $Var(\epsilon)$ gives the irreducible error that cannot be reduced from the model due to its noise variance.

$$Bias\left[m'(x)\right] = E\left[m'(x)\right] - m(x) \tag{7}$$

$$Var\left[m'(x)\right] = E\left[m'(x)^2\right] - E\left[m'(x)\right]^2 \tag{8}$$

$$E\left[(m(x) - m'(x))\,2\right] = \left(Bias\left[m'(x)\right]\right)^2 + Var\left[m'(x)\right] + Var(\epsilon). \tag{9}$$

The ensemble process of the models is defined such that, it minimizes the mean squared error of the summation of weighted models, as given in (10), where y_i denotes the predictions of model i and w_j is the corresponding weight to the model j. The weights are chosen to minimize the MSE, such that the total combination of $w_i \times y_i$ gives the minimum MSE, as given in (10). Min(.) derives the minimum MSE for different weight combinations.

$$\begin{aligned} &\text{Min}\left[MSE\left(w_1y_1 + w_2y_2 + \cdots + w_jy_j\right)\right] \\ &\text{s.t.} \quad \sum w_jy_j, w_j \geq 0, \quad \forall j = 1, \cdots, k. \end{aligned} \tag{10}$$

4.7 Data integration module

Integration of possible data types for the identification process helps to combine the important features provided by each data type and highlight their interrelationships with different functionalities. Therefore, the integrated system will provide higher throughput by getting insights into the data[59]. Therefore, in the ADHD-Care decision support system, the approach of an integrated module to classify ADHD when both fMRI and eye movement are considered for a patient. The prime goal of this module is to provide the classification results for a subject with a probability value indicating the potential of having the disorder given both fMRI and eye movement datasets.

In the data integration process, two separately generated models for fMRI and eye movement data are used for the classification process. The existing pre-processing and feature extraction methods are reused for fMRI and eye movement data separately. Since this integration module generates a combined result, a weighted approach is used to calculate the final prediction to lessen the bias of having two models contributing to the result. The weights required for these two datasets are pre-computed during the classification model generation for each dataset. The weight calculation uses metadata, i.e., the batch size, number of epochs, learning rate, model hyperparameters, and model accuracy.

The derivation process of the final prediction is given in Algorithm 1. The fMRI-Eye movement data integration module was used to derive the final prediction from the integrated fMRI and eye movement learning models. First, the previously developed fMRI and eye movement learning models were loaded to the system. Then both datasets were pre-processed, and relevant features were extracted. The extracted set of features for two datasets were used to predict the probability values of having ADHD, from each of the fMRI and eye-movement models. Then the weights for the models were loaded given the metadata as parameters. The weight calculation was carried out by a loss function named log loss or binary cross-entropy function, where the probability values of the predictions were accumulated and divided by the number of predictions (batch size). Hence, this function will return a higher weight for the lower loss function values and vice versa. Step 12) in Algorithm 1 shows the derivation of the final predictions by calculating the weighted sum of the two model predictions.

Steps 15) – 24) in Algorithm 1 gives the load_weights function. The weight is calculated by a logarithmic loss function related to the cross-entropy, named local_loss. The actual label and the predicted label of the test data are denoted by y and $p(y)$, respectively. The loss is calcu-

lated by adding the log probability of a given subject to be identified with ADHD with the log probability of the same person to be a non-ADHD subject. Then, the total loss is calculated by taking the average loss for the given batch size and the final weights are defined using the loss for an epoch. Thus, the probability values of the predictions are accumulated and divided by the number of predictions (batch size). Hence, this function will return a higher weight for the lower loss function values and vice versa.

Algorithm 1. fMRI−Eye movement data integration module

Dataset: fMRI data, eye movement data

Result: Probability value of having ADHD

1) function integration (fMRI_data, eye_movement_data){
2) fm = load_fmri_model();
3) em = load_eye_mov_model();
4) preproc_fmri = pre-process_fMRI(fMRI_data);
5) fmri_features=extract_fMRI_ features(prepocfmri);
6) fmri_prediction=fMRI_model.predict(fmri_features);
7) preproc_eye_mov=pre-process_fMRI(eye_movement_data);
8) eye_mov_features = extract_eye_mov_features(prepoc_eye_mov);
9) eye_mov_prediction=eye_mov_model.predict(eye_mov_features);
10) fmri_weight=load_weights(fm.batch_size, fm.no_of_epochs, fm.model_hyperparameters, fm.learning_rate, fm.model_accuracy);
11) eye_mov_weight = load_weights (em.batch_size, em.no_of_epochs, em.model_hyperparameters, em.learning_rate, em.model_accuracy);
12) final_prediction=((fmri_prediction × fmri_weight)+(eye_mov_prediction × eye_mov_weight))/(fmri_weight+eye_mov_weight);
13) return final_prediction;
14)}
15) function load_weights (batch_size, no_of_epochs, model_hyperparameters){
16) for i in range(no_of_epochs){
17) for j in range(batch_size){
18) y = model.hyperparameters[i][j]["label"]
19) local_loss $+=$ $y \times \log(\log(p(y)) + (1 - y) \times \log(\log(1 - p(y))$
20) }
21 loss = local_loss/batch_size
22) }
23) return loss/no_of_epochs;
24)}

4.8 Implementation tool stack

The implementation tool stack of the ADHD-Care web application consists of several technologies. The presentation layer consists of two modules, a classification manager, and a visualization manager. HTML, CSS, Javascript and jQuery are used to implement the classification manager module as those technologies are easy to maintain, update, and provide consistency in the design, while jQuery has a variety of plugins for specific needs. The health report generation which is the output of the web application was implemented using jsPDF and the html2canvas as these libraries provide functions to customize the reports.

The visualization manager module contains graphical information such as rating scale, informative and analytical graphs, which assist the user with detailed visual information. The business logic layer consists of three main modules for pre-processing, classification and evaluation.

Each module contains a separate tool stack. SPM12, Med2IMG, octave and functional connectivity toolbox for correlated and anticorrelated brain networks (CONN) are the main tools that were used for data pre-processing. The SPM12 tool is more versatile compared to other existing fMRI data pre-processors as it has many toolboxes, computational modelling and extensions. The classification module was based on the standard and prominent libraries such as pandas, NumPy and frameworks, i.e., Tensorflow, scikit-learn. The evaluation module was implemented using the scikit-learn and Keras libraries. Those python libraries provide an extensive set of features which can handle large dataset in a customizable and flexible manner.

The data access layer has a flask library to connect the front end and the back end of the ADHD-Care application. Flask is a micro-framework with minimal dependencies with external libraries, and flexible and extensible compared to the other libraries.

5 Key features of ADHD-Care

5.1 ADHD identification web application

The presented web application supports two separate classification functionalities for child and adult subjects, as the two dataset types, fMRI and eye movement data were derived from child and adult patients, respectively. After the login process, the users can select their age category and follow the proposed identification process to check the probability of them having ADHD. The fMRI data is represented in the NIfTI-1 data format (.nii), by neuroimaging informatics technology. The DSS comes under the menu item Services in the ADHD-Care web application[46].

When the raw fMRI images are added to the decision support system as the input for the identification process, the brain image passes through several pre-processing steps before the classification. The completion level of each stage in the pre-processing is displayed to the user as a step by step process in the web application. After

the pre-processing stage, the relevant connectivity features are extracted and passed to the prediction of ADHD possibility. The users can view the progress of the entire process until the classification results are displayed. The developed system relies on fMRI and eye movement data of the patients to identify ADHD and provide a severity rating.

Consequently, if a patient is identified as having a diagnosis of ADHD, then the severity level is also be displayed as high, low or medium via a rating scale rated from 0 to 1.0, indicating 0 being the lowest severity and 1.0 being the highest severity. The application provides a description of ADHD and users can view different blogs and relevant documents using the system to enrich their knowledge about ADHD, including its symptomology, current processes of identification, and possible treatments.

5.2 Report generation

When the user successfully submits the patient details along with the corresponding eye movement or fMRI data, the system automatically generates a health report. This report contains basic patient details and specific details on the disorder. ADHD-Care web application displays the health report by mentioning whether the ADHD positive or negative along with a rating score indicating the severity of the disorder for the subject. The user guide for the application is available under the menu item learn, in the ADHD-Care web application[46].

In order to support user-friendliness, a color scale is used to indicate the similarity score. For instance, if the person has a high probability of having ADHD, which gives a similarity score near to one, it indicates in red color on the color scale. If the person has a low possibility to have ADHD the scale portrays a similarity score near to zero. It indicates in blue color in the colored rating scale. Further, it gives a probability value for the predicted result. Thus, the generated health report improves usability by providing a convenient representation.

6 Evaluation and results

6.1 Accuracy result analysis

6.1.1 Evaluation measures

Several evaluation metrics including accuracy, sensitivity and specificity have been used to evaluate the classification process. The sensitivity metric also known as recall, measures the proportion of actual ADHD patients that the model can correctly predict from the actual ADHD population. While the specificity measures how well the model can predict the proportion of healthy subjects out of the actual healthy population[23,53]. The equations for sensitivity and specificity are given in (11) and (12), respectively. Accuracy is a measurement of how well the model identifies the patterns given the dataset, and predictions on the calibrated data. The accuracy is defined as stated in (13). We have used scikit-learn libraries to obtain these values.

$$Sensitivity = \frac{True\ positives\ (TP)}{True\ positives + False\ negatives\ (FN)} \tag{11}$$

$$Specificity = \frac{True\ negatives\ (TN)}{True\ negatives + False\ positives\ (FP)} \tag{12}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{13}$$

6.1.2 fMRI data processing results

We evaluated the accuracy of the fMRI data processing in the main DMN regions, generated by the CNN approach. The process can be summarized as follows. Initially, the training and testing data were randomly split to avoid the bias of the model by using 20% of the entire dataset for the validation, as given in the Pareto principle. Using 20% of the dataset as the validation portion is widely accepted and performed well in most of the practical scenarios. The seed-based correlations were extracted on selected DMN regions from the pre-processed fMRI dataset and the extracted correlation matrices were used for the classification process. Classifier models were generated for each region of the DMN areas using the CNN approach.

The correlations were converted into a series of 2D images, then fed into the CNN and obtained the classifier accuracies on the testing data sets. The number of layers and activation functions were adjusted to reduce the complexity of the model along with data augmentation to avoid the overfitting of CNN models. The obtained accuracies, specificity and sensitivity values for the classific-

Table 4 Evaluation metrics of DMN regions using fMRI data

Region	Accuracy (%)	Sensitivity (%)	Specificity (%)
MTL_L	88.12	76.89	85.51
HF_R	87.38	74.22	87.44
LTJ	85.50	72.23	86.14
MPC	85.21	72.80	84.12
PCC	84.84	71.12	82.35
RTJ	84.62	70.85	84.17
HF_L	82.37	69.12	83.52
IPC_L	51.12	41.11	76.74
MTL_R	50.50	40.45	75.81
IPC_R	50.13	49.53	74.96
Ensemble model	82.12	69.33	81.01

ation process of different DMN regions are given in Table 4.

The records of different regions are ordered based on the accuracy values. The classification process has considered different DMN regions such as left medial temporal lobe (MTL_L), right hippocampal formation (HF_R), left temporoparietal junction (LTJ), medial posterior cingulate (MPC), posterior cingulate (PCC), right temporoparietal junction (RTJ), left hippocampal formation (HF_L), left inferior parietal cortex (IPC_L), right medial temporal lobe (MTL_R), right inferior parietal cortex (IPC_R)[48].

We identify the regions with high accuracy and sensitivity values in the classification processes including MTL_L, HF_R, LTJ, MPC, PCC, RTJ and HF_L. However, some DMN regions such as IPC and MTL_R have shown low accuracy, sensitivity and specificity values by reflecting their less influence on the identification of ADHD.

Accordingly, by comparing the accuracies obtained from each model, the MTL_L region has shown the best accuracy level among other regions, in terms of ADHD classification. HF_R, MPC, LTJ regions have also shown closer accuracy levels to MTL_L. For instance, Kuang et al.[30] have shown accuracies of 34.39%, 37.04% and 37.42% for visual cortex (VC), cingulate cortex (CC) and prefrontal cortex (PC), respectively.

In Table 4, an overall accuracy of 82.12% along with 69.33% sensitivity and 81.01% specificity values were obtained by the ensemble model of all DMN regions. Here, the average accuracy value of fMRI data is 75.62% and the average value of sensitivity and specificity is 73.15%. Thus, we can see balanced evaluation values for fMRI data. The proposed ADHD-Care DSS is implemented using this ensemble model, as described in Section 4.6, where all the regions of DMN are contributed for the classification. Here, the regions with high classification accuracies are found to be more active in ADHD subjects. However, several studies have suggested considering the whole DMN region for a better choice, since this activity may vary per subject[60,61].

The same classification procedure was tested for other features for the fMRI data. fALFF and ReHo features were used with the CNN classifier to compare the seed-based approach on selected DMN areas with other feature values. The classifications with fALFF, ReHo features have given the accuracy of 67.17% and 67.30%, respectively[23]. The sensitivity and specificity rates were in the rage of 65–60%.

The evaluation of the proposed methodology in comparison with the related studies are described in Table 5. Considering the overall accuracy, sensitivity and specificity levels of all the DMN regions considered for this study, it can be concluded that the proposed method on seed correlation with CNN shows high-quality attributes in terms of accuracy, sensitivity, and specificity values compared to classification based on other features and related studies.

6.1.3 Eye movement data processing results

Different decision tree algorithms that have been used for ADHD classification with eye movement data were evaluated in terms of the accuracy as shown in Table 6. In the evaluation, the dataset was split into 2 : 1, where 66% of data are used for the training and remaining for the testing purposes, as that ratio is one of the commonly used ratios for a fair split. The metrics sensitivity and specificity values are considered to support a complete evaluation of the results.

Accordingly, an overall accuracy of 81% of the ensemble model is achieved. Random forest and J48 algorithms have performed the best in terms of the two categories of algorithms considered in this study, where the random forest classifier has the highest accuracy of 84.48% compared to the other considered classifiers. The other algorithms of REPTree, LMT, JRip, and decision stump have also performed well since they are mostly based on the decision tree approach. Random forest outperforms all of them because of the usage of a collection of trees to provide the prediction. Nevertheless, these decision tree algorithms are bound to overfit their models to the dataset. Hence, other commonly used classification algorithms like SVM and Naive Bayes have been able to perform less in terms of the classification. This may due to the higher bias and variance in data. Thus, mediator algorithms such as bagging and AdaBoost have been used, to reduce the variance, bias and the over fitness that caused by the decision tree algorithms.

In order to provide different aspects of the algorithms, an ensemble model is used to further reduce the over-fitness, bias and variance in the dataset. Moreover, accord-

Table 5 Comparison with other studies – Eye movement data

Related study	Features	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
[7]	Fixations	Bagging	78.48	78.00	–
[62]	Fixation and saccade	Tree-based classifiers	86%	–	–
[13]	Blink and saccade rates	Linear classifier	70.00	59.00	82.00
[11]	Fixations, saccades, pupil diameter	Random forest	85.31(10-folds)	–	–
			84.48(66% split)	–	–
Proposed ADHD-Care DSS	Fixations & Saccades	Ensemble model	81.02	79.56	76.89

Table 6 Evaluation metrics of eye-movement data classification

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
Random forest	84.48	81.11	79.50
REPTree	82.68	80.45	78.63
LMT	82.66	81.22	78.54
JRip	82.57	80.14	77.45
PART	82.11	80.20	77.68
J48	81.74	79.89	77.02
Bagging	81.24	79.90	76.85
Random tree	78.84	77.45	73.11
AdaBoost	73.46	71.85	69.55
Hoeffding tree	73.34	71.98	69.58
Decision stump	72.68	69.23	68.75
SVM	69.13	68.44	67.12
Naive Bayes	67.18	65.77	63.96
Ensembled model	81.02	79.56	76.89

ing to the results shown in Table 6, the average accuracy value of eye-movement data classification is 78.08% and the average value of sensitivity and specificity is 75.06%. Thus, we can see balanced evaluation values for eye-movement data.

Since this study is based on multi-classifiers, ROC curve is used to visualize the performance of the top classifiers within the given thresholds. Also, it supports the identification of the analytical ability of binary classification, considering the true-positive and false-positive rates[53]. The ROC curve which shows the trade-off between the sensitivity and the specificity shows the discriminatory ability of the classification model[11,54].

The ROC graph is shown in Fig. 11, and is generated for the ensembled model for the ADHD classification with eye movement data. The area under the ROC curve, which is 0.87 shows 87% of accuracy, indicates how accurate the prediction will be for future data using the generated ensembled model for the ADHD identification.

Further, the ROC graph of the decision tree and different classifier ruling algorithms is given in our previous work[11].

6.2 Overfitting avoidance analysis of the ensemble model

Overfitting of a machine learning model can be identified by the difference between error on the model's testing data set and error on its training data set. If the model prediction gives a higher error on the testing data set compared to the error on the training data set, it is known as a highly overfitted learning model[63]. The error value can be defined as the difference between the actual label and the model prediction for each point of data. By evaluating the percentages of training and testing accuracy values given by the generated ensemble models in this study, it has shown that there is less difference between training and testing errors than the models are considered separately.

For fMRI data, CNN models of the DMN regions such as MTL, LTJ, MPC, IPC, PCC are considered. For eye movement data classification, several learning models, e.g., random forest, Naive Bayes, AdaBoost, SVM are considered. The model training and testing accuracy were measured accordingly during and after generating the model and the prediction error was calculated after testing. The overfitting (error) values for the set of DMN regions in fMRI data and the classifiers for the eye-movement data are shown in Figs. 12 and 13, respectively.

The corresponding ensemble model has shown the minimum prediction error in each case. Thus, the ensemble model has avoided the overfitting of using one model, which is the main contribution of this study. The ensemble model of fMRI data has shown a training accuracy of 80.88% and testing accuracy of 82.12%. Similarly, the ensemble model of eye movement data has shown a training accuracy of 80.84% and testing accuracy of 81.02%.

Considering the training, testing accuracy values and

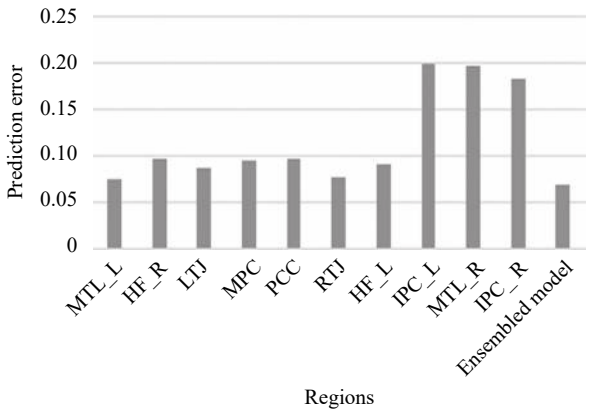

Fig. 12 Prediction error for fMRI data

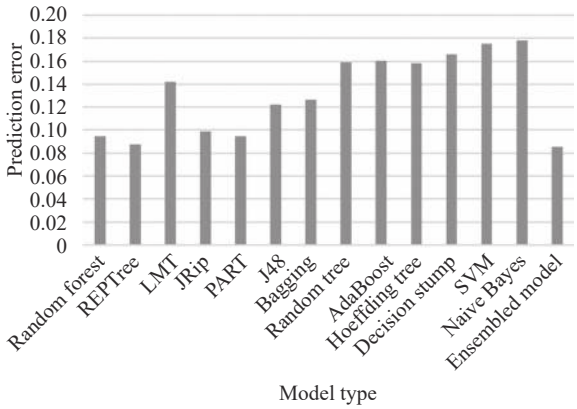

Fig. 13 Prediction error for eye-movement data

the prediction errors of both ensemble models, the proposed ensemble models reduce the prediction error compared to the other best-performing algorithms in both cases.

Further, the best performing model for fMRI data has an accuracy of 88.12% in testing, where the ensemble model has an accuracy of 80.88%, but the prediction error for the highest performing model is higher than the ensemble model. This confirms that the ensemble model for the fMRI classification has reduced the effect of overfitting in models. The same comparison can be derived for eye movement data classification as well. The highest performing algorithm of random forest with 84.48% testing accuracy has obtained a higher prediction error than the ensemble model constructed. This again confirms that the ensemble model for eye-movement classification has reduced the overfitting effects in the model generation and generalization for unseen data.

6.3 Usability measures

The usability of the application is tested using the system usability scale (SUS)[64], which is a reliable measure to assess the usability of a software application. It consists of 10 questions, including positive and negative aspects of the system. Each question is provided with a rating scale from 1 to 5, indicating strong disagreement to strong agreement. According to the scale specification, a SUS score of more than 68% is considered as above average.

The participant group consisted of 10 MRI laboratory members and physicians in local hospitals and 10 medical students, with knowledge and experience on fMRI, eye movement data, and their variations. The 20 subjects were chosen as 12 females and 8 males in the age group of 20–40, having an average of 3 years of experience in related aspects. During the SUS test, a task-list that covers the entire system features was given for each user to maintain the consistency of the study[44]. Thus, the core functionalities of the DSS, including ADHD identification, severity score computation, and report generation were tested separately for fMRI and eye movement data and their corresponding scores recorded.

The users were given the system and guided to perform an ADHD test for a given fMRI or eye movement data according to the task list. Then they were asked to work with the overall system. The results were reported to have above-average usability in both fMRI and eye movement identification. The obtained results for both the positive and negative aspects of the system are listed in Figs. 14 and 15, respectively. An average SUS score of 78.57% was observed for the fMRI data-based identification and 79.4% was obtained for eye movement data-based identification. As an overall system score, 70.2% was recorded.

Moreover, this study has used a tag cloud for further interpretation of system usability. During the usability test, the user was asked to select at least three characteristics of the system. Referring to these obtained scores and the feedback, it can be concluded that the proposed web application has high usability in practice. However, the system is not tested in actual clinical practice, due to the technical difficulties associated with the underlying infrastructure, and will be considered as a future extension.

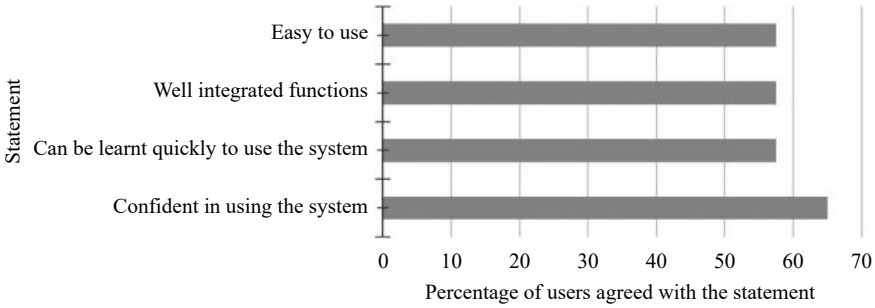

Fig. 14 SUS results for positive aspects of the system

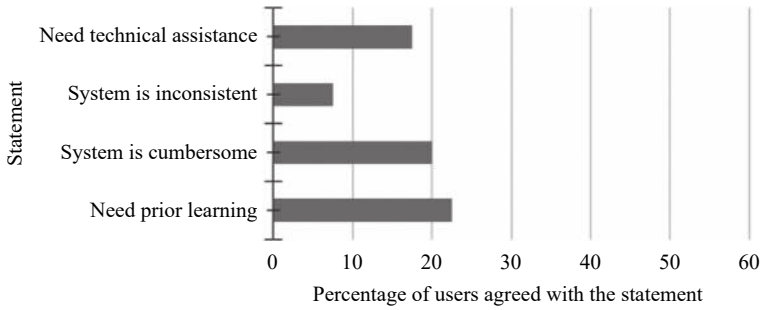

Fig. 15 SUS results for negative aspects of the system

7 Discussions

7.1 Main contributions of the proposed solution

The proposed ADHD-Care DSS has developed as two separate models in a single framework. Theoretically, fMRI data can be used for both adults and children. However, this study has used the fMRI dataset for children only, considering data availability. The ADHD classification among adults is mainly based on eye movement data, which was captured during a previous study[7]. Eye movement data for children are not available, as they cannot follow a set of defined activities to capture relevant data. Many existing studies on ADHD classification have focused on either child or adult identification separately. As per our knowledge, this is the first study that has focused on both child and adult ADHD classification using two different datasets. Thus, having an extendable system framework is one of the novelties of the proposed method. This resolves the constraint of bounding the data classification into one specific age group as in existing studies. Additionally, presenting classification results with a similarity rating scale is the main usefulness.

The classification process was applied to 10 different DMN regions to explore the usage of fMRI data in ADHD identification. These regions were selected based on the connectivity of DMN. For instance, the regions PCC, MPC, RTJ, LTJ are widely used in ADHD identification studies[5,36]. The regions such as MTL_L, MTL_R, HF_L, HF_R, IPC_L, IPC_R were selected due to their influence in connectivity[5]. Having a low classification accuracy of a brain region is considered as an indicator of having less influence on an ADHD pathology.

Additionally, the eye movement data exploration in the ADHD identification process for adults has produced high accuracy values. The data acquisition process has extracted the main features of fixations and saccades with their durations and other notable eye movement features such as pupil diameter. The standard learning models were used to develop the ensemble model that produce high accuracy, sensitivity and specificity values, by avoiding overfitting.

Further, the solution is developed by the serialized composition of a set of derived models with the novel web technology. An application programming interface (API) has developed to obtain results from each task in the processing pipeline that contains pre-processing, feature extraction and classification. Hence, the DSS can provide high-performance results in each of the pipeline tasks. This pipeline implementation was identified as a notable feature during the user survey with a maximum usability score, by analyzing the feature-wise usability test results. Also, the direct usage of neuroimaging files of fMRI data and the internal pre-processing capabilities are useful contributions of the proposed DSS.

Additionally, the use of web technologies embedding learning models in the systems has uplifted the novelty of the ADHD identification process. Since this web-based system is accessible over different platforms and geographical locations, the execution of ADHD-Care DSS in real scenarios is manageable. Further, the system usability test result verifies that the proposed system can be easily used by the practitioners, especially the psychiatrists.

7.2 Comparison of eye movement data classification with related studies

The comparison of the proposed eye movement classification results with the related studies is given in Table 5. The proposed method for the classification of ADHD using eye movement data has acquired better accuracy compared to other related studies as a result of using both fixation and saccade related gaze parameters with an ensemble classifier that associates several classification models according to their accuracy levels.

The proposed solution has generated an ensemble model by giving weights to each of the learning models that are proportional to their accuracy. Although the accuracy of the final ensemble model has reduced slightly, it has avoided overfitting. Thus, the ADHD-Care DSS has shown sufficient accuracy, specificity and sensitivity values with overfitting avoidance, compared to the related studies. Hence, this provides a reliable classification with a similarity score that indicates the possible severity of the disorder. Currently, there is a lack of related studies that have addressed the classification of ADHD based on different gaze related movements. Thus, the proposed solution further reflects the novel contribution, which uses fixation and saccade eye movement for the ADHD identification with high classification results.

7.3 Comparison of fMRI classification with related studies

The obtained results for the seed-based correlation for the main DMN regions of fMRI data are compared with the related work in Table 7. Many studies have used the ADHD-200 dataset to perform the research which compares the accuracies based on the same experimental setup. The classification accuracies in many related studies mainly depend on the selected dataset, e.g., New York University (NYU), Neuro, Oregon Health and Science University (OHSU), NeuroImage (NI), Peking and Pittsburgh, that is based on the participant data, e.g., age and gender. Also, the classification accuracy depends on the considered regions of the brain.

Compared to the other related studies, the proposed method has obtained acceptable accuracy, sensitivity and

Table 7 Comparison with other studies – fMRI data

Related study	fMRI dataset	Features	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
[40]	ADHD-200	Frequency features	Deep belief network (DBN)	OHSU 80.88	–	–
				NYU-PC 37.42	–	–
				NYU-CC 37.04	–	–
				NYU-VC 34.39	–	–
				Neuro 44.4	–	–
				Pittsburgh 55.56	–	–
				Whole-brain 44.63	–	–
[37]	ADHD-200	Deep Bayesian Network (Deep-BN) based features	SVM	NYU 64.70	43.90	68.80
				Peking-1 66.3	22.9	87.7
				KKI 59.0	55.6	83.0
[24]	ADHD-200	fLAFF, GM density	Multi-modality 3D CNN	69.15	–	–
[45]	Clinical Neuropsychological data	Test of variables of attention (TOVA), diagnosis-supported (DS)-ADHD	ANOVA, split-half method statistical analysis	TOVA 75.7	85.33	53.13
				DS-ADHD 77.57	84.72	62.86
[65]	ADHD-200	Brain region time-series signals	CNN	NYU 73.1	65.5	91.6
				NI 67.9	63.6	71.4
				Peking 62.7	48.1	79.1
[66]	ADHD-200	Feature pooling	4D CNN	71.30	73.20	69.70
[39]	Pre-processed ADHD-200	Stockwell transform, fuzzy entropies of time-frequency domain submatrices	SVM	96.68	100%	–
[6]	Pre-processed ADHD-200	Connectivity features	SVM, NB, KNN	86.00	–	–
[28]	ADHD-200 (NYU)	SBC with pre-processed fMRI	CNN	PCC 84.84	65.22	71.12
				LTJ 85.05	66.31	72.23
				RTJ 84.62	64.12	70.85
				MPC 85.21	66.41	72.80
[23]	ADHD-200 (NYU)	Connectivity features of fMRI	3D CNN	85.36	66.54	72.80
Proposed model	ADHD-200 (NYU)	Seed-based correlation	7 layered CNN	82.12	75.33	71.25

specificity values, by avoiding overfitting. The novel approach of using a seed-based correlation for DMN brain regions with the CNN model can be identified as the main reason for obtaining better results compared to other related works without considering the whole brain aspect. This study has mainly focused on DMN brain regions for the identification process. Thus, the most related and highly influenced set of features extracted from main DMN regions and the techniques that have been used to optimize the CNN model performance have led the proposed method to achieve higher accuracies.

As a summary, several related studies have considered fMRI data fed into different CNN architectures such as 3D CNN[24], and 4D CNN[66], supported by a separate CNN for the feature extraction. Moreover, the usage of SVM[6,37,39], DBN[40] for the classification have been identified as other related studies. Some of them have considered hand-crafted features like fALFF, gray matter (GM) density[24], while others have considered feature extraction using CNN or DBN[23,28,65, 66].

7.4 Future research directions

The proposed solution can be extended with the feature selection process based on different brain areas other than DMN. Also, the extracted time series from fMRI data during the feature extraction can be used to explore the patterns in ADHD subjects. The focus of the eye movement data can be extended to obtain features on ROIs, saccade speed derivations, etc.

Additionally, with an appropriate dataset, this approach can be extended to multi-class classification to identify different ADHD sub-types such as ADHD-Inattentive, ADHD-hyperactive/Impulsive, and ADHD-combined. This will help to improve the severity score values.

We have tested the current system for the classification of the data types separately, due to the unavailability of eye movement and fMRI data that belong to the same set of users. The integration of eye movement and fMRI data that belongs to the same set of user group can also be conducted as future research. This can be used to enhance the overall prediction accuracy and validate the classification accuracy of each data type.

Moreover, the learning model can be enhanced with emerging techniques such as deep learning, artificial intelligence, probabilistic programming, parallel processing[67–69]. Accordingly, the approach can be extended to a generic framework that supports the processing and classification of different related data types and learning models, which is another possible research direction[6]. Moreover, having new usability features such as batch processing for datasets will enhance the engagement of the research community in ADHD pathology. Further, the application can be improved to use in real clinical practices with the support of technology and finance aspects.

8 Conclusions

Applying research into the development is useful for the growth of biomedical informatics in clinical practice. ADHD is a common neurological disorder among children and requires early detection and treatment to mitigate negative long-term and short-term outcomes. The identification process of ADHD still relies on flawed and subjective methods where manual clinical approaches are applied. This motivated the development of tools supporting an automated decision support system. This system has aimed to achieve a systematic, valid, and reliable measure to classify ADHD. This was developed by classifying fMRI and eye movement data that result in high accuracy and sensitivity values.

This research addresses the identification of ADHD by constructing an approach using fMRI data for children and eye movement data for adults. In order to explore the usage of fMRI data in ADHD identification, the connectivity of DMN was studied. The overall classification accuracy obtained was 82.12% from the CNN approach for the ensemble model. Identification of the main DMN region coordinates providing high classification accuracies is one of the contributions of this study. Additionally, the usage of eye movement data in adult ADHD identification has been supported as successful and as another novel contribution. The combination of fixations and saccades in the classification process has yielded robust accuracy, sensitivity and specificity values verifying the applicability of eye movements in ADHD identification. The generated ensemble model for the eye-movement data classification has shown an overall accuracy of 81%. Additionally, the ensemble models were generated to avoid overfitting of the individual models and classifiers.

A DSS was developed using the generated models and further assisted by notable features such as severity measurement, report generation, and an integration module. The ADHD-Care DSS has achieved a system usability score of 70.2%. The combination of both fMRI and eye movement data for the classification process in the DSS reflects the originality of this study to derive accurate results for separate age groups with precise severity scores. Thus, with the state-of-art results from the model evaluations and usability testing, it can be concluded that ADHD-Care DSS can be used to provide an accurate, user-friendly experience for the research community and practitioners in clinical practices.

Acknowledgements

This work was supported by Old Dominion University, Norfolk, Virginia, USA and University of Moratuwa, Sri Lanka. We thank the participants of the system usability study.

References

[1] C. Sridhar, S. Bhat, U. R. Acharya, H. Adeli, G. M. Bairy. Diagnosis of attention deficit hyperactivity disorder using imaging and signal processing techniques. *Computers in Biology and Medicine*, vol. 88, pp. 93–99, 2017. DOI: 10.1016/j.compbiomed.2017.07.009.

[2] D. A. Meedeniya, I. D. Rubasinghe. A review of supportive computational approaches for neurological disorder identification. *Interdisciplinary Approaches to Altering Neurodevelopmental Disorders*, T. Wadhera, D. Kakkar, Eds., IGI Global, Chapter 16, Hershey, USA: IGI Global, pp.271–302, 2020. DOI: 10.4018/978-1-7998-3069-6.ch016.

[3] B. Zablotsky, L. I. Black, M. J. Maenner, L. A. Schieve, M. L. Danielson, R. H. Bitsko, S. J. Blumberg, M. D. Kogan, C. A. Boyle. Prevalence and trends of developmental disabilities among children in the United States: 2009–2017. *Pediatrics*, vol. 144, no. 4, Article number e20190811, 2019. DOI: 10.1542/peds.2019-0811.

[4] S. De Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna. A survey of attention deficit hyperactivity disorder identification using psychophysiological data. *International Journal of Online and Biomedical Engineering*, vol. 15, no. 13, pp. 61–76, 2019. DOI: 10.3991/ijoe.v15i13.10744.

[5] L. Q. Uddin, A. M. C. Kelly, B. B. Biswal, D. S. Margulies, Z. Shehzad, D. Shaw, M. Ghaffari, J. Rotrosen, L. A. Adler, F. X. Castellanos, M. P. Milham. Network homogeneity reveals decreased integrity of default-mode network in ADHD. *Journal of Neuroscience Methods*, vol. 169, no. 1, pp. 249–254, 2008. DOI: 10.1016/j.jneumeth.2007.11.031.

[6] I. D. Rubasinghe, D. A. Meedeniya. Automated neuroscience decision support framework. *Deep Learning Techniques for Biomedical and Health Informatics*, B. Agarwal, V. E. Balas, L. C. Jain, R. C. Poonia, Manisha, Eds., Cambridge, USA: Academic Press, pp.305–326, 2020. DOI: 10.1016/b978-0-12-819061-6.00013-6.

[7] A. M. P. Michalek, G. Jayawardena, S. Jayarathna. Predicting ADHD using eye gaze metrics indexing working memory capacity. *Computational Models for Biomedical Reasoning and Problem Solving*, C. H. Chen, S. C. S. Ch-

eung, Eds., Hershey: IGI Global, Chapter 3, pp.66–88, 2019. DOI: 10.4018/978-1-5225-7467-5.ch003.

[8] N. N. J. Rommelse, S. Van Der Stigchel, J. A. Sergeant. A review on eye movement studies in childhood and adolescent psychiatry. *Brain and Cognition*, vol. 68, no. 3, pp. 391–414, 2008. DOI: 10.1016/j.bandc.2008.08.025.

[9] P. Deans, L. O'Laughlin, B. Brubaker, N. Gay, D. Krug. Use of eye movement tracking in the differential diagnosis of attention deficit hyperactivity disorder (ADHD) and reading disability. *Psychology*, vol. 1, no. 4, pp. 238–246, 2010. DOI: 10.4236/psych.2010.14032.

[10] S. Van Der Stigchel, M. Meeter, J. Theeuwes. Eye movement trajectories and what they tell us. *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 5, pp. 666–679, 2006. DOI: 10.1016/j.neubiorev.2005.12.001.

[11] S. De Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna, A. M. P. Michalek, G. Jayawardena. A rule-based system for ADHD identification using eye movement data. In *Proceedings of Moratuwa Engineering Research Conference*, IEEE, Moratuwa, Sri Lanka, pp.538–543, 2019. DOI: 10.1109/mercon.2019.8818865.

[12] K. Krejtz, A. T. Duchowski, A. Niedzielska, C. Biele, I. Krejtz. Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS One*, vol. 13, no. 9, Article number e0203629, 2018. DOI: 10.1371/journal.pone.0203629.

[13] M. Fried, E. Tsitsiashvili, Y. S. Bonneh, A. Sterkin, T. Wygnanski-Jaffe, T. Epstein, U. Polat. ADHD subjects fail to suppress eye blinks and microsaccades while anticipating visual stimuli but recover with medication. *Vision Research*, vol. 101, pp. 62–72, 2014. DOI: 10.1016/j.visres.2014.05.004.

[14] R. G. Ross, J. G. Harris, A. Olincy, A. Radant. Eye movement task measures inhibition and spatial working memory in adults with schizophrenia, ADHD, and a normal comparison group. *Psychiatry Research*, vol. 95, no. 1, pp. 35–42, 2000. DOI: 10.1016/s0165-1781(00)00153-0.

[15] R. G. Ross, A. Olincy, J. G. Harris, B. Sullivan, A. Radant. Smooth pursuit eye movements in schizophrenia and attentional dysfunction: Adults with schizophrenia, ADHD, and a normal comparison group. *Biological Psychiatry*, vol. 48, no. 3, pp. 197–203, 2000. DOI: 10.1016/s0006-3223(00)00825-8.

[16] D. P. Munoz, I. T. Armstrong, K. A. Hampton, K. D. Moore. Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder. *Journal of Neurophysiology*, vol. 90, no. 1, pp. 503–514, 2003. DOI: 10.1152/jn.00192.2003.

[17] G. J. Hyun, J. W. Park, J. H. Kim, K. J. Min, Y. S. Lee, S. M. Kim, D. H. Han. Visuospatial working memory assessment using a digital tablet in adolescents with attention deficit hyperactivity disorder. *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 137–143, 2018. DOI: 10.1016/j.cmpb.2018.01.022.

[18] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal. *Data Mining*: *Practical Machine Learning Tools and Techniques*, 4th ed., San Francisco, USA: Morgan Kaufmann, 2016.

[19] M. Kantardzic. *Data Mining*: *Concepts*, *Models*, *Methods*, *and Algorithms*, 3rd ed., Hoboken, USA: John Wiley & Sons, 2019.

[20] Y. H. Shi, W. M. Zeng, N. Z. Wang, D. T. L. Chen. A novel fMRI group data analysis method based on data-driven reference extracting from group subjects. *Computer Methods and Programs in Biomedicine*, vol. 122, no. 3, pp. 362–371, 2015. DOI: 10.1016/j.cmpb.2015.09.002.

[21] NITRC: ADHD-200, 2019, [Online], Available: https://www.nitrc.org/ir/app/template/, May 2, 2020.

[22] B. Jie, M. X. Liu, D. G. Shen. Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease. *Medical Image Analysis*, vol. 47, pp. 81–94, 2018. DOI: 10.1016/j.media.2018.03.013.

[23] S. De Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna. fMRI feature extraction model for ADHD classification using convolutional neural network. *International Journal of E-Health and Medical Communications*, vol.12, no.1, pp.81–105, 2021. DOI: 10.4018/IJEHMC.2021010106.

[24] L. Zou, J. N. Zheng, C. Y. Miao, M. Mckeown, Z. J. Wang. 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. *IEEE Access*, vol. 5, pp. 23626–23636, 2017. DOI: 10.1109/access.2017.2762703.

[25] V. Subbaraju, M. B. Suresh, S. Sundaram, S. Narasimhan. Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach. *Medical Image Analysis*, vol. 35, pp. 375–389, 2017. DOI: 10.1016/j.media.2016.08.003.

[26] V. Subbaraju, S. Sundaram, S. Narasimhan. Identification of lateralized compensatory neural activities within the social brain due to autism spectrum disorder in adolescent males. *European Journal of Neuroscience*, vol. 47, no. 6, pp. 631–642, 2018. DOI: 10.1111/ejn.13634.

[27] H. Dhayne, R. Haque, R. Kilany, Y. Taher. In search of big medical data integration solutions − A comprehensive survey. *IEEE Access*, vol. 7, pp. 91265–912900, 2019. DOI: 10.1109/ACCESS.2019.2927491.

[28] G. Ariyarathne, S. De Silva, S. Dayarathna, D. Meedeniya, S. Jayarathne. ADHD identification using convolutional neural network with seed-based approach for fMRI data. In *Proceedings of the 9th International Conference on Software and Computer Applications*, ACM, Langkawi, Malaysia, pp.31–35, 2020. DOI: 10.1145/3384544.3384552.

[29] X. L. Peng, P. Lin, T. S. Zhang, J. Wang. Extreme learning machine-based classification of ADHD using brain structural MRI data. *PLoS One*, vol. 8, no. 11, Article number e79476, 2013. DOI: 10.1371/journal.pone.0079476.

[30] D. P. Kuang, X. J. Guo, X. An, Y. L. Zhao, L. H. He. Discrimination of ADHD based on fMRI data with deep belief network. *Intelligent Computing in Bioinformatics*, D. S. Huang, K. Han, M. Gromiha, Eds., Cham, Switzerland: Springer, pp.225–232, 2014. DOI: 10.1007/978-3-319-09330-7_27.

[31] F. X. Castellanos, D. S. Margulies, C. Kelly, L. Q. Uddin, M. Ghaffari, A. Kirsch, D. Shaw, Z. Shehzad, A. Di Martino, B. Biswal, E. J. S. Sonuga-Barke, J. Rotrosen, L. A. Adler, M. P. Milham. Cingulate-precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biological Psychiatry*, vol. 63, no. 3, pp. 332–337, 2008. DOI: 10.1016/j.biopsych.2007.06.025.

[32] C. Fassbender, H. Zhang, W. M. Buzy, C. R. Cortes, D. Mizuiri, L. Beckett, J. B. Schweitzer. A lack of default network suppression is linked to increased distractibility in ADHD. *Brain Research*, vol. 1273, pp. 114–128, 2009. DOI: 10.1016/j.brainres.2009.02.070.

[33] A. M. S. Aradhya, A. Joglekar, S. Suresh, M. Pratama.

Deep transformation method for discriminant analysis of multi-channel resting state fMRI. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, AAAI, Hawaii, USA, pp.2556–2563, 2019. DOI: 10.1609/aaai.v33i01.33012556.

[34] A. M. S. Aradhya, V. Subbaraju, S. Sundaram and N. Sundararajan. Regularized spatial filtering method (R-SFM) for detection of attention deficit hyperactivity disorder (ADHD) from resting-state functional magnetic resonance imaging (rs-fMRI). In *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, Honolulu, USA, USA, pp.5541–5544, 2018. DOI: 10.1109/embc.2018.8513522.

[35] K. Konrad, S. B. Eickhoff. Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder *Human Brain Mapping*, vol. 31, no. 6, pp. 904–916, 2010. DOI: 10.1002/hbm.21058.

[36] I. A. Strigo, S. C. Matthews, A. N. Simmons. Decreased frontal regulation during pain anticipation in unmedicated subjects with major depressive disorder. *Translational Psychiatry*, vol. 3, no. 3, Article number e239, 2013. DOI: 10.1038/tp.2013.15.

[37] A. J. Hao, B. L. Ha, C. H. Yin. Discrimination of ADHD children based on deep Bayesian network. In *Proceedings of IET International Conference on Biomedical Image and Signal Processing*, IEEE, Beijing, China, 2015. DOI: 10.1049/cp.2015.0764.

[38] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Muller, G. Candrian. Machine learning approach for classification of ADHD adults. *International Journal of Psychophysiology*, vol. 93, no. 1, pp. 162–166, 2014. DOI: 10.1016/j.ijpsycho.2013.01.008.

[39] S. Sartipi, H. Kalbkhani, P. Ghasemzadeh, M. G. Shayesteh. Stockwell transform of time-series of fMRI data for diagnoses of attention deficit hyperactive disorder. *Applied Soft Computing*, vol. 86, Article number 105905, 2020. DOI: 10.1016/j.asoc.2019.105905.

[40] D. P. Kuang, L. H. He. Classification on ADHD with deep learning. In *Proceedings of International Conference on Cloud Computing and Big Data*, IEEE, Wuhan, China, pp.27–32, 2014. DOI: 10.1109/ccbd.2014.42.

[41] G. Deshpande, P. Wang, D. Rangaprakash, B. Wilamowski. Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE Transactions on Cybernetics*, vol. 45, no. 12, pp. 2668–2679, 2015. DOI: 10.1109/tcyb.2014.2379621.

[42] G. Brihadiswaran, D. Haputhanthri, S. Gunathilaka, D. Meedeniya, S. Jayarathna. EEG-based processing and classification methodologies for autism spectrum disorder: A review. *Journal of Computer Science*, vol. 15, no. 8, pp. 1161–1183, 2019. DOI: 10.3844/jcssp.2019.1161.1183.

[43] V. Sachnev, S. Suresh, N. Sundararajan, B. S. Mahanand, M. W. Azeem, S. Saraswathi. Multi-region risk-sensitive cognitive ensembler for accurate detection of attention-Deficit/Hyperactivity disorder. *Cognitive Computation*, vol. 11, no. 4, pp. 545–559, 2019. DOI: 10.1007/s12559-019-09636-0.

[44] M. Delavarian, F. Towhidkhah, P. Dibajnia, S. Gharibzadeh. Designing a decision support system for distinguishing ADHD from similar children behavioral disorders. *Journal of Medical Systems*, vol. 36, no. 3, pp. 1335–1343, 2010. DOI: 10.1007/s10916-010-9594-9.

[45] K. C. Chu, Y. S. Huang, C. F. Tseng, H. J. Huang, C. H. Wang, H. Y. Tai. Reliability and validity of DS-ADHD: A decision support system on attention deficit hyperactivity disorders. *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 241–248, 2017. DOI: 10.1016/j.cmpb.2016.12.003.

[46] ADHD-Care, 2019, [Online], Available: http://blooming-sands-73478.herokuapp.com, May 2, 2020.

[47] T. J. Andrews, S. D. Halpern, D. Purves. Correlated size variations in human visual cortex, lateral geniculate nucleus, and optic tract. *The Journal of Neuroscience*, vol. 17, no. 8, pp. 2859–2868, 1997. DOI: 10.1523/jneurosci.17-08-02859.1997.

[48] X. Nie, Y. Shao, S. Y. Liu, H. J. Li, A. L. Wan, S. Nie, D. C. Peng, X. J. Dai. Functional connectivity of paired default mode network subregions in primary insomnia. *Neuropsychiatric Disease and Treatment*, vol. 11, pp. 3085–3093, 2015. DOI: 10.2147/ndt.s95224.

[49] V. K. Ha, J. C. Ren, X. Y. Xu, S. Zhao, G. Xie, V. Masero, A. Hussain. Deep learning based single image super-resolution: A survey. *International Journal of Automation and Computing*, vol. 16, no. 4, pp. 413–426, 2019. DOI: 10.1007/s11633-019-1183-x.

[50] T. Honderich. *The Oxford Companion to Philosophy*, 2nd ed., Oxford, UK: Oxford University Press, 2005.

[51] I. Zaidi, M. Chtourou, M. Djemel. Robust neural control of discrete time uncertain nonlinear systems using sliding mode backpropagation training algorithm. *International Journal of Automation and Computing*, vol. 16, no. 2, pp. 213–225, 2017. DOI: 10.1007/s11633-017-1062-2.

[52] S. Ruder. An overview of gradient descent optimization algorithms, [Online], Available: https://arxiv.org/abs/1609.04747, May 2, 2020.

[53] X. H. Zhou, N. A. Obuchowski, D. K. McClish. *Statistical Methods in Diagnostic Medicine*, 2nd ed., Hoboken, USA: Wiley, 2011.

[54] M. H. Zweig, G. Campbell. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, vol. 39, no. 4, pp. 561–577, 1993. DOI: 10.1093/clinchem/39.4.561.

[55] I. Unal. Defining an optimal cut-point value in ROC analysis: An alternative approach. *Computational and Mathematical Methods in Medicine*, vol. 2017, Article number 3762651, 2017. DOI: 10.1155/2017/3762651.

[56] K. H. Zou, C. R. Yu, K. Z. Liu, M. O. Carlsson, J. Cabrera. Optimal thresholds by maximizing or minimizing various metrics via ROC-type analysis. *Academic Radiology*, vol. 20, no. 7, pp. 807–815, 2013. DOI: 10.1016/j.acra.2013.02.004.

[57] R. Fluss, D. Faraggi, B. Reiser. Estimation of the Youden index and its associated cutoff point. *Biometrical Journal*, vol. 47, no. 4, pp. 458–472, 2005. DOI: 10.1002/bimj.200410135.

[58] N. J. Perkins, E. F. Schisterman. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology*, vol. 163, no. 7, pp. 670–675, 2006. DOI: 10.1093/aje/kwj063.

[59] I. Subramanian, S. Verma, S. Kumar, A. Jere, K. Anamika. Multi-omics data integration, interpretation, and its application. *Bioinformatics and Biology Insights*, vol. 14, pp. 1–24, 2020. DOI: 10.1177/1177932219899051.

[60] S. V. Faraone, P. Asherson, T. Banaschewski, J. Bieder-

man, J. K. Buitelaar, J. A. Ramos-Quiroga, L. A. Rohde, E. J. S. Sonuga-Barke, R. Tannock, B. Franke. Attention-deficit/hyperactivity disorder. *Nature Reviews Disease Primers*, vol. 1, Article number 15020, 2015. DOI: 10.1038/nrdp.2015.20.

[61] E. Hoekzema, S. Carmona, J. A. Ramos-Quiroga, V. Richarte Fernandez, R. Bosch, J. C. Soliva, M. Rovira, A. Bulbena, A. Tobena, M. Casas, O. Vilarroya. An independent components and functional connectivity analysis of resting state fMRI data points to neural network dysregulation in adult ADHD. *Human Brain Mapping*, vol. 35, no. 4, pp. 1261–1272, 2014. DOI: 10.1002/hbm.22250.

[62] Z. Y. Mao, Y. Su, G. Q. Xu, X. P. Wang, Y. Huang, W. H. Yue, L. Sun, N. X. Xiong. Spatio-temporal deep learning method for ADHD fMRI classification. *Information Sciences*, vol. 499, pp. 1–11, 2019. DOI: 10.1016/j.ins.2019.05.043.

[63] Q. Xu, M. Zhang, Z. H. Gu, G. Pan. Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs. *Neurocomputing*, vol. 328, pp. 69–74, 2019. DOI: 10.1016/j.neucom.2018.03.080.

[64] J. Sauro. Measuring usability with the system usability scale (SUS): A practical guide to measuring usability, 2019, [Online], Available: https://measuringu.com/sus/, 4, 2019.

[65] A. Riaz, M. Asad, S. M. M. R. Al Arif, E. Alonso, D. Dima, P. Corr, G. Slabaugh. Deep fMRI: AN end-to-end deep network for classification of fMRI data. In *Proceedings of the 15th International Symposium on Biomedical Imaging*, IEEE, Washington, USA, pp.1419−1422, 2018. DOI: 10.1109/isbi.2018.8363838.

[66] G. Jayawardena, A. Michalek, S. Jayarathna. Eye tracking area of interest in the context of working memory capacity tasks. In *Proceedings of the 20th International Conference on Information Reuse and Integration for Data Science*, IEEE, Los Angeles, USA, pp.208−215, 2019. DOI: 10.1109/iri.2019.00042.

[67] I. Rubasinghe, D. Meedeniya. Ultrasound nerve segmentation using deep probabilistic programming. *Journal of ICT Research and Applications*, vol. 13, no. 3, pp. 241–256, 2019. DOI: 10.5614/itbj.ict.res.appl.2019.13.3.5.

[68] Z. J. Yao, J. Bi, Y. X. Chen. Applying deep learning to individual and community health monitoring data: A survey. *International Journal of Automation and Computing*, vol. 15, no. 6, pp. 643–655, 2018. DOI: 10.1007/s11633-018-1136-9.

[69] D. Haputhanthri, G. Brihadiswaran, S. Gunathilaka, D. Meedeniya, S. Jayarathna, M. Jaime, C. Harshaw. Integration of facial thermography in EEG-based classification of ASD. *International Journal of Automation and Computing*, to be published. vol.17, no. 6, pp. 837−854 DOI: 10.1007/s11633-020-1231-6.

Senuri De Silva is a bachelor student in computer science and engineering at Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka.

Her research interests include biomedical, machine learning and data mining.

E-mail: ktsdesilva.15@cse.mrt.ac.lk
ORCID iD: 0000-0002-4526-9586

Sanuwani Dayarathna is a bachelor student in computer science and engineering at Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka.

Her research interests include biomedical, machine learning and data mining.

E-mail:sanuwaniudara.15@cse.mrt.ac.lk
ORCID iD: 0000-0001-8669-2893

Gangani Ariyarathne is a bachelor student in computer science and engineering at Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka.

Her research interests include machine learning, data mining, biomedical, and computer security.

E-mail: gcariyarathne.15@cse.mrt.ac.lk
ORCID iD: 0000-0003-4205-6574

Dulani Meedeniya received the Ph. D. degree in computer science from University of St Andrews, UK in 2013. She is a senior lecturer in Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. She is a Fellow of Higher Education Academy, UK, Member of The Institution of Engineering and Technology, Member of Institute of Electrical and Electronics Engineers, and a Charted Engineer registered at Engineering Council, UK.

Her research interests include software modelling and design, workflow tool support for bioinformatics, data visualization and recommender systems.

E-mail: dulanim@cse.mrt.ac.lk (Corresponding author)
ORCID iD: 0000-0002-4520-3819

Sampath Jayarathna received the Ph. D. degree in computer science from the Texas A&M University – College Station, USA in 2016. He is an assistant professor in computer science at Old Dominion University, USA, where he is associated with Web Science and Digital Libraries (WS-DL) research group. He is a member of Association for Computing Machinery (ACM), IEEE, and Scientific Research Honor Society (Sigma XI).

His research interests include machine learning, information retrieval, data science, eye tracking, and brain-computer interfacing.

E-mail: sampath@cs.odu.edu
ORCID iD: 0000-0002-4879-7309

Anne M. P. Michalek received the Ph. D. degree in special education from Old Dominion University, USA in 2012. She is an associate professor of communication disorders and special education at Old Dominion University, USA.

Her research interests include multi-disciplinary approaches using biomedical technologies and instructional tools to improve outcomes for at-risk students and students with ADHD and autism spectrum disorder (ASD).

E-mail: aperrott@odu.edu
ORCID iD: 0000-0001-6850-3948