FOCUS: Detecting ADHD Patients by an EEG-Based Serious Game

Alaa Eddin Alchalabi, Shervin Shirmohammadi, *Fellow, IEEE*, Amer Nour Eddin, and Mohamed Elsharnouby

***Abstract*— Attention deficit hyperactivity disorder (ADHD), categorized by the lack of attention and focus, is one of the most common cognitive disorders. Since electroencephalogram (EEG) signals carry wide-ranging insights about cognition skills, the potential of using EEG signals to detect ADHD has a significant potential. EEG can be recorded utilizing wireless EEG reading devices often used by brain–computer interface researchers. In parallel-to-affordable EEG devices, serious games have been recently employed in the rehabilitation of multiple cognitive deficits. In this paper, we put the two things together, and we investigate the integration of an EEG-controlled serious game that trains and strengthens patients' attention ability while using machine learning to detect their attention level. Our pilot experiments with healthy individuals show an accuracy of up to 96% in classifying the EEG data to detect the correct game control type during gameplay, while our extended experiments with ADHD patients show an accuracy of up to 98% with a standard uncertainty of 0.16% in detecting ADHD patients.**

***Index Terms*— Attention deficit hyperactivity disorder (ADHD) detection, brain–computer interface, brain-controlled games, electroencephalogram (EEG) classification, serious games.**

I. INTRODUCTION

IN THE highly cited survey by Rego *et al.* [1], attention is considered one of the cognitive abilities alongside concentration, problem-solving, judgment, and language. Attention deficit hyperactivity disorder (ADHD) is a mental disorder, that is, identified with levels of inattention, impulsivity, and hyperactivity [2]. It mostly occurs in children, while symptoms may take effect in later adultery stages too. A survey conducted by the Centers for Disease Control and Prevention, Atlanta, GA, USA, showed that 11% of children (4–17 years old) in the USA were diagnosed with ADHD as of 2011 [3]. ADHD's presence in children leads to hindering academic achievement and social interactions.

Cognitive training is a commonly used therapy for ADHD patients. Cognitive training with different durations and intensities can lead to an increase in cognitive abilities, including attention. However, the main challenges with such training are illuminating the repetitive scenes, stimulating the users' interests, encouraging more engagement, and adjusting the difficulty level. While state-of-the-art treatments for people suffering from inattentive targets enhancing attentive skills, they lack the motivational aspect. With children, especially there is an existing dilemma of targeting the attentiveness level in therapies while avoiding distraction.

With the introduction of serious games (games used for nonentertainment purposes) in the cognitive therapies field, reinforcing users' motivation and engagement are no longer an obstacle due to the ability of accommodating different spans of ages and targeting various cognitive impediments. Using a serious game, one could engage the patient with an entertaining game while measuring his or her brain activity through an electroencephalogram (EEG) recording device. In addition, the recently introduced wireless EEG recording devices have become more affordable and available on a mass scale. But since the amount of information in EEG signals is vast, it will be too cumbersome for human processing to detect abnormalities.

This is where machine learning (ML) can become useful. The recent advances in artificial intelligence and ML, which nowadays are faster and more accurate in classifying data, can be leveraged to detect ADHD patterns from EEG data. In fact, recent research uses ML and statistical techniques to detect several neurological diseases, such as, epilepsy and becker muscular dystrophy [4], and psychological disorders, such as ADHD [4]–[6], schizophrenia [7], and autism [6].

The works which use ML to detect ADHD can be divided into those that use the EEG signal [8], [9] and those which do not [6], [10], [11]. Our work is in the former category; however, it is the only one that uses serious games to more easily engage a patient and acquire the EEG signal. Since patients, especially children, like to play games, using a serious game will significantly reduce the "white coat effect" which hinders detections in existing pen-and-paper screening tests [12], especially for children who might perceive the doctor as a punisher [13]. The game atmosphere eases that helping the detection process. In addition, the serious game will also allow us to remove the "training effect" by having a training phase before actual measurements. These two factors together will enable us to make measurements over more realistic and unbiased data, leading to more accurate detections. In that sense, our work is unique.

To the best of our knowledge, this is the first work that has laid the groundwork for integrating an ML classifier with a serious game to detect ADHD. In addition, as will be shown in Section II, this is the first work in the instrumentation and

Manuscript received March 22, 2018; accepted May 06, 2018. Date of publication June 1, 2018; date of current version June 7, 2018. This work was supported by the European ERASMUS+ Project titled "Intelligent Serious Games for Social and Cognitive Competence" under Project KA201-022247. The Associate Editor coordinating the review process was Dr. Domenico Grimaldi. *(Corresponding author: Alaa Eddin Alchalabi.)*

A. E. Alchalabi and S. Shirmohammadi are with the Multimedia Systems Lab, Istanbul Sehir University, Istanbul 34865, Turkey, and also with the Distributed and Collaborative Virtual Environment Research Laboratory, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: aalch040@uottawa.ca; shervin@discover.uottawa.ca).

A. N. Eddin and M. Elsharnouby are with the Multimedia Systems Lab, Istanbul Sehir University, Istanbul 34865, Turkey (e-mail: amernoureddin@std.sehir.edu.tr; mohamedelsharnouby@std.sehir.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2018.2838158

Fig. 1. EMOTIV EPOC + EEG reader.

measurement literature that has attempted to measure attention to detect ADHD. Our work can give an accuracy of up to 98% with a standard uncertainty of 0.16%, which is higher than existing works in using ML to detect ADHD, the best of which gives an accuracy of 95.6%. It must be mentioned that our serious game, called FOCUS, has been developed within the framework of the European ERASMUS+ Project titled "Intelligent Serious Games for Social and Cognitive Competence" [14].

The rest of this paper is organized as follows. Section II will present the related works, while Section III describes the game, our data collection methodology, and data preprocessing. Section IV will illustrate the proposed classification models, experiments, and results. Finally, the paper is concluded in Section V where future work opportunities are also presented.

II. Related Work

Neurofeedback was originated in the late 1960s [15] and has been utilized to train the ability of self-control via a real-time analysis of EEG brain signals, magnetoencephalography, and real-time functional magnetic resonance imaging [16]. Various types of therapeutic enhancements have been recorded after neurofeedback training (NFT). NFT could be implemented in association with a normalization of the quantitative EEG frequency ranges, which is a protocol used in signal processing of EEG signals [17]. EMOTIV EPOC+, a 14-channel wireless EEG system by EMOTIV Inc., San Francisco, CA, USA, is designed for brain–computer interface-related research. The device can record EEG frequency bands' data via open-source software provided by the company [18]. Fig. 1 shows the EMOTIV EPOC+ headset in action.

The EEG waves are generally divided into multiple frequency bands: δ-band <4 Hz, θ-band 4–7 Hz, α-band 8–12 Hz, β-band 12–30 Hz, and γ-band >30 Hz. Using EMOTIV, these frequency bands are generated by 0.5-s time step and 2 s of data window size. Each frequency band is associated with a unique brain function: δ-band is dominant in children during sleeping and is related to linguistic acquisition [19], θ-band is predominant in EEG during drowsiness states, α-band is important in relaxation, β-band is linked to fast activities, and γ-band is related to problem solving and memory [20].

Using multivariate analyses and advanced studies of EEG signal generators, EEG remains a strong candidate for a spot in the clinical setting, depending on continued efforts to capture additional sources of heterogeneity in ADHD [21].

A study aimed at using quantified EEG data to analyze subjects learning status gave optimum accuracy when all the five frequency bands' features were used; however, there was a difference on influencing the classification accuracy by each of the features [5]. The δ-band is shown to have the most significant effect on the accuracy by up to 6%. The results indicate that the EEG signals of attention are quite easier to detect compared with those of inattention.

Another study proved that during attentive states, subjects with ADHD are characterized by an under EEG activated state with noticeable specific differences [22]. Findings provided a rationale for applying NFT protocols aiming to study θ activity and θ/β ratio in children with ADHD to achieve a better attentive state.

Another study on ADHD recorded a growth of front-central θ-band activity and increased θ-to-β (θ/β) power ratio during rest compared to non-ADHD [23]. This study showed that the intensive analysis EEG data can describe the neurophysiological differences between ADHD and non-ADHD subjects, which can lead to more accurate diagnostic measures and effective NFT approaches for ADHD patients.

In the literature, a few recent research papers have reported their methodologies and results in detecting ADHD. Most of these works have utilized ML techniques on EEG data and others have used MRI data, while none have used a serious game. Duda *et al.* [6] used data from surveys given to the parents of ADHD and autism spectrum disorder (ASD) children that mainly focus on the child's behavior. This study utilized the acquired data and built classification models (logistic regression, support vector machine (SVM), linear discriminant analysis, and categorical Lasso) that aimed to distinguish ADHD and ASD, reporting accuracies of up to 96.5%. Peng *et al.*[10] used the MRI data in detecting ADHD patients using extreme learning machine, linear SVM, and radial basis function (RBF)-SVM, and it was able to correctly classify ADHD by: 90.18%, 84.73%, and 86.55%, respectively.

Ahmadlou and Adeli [8] recorded the EEG signals of ADHD and non-ADHD subjects with eyes closed. The data were classified correctly with an accuracy of 95.6% using an RBF neural network classifier. Similarly, Abibullaev and An [9] recorded the EEG signals of ADHD and non-ADHD patients during a cognitive task, and then used the data to correctly classify 95.4% of the data using an RBF-SVM classifier.

There has been much work done on EEG in the instrumentation and measurement community, but none have focused on attention measurement for ADHD. For example, detection and analysis of seizure were performed using EEG [24], [25]. Ahmad *et al.*[26] utilizes SVM classifiers and develops a model to recognize the cognitive and resting state of the brain. As a final example, Lay-Ekuakille *et al.* [27] introduced an implantable microapparatus encompassed under the scalp aims at monitoring and retrieving electrical cerebral

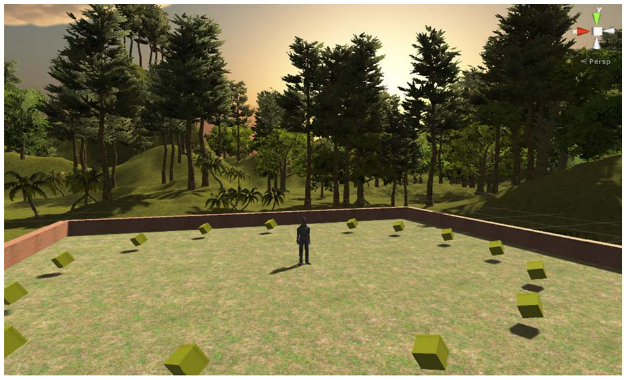

Fig. 2. FOCUS game [31].

activities. In addition, a few studies have employed EEG in several medical applications. A recent study introduced a novel method for excluding muscle artifacts in EEG data based on the independent vector analysis [28]. Another older study utilized multiscale entropy and autoregressive models in order to propose an automatic sleep-scoring using a single-channel EEG [29]. In addition, another study employed EEG in developing a portable wireless closed-loop seizure controller, and obtained commendable results [30]. Our work adds to this body of the literature by presenting a measurement approach to detect ADHD.

III. Game and Data Preprocessing

In this section, a brief explanation of the FOCUS game will be presented, followed by the data wrangling techniques that we used. For more detailed understanding of the game, please refer to [31].

A. Game FOCUS

We designed FOCUS using the unity game engine. The game digitally mimics a few existing clinical and rehabilitation therapies. The details of the game can be seen in [31]. Here, we present a brief summary of how it works, and refer readers to [31] for more details. The game is EEG controlled and can be played using the EMOTIV EPOC+ kit. The game challenges the player to move an avatar, by focusing and using mental commands, to collect all the cubical pickups in the shortest time possible. As mentioned before, in [31], we showed that the brain-controlled mode had an increase of 10% increase in engagement and 8% in focus over the keyboard-controlled mode, both with healthy subjects. The game environment is shown in Fig. 2. The wearable wireless EEG device, EMOTIV which is shown in Fig. 1, was used to control the serious game via the open-source SDK provided by the company that is compatible with the famous unity 3-D game engine. Last, the raw EEG data were extracted and recorded during the testing sessions using python scripts run in the background and classification models were built also using python.

In this paper, we extended our previous work in two aspects. Subjects were asked to use two ways of game control: controlling the character with keyboard arrows and controlling the character using the EEG device. First, we tried to classify the EEG signals according to their attention states, a.k.a. labels, which in our case are two labels: keyboard controlled and EMOTIV controlled. The presumption is that playing the keyboard-controlled game and the EMOTIV-controlled game has different attention levels, so they can be separable. Second, since ADHD patients have problems with attention and different EEG characteristics, we assume that it is possible to detect persons suffering from ADHD by classifying their EEG signals.

For the remainder of this paper, we focus on the ML classification models that were built using the data recordings of our test subjects. We will start by explaining the data wrangling techniques applied to the data. We will then present our preliminary study, done before access to ADHD patients became possible, in which we tried to classify between the data recorded by the EEG-controlled mode and the keyboard-controlled mode, as they represent different types of attention levels. We will then show our extended study, using actual ADHD subjects, and we will explain and analyze our proposed ADHD classification models.

B. Data Wrangling

Some data wrangling techniques were applied to the data before building the classification models. The techniques aimed at enhancing the performance of the model and obtaining the best accuracy possible.

1) Frequency Bands' Data Filtering: The data used in the classification models were only the EEG frequency bands' data, since the frequency bands' data is more compact in size because of applying fast Fourier transform, conserving the same wave characteristics. In addition, the previous research recorded some abnormal activities in the θ/β bands for ADHD diagnosed individuals.

2) Data Windowing: Data windowing was employed on the data in order to improve the accuracy of the models. Each continuous recording session was divided into chunks by a window of n-rows (representing a time step). Each chunk was considered a single sample.

The windows applied were non-overlapping, in order to eliminate the duplication in the data. Since our data are small, using overlapping windows will result in having a biased result due to the duplication of the data. After applying the data windowing of size 5, the total number of data samples reached 57 000 samples.

3) Feature Extraction (Extra Attributes): In this paper, we would name the extracted features as attributes in order to not cause confusion with the features that are input to the model. The attributes in the data represent only the five frequency bands. As explained in detail in [32], the ratio of θ/β subbands gave the highest accuracies, a fact that is also confirmed by [22] and [23]. According to [22], the ratio between θ- and β- bands relates to the attention levels of the person. Loo and Makeig [23] proved that ADHD patients are more likely to produce comparatively less of the higher frequency band (β) and more of the lower frequency band (θ); therefore, the ratio θ/β is quite significant for such research. Hasegawa and Oguri [33] showed that the ratio between α- and β-bands carries the same importance and relates to the attention levels.

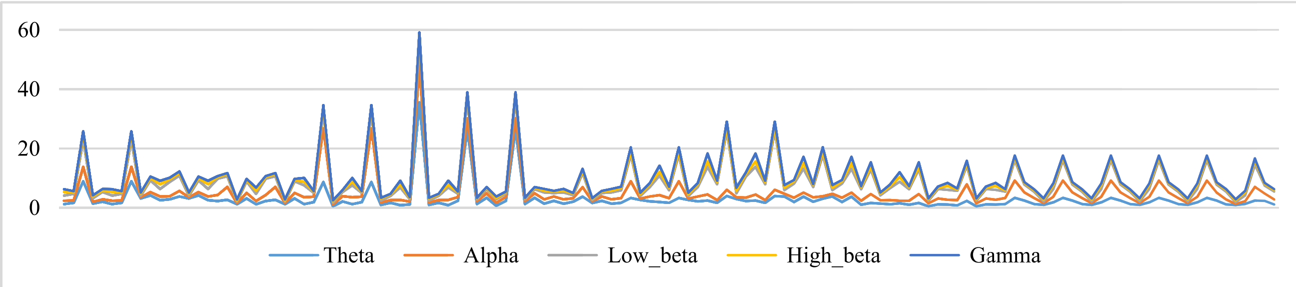

Fig. 3. Data from a sample recording session during FOCUS gameplay that illustrates different frequency bands.

TABLE I
SAMPLE DATA RECORDING DURING FOCUS GAMEPLAY

T	**Alpha**	**High Beta**	**Low Beta**	**Theta**	**Gamma**
0.5	4.359	0.878	2.307	0.505	0.382
1.0	6.616	1.362	2.756	0.634	0.309
1.5	62.127	18.296	15.896	7.068	2.527
2.0	1.457	1.559	1.474	1.360	1.134
2.5	717.397	255.905	344.340	118.098	31.149
3.0	4.359	0.878	2.307	0.505	0.382
3.5	6.616	1.362	2.756	0.634	0.309
4.0	62.127	18.296	15.896	7.068	2.527
4.5	1.457	1.559	1.474	1.360	1.134
5.0	717.397	255.905	344.340	118.098	31.149

Our recorded data contained high β and low β, and consequently, four extra attributes were added: θ/high β, θ/low β, α/high β, and α/low β. Table I, in addition to Fig. 3, exemplifies a sample recording session during gameplay that illustrates different frequency bands and some of the data collected.

Since we are using SVMs as classification models, it is a requirement that each data sample to be represented as a vector of features. Since we have nine features (after adding the new ratio between the bands described above), we flatten all the features in a vector. With the addition of extra data attributes, the vector length will be calculated by the window size multiplied by the total number of attributes as in the following equation:

$$\text{Vector Length} = \text{Window size} \times \text{Num of Attributes}. \quad (1)$$

After applying the above data wrangling techniques, each data point was labeled accordingly, and models were built. The model was trained on the training data folds; then, they were tested on the remaining fold.

IV. PROPOSED CLASSIFICATION MODELS, EXPERIMENTS, AND RESULTS

A. *EMOTIV-Controlled Versus Keyboard-Controlled Classification (Healthy Subjects)*

In the first stage of this work, healthy people were tested by assigning them tasks which require levels of attentiveness. Since keyboard-controlled recordings showed lower attention levels than EMOTIV-controlled recordings [31], detecting the control type of the game, irrespective of whether the player is ADHD or non-ADHD, gives a powerful insight into our method's ability to classify EEG recordings based on the concentration levels.

1) Prerecording Instructions: Instructions were given to subjects on how to play the FOCUS game, using both the keyboard and the EMOTIV for control. Subjects were asked to play a nonrecorded EMOTIV-controlled game at the start in order to remove the training effect. Once the subjects were confident of being able to control the avatar via EMOTIV, actual recording was initiated.

For these models, we used five healthy subjects (males, age range 19–26). Each test subject was asked to play the game using the arrow buttons on the keyboard to collect the pickups. Their EEG signals were recorded while playing. After that, the subject was asked to play the same game without the keyboard, and using the EMOTIV. Again, their EEG signals were recorded while playing. The subjects were asked to repeat the process 5 times for the keyboard-controlled game and then 5 times for the EMOTIV-controlled game, resulting in 10 recordings per person and 50 sessions in total while having a 2–3-min break after each recording session.

The average time measured was 1 min for a keyboard-controlled recording, and 2.5 min for an EMOTIV-controlled recording. The EMOTIV-controlled game showed a higher average time because controlling a game with EEG was a new concept to all subjects. Since the EMOTIV-controlled game requires higher levels of attention and focus in the game, it is challenging for the subjects to get the character moving without pressing a button. The presented EEG-control concept is also not perfectly accurate, which introduces a lag or a jitter that adds to the average time measured to complete the task in the gameplay.

While subjects were playing, two different scripts in the background collected the EEG frequency bands (α, β, θ, and γ). The data were then labeled according to the game control type (keyboard as 0 or EMOTIV as 1). Since playing the FOCUS game with EMOTIV rather than the keyboard leads to a significant engagement in focus and attention abilities of healthy individuals, we could say that the ability to detect the game control type with healthy subjects by EEG data is, conceptually, similar to the detection and diagnosis of attention deficit individuals. For more details about the game controlling mechanisms and data acquisition techniques, we refer the readers to [32].

The following sections present the results for each of our classification models.

TABLE II
RBF MODEL RESULTS

Window size	Count	Mean	STD	Median	Max
50	5	0.8110	0.0179	0.8129	0.8295
20	5	0.8769	0.0068	0.8761	0.8840
5	5	0.8926	0.0016	0.8927	0.8944

2) Support Vector Machine (SVM) Models: In this model, SVM used to classify data samples to one of two labels: keyboard controlled and EMOTIV controlled. Multiple SVM models were built tuning the models' parameters and changing the testing criteria. Data windowing, data subsampling, data shuffling, feature extraction, and cross-validation were utilized.

It is important to mention that the C parameter of the SVM (the regularization parameter that prevents overfitting and/or underfitting) was tuned iteratively without overfitting the data and it occurred at C equal to 3. Also, the accuracies obtained are calculated simply by dividing the true positives over the total number of data points as in (2). The true positives include the correctly classified samples from both brain-controlled and keyboard-controlled data points. We have also recorded the precision, the recall, and the F1-score for the models

$$\text{Accuracy} = \text{True Positive/Total Data Points.} \tag{2}$$

We then designed multiple different SVM models, for the purpose of analyzing the data and finding the best for attention classification. Described next are a few of the models; they are the ones that were repeated for ADHD classification in Section IV-B. To see other models used with healthy subjects only, we refer the readers to [32].

a) Radial basis function (RBF) kernel model: More sophisticated kernel than the linear kernel was used, in addition to applying the non-overlapping windows. Data were split into different ranges of windows: 50, 20, and 5 steps producing in approximately 2000, 5000, and 20 000 data samples, respectively. The idea is to choose the best fit for this experiment in order to utilize it later with the other experiments.

Each of the created models was tested by fivefold cross-validation. The average, standard uncertainty, min, max, and median of the accuracies on the test samples of the fivefold are shown in Table II.

According to the obtained results, the window size 5 produced the best mean accuracy of 89.26%. Therefore, the windows size 5 will be used in all later experiments unless it is mentioned otherwise.

To validate the accuracies obtained previously, the confusion matrix of the model with window size 5 was calculated. The confusion matrix accounts the precision, recall, and F1-score of each of keyboard and EMOTIV labels. These are shown in Table III.

b) RBF kernel model with extra attributes: We added the four extra features to the data after applying feature extraction which increased the features count to nine. Data samples were shuffled, balanced, and inputted to the model. Fivefold cross-validation was used. The obtained accuracies' mean was

TABLE III
CONFUSION MATRIX OF RBF MODEL (W SIZE = 5)

Labels	Precision	Recall	F1-Score	Samples count
Keyboard	0.86	0.92	0.89	1900
Emotiv	0.93	0.86	0.89	2115
Avg. / **Total**	0.89	0.89	0.89	**4015**

TABLE IV
RBF MODEL'S RESULTS

Window size	Count	Mean	STD	Median	Max
5	5	0.9610	0.0027	0.9601	0.9644

96.1%. The accuracies' summary on the test samples of the fivefold is shown in Table IV.

c) One subject's data holdout: For a deeper understanding of the obtained results, a different testing methodology is needed. This model was built to illustrate how well the SVM model can generalize on data from subjects that it did not encounter. A similar model to the previous model was implemented with different testing methodologies. The model was trained on data from four different subjects and tested on the fifth subject's data. The methodology was iterated 5 times for each of the five test subjects. The goal was to verify the ability of the model to generalize on data from individuals who were not seen before.

Iteration is similar to the cross-validation approach which averages out the accuracies. The results of the fivefold were: 70.86%, 62.45%, 60.66%, 58.56%, and 46.39%, with an average of 59.78%. The testing showed relatively lower accuracies than the previous models. This could be the result of having a small number of subjects.

B. ADHD Versus Non-ADHD Classification

At this stage, we started testing with ADHD persons. So, this section presents the models that detect ADHD patients. Using the same serious game, FOCUS, and the same methodology explained earlier, we tested four ADHD subjects. The same testing methodology was applied to ADHD patients; whereby, they were asked to play the game using both control methods: keyboard and EMOTIV. Two different classifiers were built, separately, using the data recorded from two different methods.

The extended experiments done in this section include testing the game with four subjects: two males (18- and 23-year old) and two females (21- and 22-year old), who were clinically diagnosed as suffering from ADHD at an early age. The four ADHD subjects will be referred to as P1, P2, P3, and P4. We asked about the medical history of the subjects before doing the experiments. P1 was reported to have a higher rate of hyperactivity than his/her peers at an early age. P2 was clinically trained to improve his/her focusing abilities, which helped him/her to bypass the obstacles of the ADHD effects, while the other ADHD subjects were not. P3 and P4 did not report any medical history other than their previous medical ADHD diagnosis test.

1) Data Recordings and Data Labeling: The same prerecording instructions were presented to ADHD patients before the recording sessions started. Subjects played a nonrecorded EMOTIV-controlled game in the beginning to make sure that patients are confident how to control the game, and to remove the training effect. Once they were confident of controlling the avatar using the EMOTIV, the recorded sessions started.

Similar to Section IV-A, each test subject was asked to play the FOCUS game using the arrow buttons on the keyboard to control the avatar collecting the pickups. After that, the subject was asked to play the same game using the EMOTIV kit to control the avatar movement. The subject was asked to repeat the two games 6 times, producing six recording sessions for each of the keyboard-controlled and EMOTIV-controlled games, resulting in 12 recordings per person and 48 sessions in total. The average time measured was 1 min for a keyboard-controlled recording and 4 min for an EMOTIV-controlled recording. As expected, the average duration for the EMOTIV-controlled recording by ADHD persons is higher than that of healthy subjects, due to ADHD persons' difficulty in focusing.

Before designing the model, data were labeled according to the subject (non-ADHD-0 and ADHD-1). Since ADHD patients have lower attention spans and therefore have different EEG patterns, our goal is to try to detect ADHD patients using the EEG data that is recorded during gameplay since attention deficit individuals have, conceptually, similar EEG patterns that could be diagnosed.

Next, we repeat the same models from Section IV-A, but this time used differently to detect ADHD.

2) Support Vector Machine (SVM) Models: To detect ADHD, data samples will be labeled 0-non-ADHD sample and 1-ADHD sample. Since recording sessions were using two different control methods, all the models that use different data divisions will be done separately on keyboard-controlled and EMOTIV-controlled data. Accordingly, the results will be compared on this basis.

Since the RBF kernel gave the best results in the models presented above, all the upcoming models will use the same kernel. Also, in most of the models, the extra attributes mentioned in Section III-B3 were also used since they are proven to increase the accuracy in previous models.

Since we had five non-ADHD subjects and four ADHD subjects, the non-ADHD samples were slightly more than the ADHD samples. The four ADHD subjects will be referred to as P1, P2, P3, and P4, while the five non-ADHD subjects will be referred to as S1, S2, S3, S4, and S5. It is very important to note that unlike other ADHD subjects, P2 was clinically diagnosed with ADHD at a young age and since the early stages, he/she was going to a therapist and had learned some techniques to help with attention. In other words, P2 knew how to focus, much better than other ADHD subjects. As we will see, this will have a significant role in the results as it fooled some of our detection models into believing that P2 is not ADHD.

a) RBF kernel with extra attributes: First, the model was trained on the data from EMOTIV-controlled sessions. The total number of data samples after applying data wrangling techniques on EMOTIV recording sessions was 37 725 samples, where each model has 7545 samples in each fold of the fivefold cross-validation to test the model on.

TABLE V
RBF KERNEL MODEL'S RESULTS USING EMOTIV DATA

Count	Mean	STD	Median	Max
5	0.98624	0.0016	0.98701	0.98754

TABLE VI
RBF KERNEL MODEL'S RESULTS USING KEYBOARD DATA

Count	Mean	STD	Median	Max
5	0.98232	0.0027	0.98381	0.98382

Data samples from the two classes were balanced throughout the folds (7545 total samples) such that in each fold non-ADHD samples ranged between 4126 and 4176 samples (54.68%–55.35%) while ADHD samples ranged between 3378 and 3419 samples (44.77%–45.31%). Also, fivefold cross-validation was applied as a testing methodology for the model.

The mean of the accuracies obtained was 98.62% with an uncertainty of 0.16% for the fivefold cross-validation. The average, standard uncertainty, min, max, and median of the accuracies on the test samples of the fivefold are shown in Table V.

For the keyboard-controlled sessions, the total number of data samples after applying data wrangling on keyboard recording sessions was 16 065 samples, where each model has 3213 samples in each fold of the fivefold cross-validation to test the model on.

Similarly, data samples from the two classes were balanced throughout the folds (3213 total samples) such that in each fold non-ADHD samples ranged between 1922 and 2016 samples (59.82%–62.75%) while ADHD samples ranged between 1197 and 1291 samples (37.25%–40.18%). Also, a fivefold cross-validation was applied as a testing methodology for the model.

The mean of the accuracies obtained was 98.23% with a standard uncertainty of 0.27% for the fivefold cross-validation applied. The average, standard uncertainty, min, max, and median of the accuracies on the test samples of the fivefold are shown in Table VI.

b) One ADHD subject's data holdout: In this model, only ADHD subject's data were held out since the overall goal is ADHD diagnosis. Similar to what was done previously, one of the ADHD subject's data was held out of the training phase and then the model was tested. This process was iterated 4 times taking each of the ADHD subjects as a test set while the non-ADHD subjects' data were all kept in the training set. This way of testing helped us draw some conclusions about the ability of classification algorithms and how it performs on data from a subject whom it did not encounter and was not trained on. Similar to the other models, the same process was repeated twice; once with EMOTIV-controlled recordings and another with the keyboard-controlled sessions.

TABLE VII
ONE-HOLDOUT MODEL'S RESULTS USING EMOTIV DATA

Subject held-out	P1	P2	P3	P4
Test Accuracy	0.7219	0.3875	0.8671	0.8124
No. of Test Data Samples	6310	4356	1648	4681

TABLE VIII
ONE-HOLDOUT MODEL'S RESULTS USING KEYBOARD DATA

Subject held-out	P1	P2	P3	P4
Test Accuracy	0.9060	0.1513	0.7789	0.5529
No. of Test Data Samples	2426	1051	570	2154

The results obtained from the four models were somehow different from each other, for both EMOTIV-controlled-based models and keyboard controlled.

For EMOTIV-controlled data, Table VII illustrates the obtained training accuracy, test accuracy, and the number of samples in the test set for each of the subjects.

The total number of data samples was 37 725 samples, and the number of test data samples was presented in order to validate the credibility of the accuracies obtained. Test data ranged between 4.37% and 16.73% of the whole data.

Test accuracies varied from one subject to another. For P1, P3, and P4, at least two-thirds of the data were labeled as ADHD data, and therefore, the subjects should be diagnosed as ADHD subjects. Generally, the line should be drawn at 50%, but that is still a naïve assumption due to the difficulty of accessing ADHD subjects and collecting more data.

The results obtained from subject P2 were around 39%, which means that around two-fifths of the data were predicted to be ADHD. Since there is not enough data, we could go with the assumption that around 40% of the data is enough to diagnose a person with ADHD. Another justification to the result is that, as mentioned earlier, P2 was clinically trained to improve his/her focusing abilities, which helped him to bypass the obstacles of the ADHD effects, while the other ADHD subjects were not. The results in the upcoming models will support this claim.

On the hand, for keyboard-controlled data, Table VIII illustrates the obtained training accuracy, test accuracy, and the number of samples in the test set for each of the subjects.

The total number of data samples is 16 065 samples. Test data ranged between 3.55% and 15.10% of the whole data.

Test accuracies varied from a subject to another. For P1, P3, and P4, respectively, 90%, 78%, and 55% of their data was labeled as ADHD data, and therefore, the subjects should be diagnosed as ADHD subjects.

The results of P2 showed that only 15% of his/her data was labeled as ADHD data. This result supports our claim earlier, and it shows clearly that the data recorded from P2 is quite different from that of other ADHD subjects.

By comparing EMOTIV-based models to keyboard-based models, the earlier has an average accuracy of 69.72% for all subjects while the later has an average accuracy of 59.73%.

TABLE IX
TWO-HOLDOUT EMOTIV MODEL'S RESULTS

Subject held-out	1- S1, P1		2- S2, P2		3- S3, P3		4- S4, P4	
	Non	ADHD	Non	ADHD	Non	ADHD	Non	ADHD
Precision	0.49	0.69	0.41	0.44	0.96	0.53	0.55	0.52
(avg)	**0.62**		**0.43**		**0.85**		**0.54**	
Recall	0.39	0.77	0.43	0.42	0.73	0.90	0.15	0.89
(avg)	**0.63**		**0.43**		**0.77**		**0.53**	
F1-score	0.43	0.73	0.42	0.43	0.83	0.66	0.23	0.66
(avg)	**0.62**		**0.43**		**0.79**		**0.45**	
Test Accuracy	0.6327		0.4251		0.7713		0.5273	
No. of Test Samples	3566	6310	4082	4356	4922	1648	4410	4681
(total)	**9876**		**8438**		**6570**		**9091**	

TABLE X
TWO-HOLDOUT KEYBOARD MODEL'S RESULTS

Subject held-out	S1, P1		S2, P2		S3, P3		S4, P4	
	Non	ADHD	Non	ADHD	Non	ADHD	Non	ADHD
Precision	0.90	0.66	0.64	0.33	0.94	0.63	0.71	0.64
(avg)	**0.77**		**0.53**		**0.87**		**0.68**	
Recall	0.41	0.96	0.80	0.18	0.87	0.80	0.74	0.60
(avg)	**0.71**		**0.58**		**0.86**		**0.68**	
F1-score	0.56	0.78	0.71	0.23	0.91	0.71	0.73	0.62
(avg)	**0.68**		**0.54**		**0.86**		**0.68**	
Training Accuracy	0.9955		0.9997		0.9959		0.9969	
Test Accuracy	0.7085		0.5763		0.8564		0.6815	
No. of Test Samples	2054	2426	1897	1051	2084	570	2786	2154
(total)	**4480**		**2948**		**2654**		**4940**	

EMOTIV-based models clearly outperformed the keyboard-based models which complements the scope of the EMOTIV-controlled-based game in the diagnosis of ADHD subjects.

c) Two subjects' data holdout (one ADHD and one non-ADHD): In this model, two subjects were randomly chosen from each the ADHD and non-ADHD groups. The model was built using the data of the rest of the subjects, and then it was tested and evaluated using the two subjects' data. The process was iterated 4 times, taking P1 and S1, P2 and S2, P3 and S3, and P4 and S4, as testing sets, respectively, in each of the four iterated models. The goal here is to use the same testing methodology that was used earlier with the addition of the ADHD subjects' data which will help in drawing some conclusions. This methodology of testing, which includes a non-ADHD subject's data to the test set, is important so we can generalize the results obtained earlier.

Table IX contains the results obtained from the four EMOTIV-based models, which contains the precision, recall, and F1-score for both the ADHD and non-ADHD classes.

While the total number of samples is 37 725, the test data samples ranged between 17.42% and 26.18% of the whole data. The overall testing accuracy is better than a random model with the exception of the second model, while the reasoning might be the data recorded by subject P2. The highest accuracy obtained was 77.13%, while the lowest being 42.51%. Since ADHD is a psychological disorder, the low results obtained by the fourth model could be explained as S4 suffering from ADHD while she/he is labeled as a non-ADHD subject in our experiment.

TABLE XI
RESULTS' SUMMARY

Classification method	Testing methodology	Samples' Count (Test Samples)	Testing Accuracy (avg)
RBF Kernel Model Results Using Emotiv Data	5-folds cross-validation	37,725 (7545)	98.62%
RBF Kernel Model Results Using Keyboard Data	5-folds cross-validation	16,065 (3213)	98.23%
One ADHD Subject's Data Holdout Using Emotiv Data	1-subject data holdout	37,725 (4.37% - 16.73%)	69.72%
One ADHD Subject's Data Holdout Using Keyboard Data	1-subject data holdout	16,065 (3.55% - 15.10%)	59.73%
Two Subjects' Data Holdout (one ADHD and one non-ADHD Using Emotiv Data	2-subjects data holdout	37,725 (17.42% - 26.18)	58.91%
Two Subjects' Data Holdout (one ADHD and one non-ADHD Using Keyboard Data	2-subjects data holdout	16,065 (16.52% - 30.75%)	70.57%

TABLE XII
SIMILAR WORKS' RESULTS

Paper	Data Used	Classification Technique	Serious Game?	EEG	Diagnosis accuracy
Our work	EEG recording during gameplay	RBF SVM Classifier	Yes	Yes	98.62%
[8]	EEG data with eye closed	RBF Neural Network Classifier	No	Yes	95.6%
[9]	EEG recording during a cognitive task	RBF SVM Classifier	No	Yes	95.4% (4.6% error)
[6]	ADHD vs. ASD Using Behavioral Data	Logistic Regression, SVM, Linear discriminant analysis, and Categorical Lasso	No	No	96.2% – 96.5%
[10]	MRI Data	ELM, Linear-SVM, and RBF-SVM	No	No	90.18%, 84.73%, and 86.55% respectively.

On the other hand, Table X contains the results obtained from the four keyboard-based models, which contains the precision, recall, and F1-score for both the ADHD and non-ADHD classes.

While the total number of samples is 16 065, the test data samples ranged between 16.52% and 30.75% of the whole data. The overall testing accuracy is better than that of a random model with the highest being 85.6% and the lowest being 57.6%.

By comparing EMOTIV-based models to keyboard-based models, the former has an average accuracy of 58.91% for all subjects while the latter has an average accuracy of 70.57%. In this case, keyboard-based models performed better than EMOTIV-based models and the reasoning could be the existence of the subjects P2 and S4 whose data did not perform well while testing the classifier.

C. Summary of Results and Comparison

In Table XI, we summarize the ADHD versus non-ADHD classification models and the obtained results using our proposed method.

To compare the obtained results against the existing literature, similar projects were chosen, and the results are summarized and compared with our work in Table XII. Our work not only shows a higher diagnosis accuracy, but also stands out being the only work that utilizes a serious game, offering the aforementioned benefits. It should be taken into account that the tests in each of these works are obviously not on the same human subjects.

V. CONCLUSION

In this paper, we studied a new measurement system, consisting of an EEG recorder, a serious game, and ML classifier, to detect ADHD patients. The novelty of this paper is not only being the first work that laid the groundwork for the creation of an EEG-controlled serious game for detecting ADHD patients, but also the obtained classification results were promising, and they illustrate a strong feasibility of a robust classification system.

Our pilot experiments with healthy individuals showed an accuracy of up to 96% in classifying the EEG data to detect the correct attention state during gameplay, and our extended experiments with ADHD patients show an accuracy up to 98% in classifying the patients' EEG data.

Planning to use more interactive game scenarios is one of the future goals of this research. More interactive scenarios might have a positive effect on ADHD patients' attention span. In addition, creating serious games that are supported by virtual reality glasses, such as the Oculus Rift and HTC VIVE, might also enrich the interaction and the engagement during the rehabilitation session. Last, creating a multiplayer platform that support EEG-based games is a long-term goal.

ACKNOWLEDGMENT

This work has been done under ethical approval from the Istanbul Sehir University, Istanbul, Turkey, permission number 09/2017.

REFERENCES

[1] P. Rego, P. M. Moreira, and L. P. Reis, "Serious games for rehabilitation: A survey and a classification towards a taxonomy," in *Proc. 5th Iberian Conf. Inf. Syst. Technol.*, 2010, pp. 1–6.
[2] M. van Lambalgen, C. van Kruistum, and E. Parigger, "Discourse processing in attention-deficit hyperactivity disorder (ADHD)," *J. Logic, Lang. Inf.*, vol. 17, no. 4, pp. 467–487, Oct. 2008.
[3] Centers for Disease Control and Prevention. *Key Findings: Newborn Screening for Critical Congenital Heart Defects Now Common Throughout United States | NCBDDD | CDC*. Accessed: Nov. 15, 2016. [Online]. Available: http://www.cdc.gov/ncbddd/heartdefects/features/kf-newborn-heart-screening-commom-us.html
[4] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of BMD and ADHD patients using their EEG signals," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1956–1963, 2011.
[5] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors," *Sensors*, vol. 13, no. 8, pp. 10273–10286, Aug. 2013.
[6] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioral distinction of autism and ADHD," *Transl. Psychiatry*, vol. 6, no. 2, p. e732, Feb. 2016.

[7] M. Sabeti, S. D. Katebi, R. Boostani, and G. W. Price, "A new approach for EEG signal classification of schizophrenic and control participants," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2063–2071, 2011.

[8] M. Ahmadlou and H. Adeli, "Wavelet-synchronization methodology: A new approach for EEG-based diagnosis of ADHD," *Clin. EEG Neurosci.*, vol. 41, no. 1, pp. 1–10, Jan. 2010.

[9] B. Abibullaev and J. An, "Decision support algorithm for diagnosis of ADHD using electroencephalograms," *J. Med. Syst.*, vol. 36, no. 4, pp. 2675–2688, Aug. 2012.

[10] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural MRI data," *PLoS ONE*, vol. 8, no. 11, p. e79476, Nov. 2013.

[11] D. A. Fair *et al.*, "Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data," *Frontiers Syst. Neurosci.*, vol. 6, p. 80, Feb. 2013.

[12] K. Boletsis, "Augmented reality serious gaming for cognitive health," Ph.D. dissertation, Faculty Comput. Sci. Media Technol., Norwegian Univ. Sci. Technol., Trondheim, Norway, 2016.

[13] S. Göbel and J. Wiemeyer, *Games for Training, Education, Health and Sports: 4th International Conference on Serious Games (GameDays)* (Lecture Notes in Computer Science), Darmstadt, Germany, vol. 8395, Apr. 2014, p. 144.

[14] *ERASMUS+ Strategic Partnership Project: Intelligent Serious Games for Social and Cognitive Competence*. Accessed: Nov. 23, 2016. [Online]. Available: http://www.isg4competence.com/

[15] D. C. Hammond, "What is neurofeedback?" *J. Neurotherapy*, vol. 10, no. 4, pp. 25–36, 2007.

[16] C.-H. Han, J.-H. Lim, J.-H. Lee, K. Kim, and C.-H. Im, "Data-driven user feedback: An improved neurofeedback strategy considering the interindividual variability of EEG features," *BioMed Res. Int.*, vol. 2016, pp. 1–7, Jul. 2016.

[17] D. Vernon *et al.*, "The effect of training distinct neurofeedback protocols on aspects of cognitive performance," *Int. J. Psychophysiol.*, vol. 47, no. 1, pp. 75–85, 2003.

[18] EMOTIV. *EMOTIV—Brainwear Wireless EEG Technology*. Accessed: Nov. 15, 2016. [Online]. Available: https://www.emotiv.com/

[19] B. Penolazzi, C. Spironelli, and A. Angrilli, "Delta EEG activity as a marker of dysfunctional linguistic processing in developmental dyslexia," *Psychophysiology*, vol. 45, no. 6, pp. 1025–1033, Nov. 2008.

[20] Q. Wang, O. Sourina, and M. K. Nguyen, "EEEG-based 'serious' games design for medical applications," in *Proc. Int. Conf. Cyberworlds*, 2010, pp. 270–276.

[21] A. Lenartowicz and S. K. Loo, "Use of EEG to diagnose ADHD," *Current Psychiatry Rep.*, vol. 16, no. 11, p. 498, Nov. 2014.

[22] H. Heinrich, K. Busch, P. Studer, K. Erbe, G. H. Moll, and O. Kratz, "EEG spectral analysis of attention in ADHD: Implications for neurofeedback training?" *Frontiers Hum. Neurosci.*, vol. 8, p. 611, Aug. 2014.

[23] S. K. Loo and S. Makeig, "Clinical utility of EEG in attention-deficit/hyperactivity disorder: A research update," *Neurotherapeutics*, vol. 9, no. 3, pp. 569–587, Jul. 2012.

[24] R. Janca, P. Jezdik, R. Cmejla, R. Glajcar, P. Krsek, and A. Jahodova, "Seizure onset zone detection and localization in iEEG using DTF," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2011, pp. 93–97.

[25] N. Mammone, A. Lay-Ekuakille, F. C. Morabito, A. Massaro, S. Casciaro, and A. Trabacca, "Analysis of absence seizure EEG via permutation entropy spatio-temporal clustering," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2011, pp. 532–535.

[26] R. F. Ahmad, A. S. Malik, H. U. Amin, N. Kamel, and F. Reza, "Classification of cognitive and resting states of the brain using EEG features," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, May 2016, pp. 1–5.

[27] A. Lay-Ekuakille *et al.*, "EEG signal processing and acquisition for detecting abnormalities via bio-implantable devices," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, May 2016, pp. 1–5.

[28] X. Chen, H. Peng, F. Yu, and K. Wang, "Independent vector analysis applied to remove muscle artifacts in EEG data," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1770–1779, Jul. 2017.

[29] S.-F. Liang, Y.-H. Kuo, Y.-H. Hu, Y.-H. Pan, and Y.-H. Wang, "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1649–1657, Jun. 2012.

[30] C.-P. Young, S.-F. Liang, D.-W. Chang, Y.-C. Liao, F.-Z. Shaw, and C.-H. Hsieh, "A portable wireless online closed-loop seizure controller in freely moving rats," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 513–521, Feb. 2011.

[31] A. E. Alchalabi, A. N. Eddin, and S. Shirmohammadi, "More attention, less deficit: Wearable EEG-based serious game for focus improvement," in *Proc. 5th IEEE Conf. Serious Games Appl. Health (SeGAH)*, Apr. 2017, pp. 1–8.

[32] A. E. Alchalabi, M. Elsharnouby, S. Shirmohammadi, and A. N. Eddin, "Feasibility of detecting ADHD patients' attention levels by classifying their EEG signals," in *Proc. 12th Annu. IEEE Int. Symp. Med. Meas. Appl.*, May 2017, pp. 314–319.

[33] C. Hasegawa and K. Oguri, "The effects of specific musical stimuli on driver's drowsiness," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2006, pp. 817–822.

Alaa Eddin Alchalabi received the M.Sc. degree in electronics and computer engineering from Istanbul Sehir University, Istanbul, Turkey. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Ottawa, Ottawa, ON, Canada.

He is currently a Research Assistant with the Distributed and Collaborative Virtual Environment Research Laboratory, University of Ottawa, where he is involved in machine learning and multimedia systems.

Shervin Shirmohammadi (M'04–SM'04–F'17) received the Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada.

He is currently a Professor with the School of Electrical Engineering and Computer Science, University of Ottawa, where he is also the Director of the Distributed and Collaborative Virtual Environment Research Laboratory, focusing on multimedia systems and networks, including measurement techniques for network, video streaming, and health systems. The results of his research, funded by more than U.S. $13 million from public and private sectors, have led to over 300 publications, over 70 researchers trained at the postdoctoral, Ph.D., and master's levels, and over 20 patents and technology transfers to the private sector.

Dr. Shirmohammadi was a recipient of a number of awards. He is an IEEE Fellow for contributions to multimedia systems and network measurements, a Senior Member of the ACM, a University of Ottawa Gold Medalist, and a licensed Professional Engineer in Ontario. He has been an IMS AdCom Member since 2014, and served as the Vice President for its Membership Development Committee from 2014 to 2017, as well as a member of the IEEE I^{2}MTC Board of Directors from 2014 to 2016. He was the Associate Editor-in-Chief of the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT in 2013 and 2016, and the *IEEEInstrumentation and Measurement Magazine* in 2014 and 2015. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, and an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and ACM TRANSACTIONS ON MULTIMEDIA COMPUTING, COMMUNICATIONS, AND APPLICATIONS.

Amer Nour Eddin received the bachelor's degree in computer science and engineering from Istanbul Sehir University, Istanbul, Turkey.

He was a Research Assistant with the Multimedia Systems Lab., Istanbul Sehir University. His research interests included creating 3-D serious games for children with cognitive disabilities.

Mohamed Elsharnouby received the bachelor's degree in medicine from the University of Tanta, Tanta, Egypt, and the M.Sc. degree in electronics and computer engineering from Istanbul Sehir University, Istanbul, Turkey.

His research interests included networks and adaptive video streaming.