Intelligent Automation And Soft Computing, 2018
Copyright © 2018, TSI® Press
Vol.24,no.4,891–905

Statistical Analysis and Multimodal Classification on Noisy Eye Tracker and Application Log Data of Children with Autism and ADHD

Mahiye Uluyagmur Ozturk[a], Ayse Rodopman Arman[b], Gresa Carkaxhiu Bulut[c], Onur Tugce Poyraz Findik[b], Sultan Seval Yilmaz[d], Herdem Aslan Genc[b], M. Yanki Yazgan[e,f], Umut Teker[a], Zehra Cataltepe[a]

[a]Computer Engineering Department Istanbul Technical University Istanbul, Turkey;
[b]Marmara University Medical Faculty Child and Adolescent Psychiatry, Istanbul, Turkey;
[c]Mus State Hospital Child and Adolescent Psychiatry, Mus, Turkey;
[d]TC Health Municipality Medeniyet University Child and Adolescent Psychiatry, Istanbul Turkey;
[e]Güzel Günler Polyclinics, Istanbul, Turkey;
[f]Yale Child Study Center, New Haven, CT, USA

ABSTRACT
Emotion recognition behavior and performance may vary between people with major neurodevelopmental disorders such as Autism Spectrum Disorder (ASD), Attention Deficit Hyperactivity Disorder (ADHD) and control groups. It is crucial to identify these differences for early diagnosis and individual treatment purposes. This study represents a methodology by using statistical data analysis and machine learning to provide help to psychiatrists and therapists on the diagnosis and individualized treatment of participants with ASD and ADHD. In this paper we propose an emotion recognition experiment environment and collect eye tracker fixation data together with the application log data (APL). In order to detect the diagnosis of the participant we used classification algorithms with the Tomek links noise removing method. The highest classification accuracy results were reported as 86.36% for ASD vs. Control, 81.82% for ADHD vs. Control and 70.83% for ASD vs. ADHD. This study provides evidence that fixation and APL data have distinguishing features for the diagnosis of ASD and ADHD.

KEYWORDS: **Classification of Medical Diagnosis, Emotion Recognition Ability, Eye Tracking, Noise Removal.**

1 INTRODUCTION

ONE of the essential components of social interaction is recognizing emotions since emotion recognition provides the understanding of other people's intentions accurately and reacting to them appropriately (Kuusikko et al., 2009). Emotion recognition processes carry clues about one's overall emotional well-being. Interpreting the facial expressions makes it easier to communicate with other people (Delaherche et al., 2013). Deficits in social interaction and social attitude are critical symptoms of children and adults with Autism Spectrum Disorder (ASD) or Attention Deficit Hyperactivity Disorder (ADHD), increasing the importance of emotion recognition for people with these major neuro-developmental disorders.

In this work, we focus on the analysis of emotion recognition behavior of the ASD, ADHD and the control groups. ASD is a complex neuro-developmental disorder that usually surfaces during the first year of life (American Psychiatric Association, 2000). Some characteristics of ASD are as follows: Difficulties in social communication and interactions, the problem in conducting and sustaining a relationship, disability in establishing eye contact. These could lead to impairment in understanding the emotion and intention of others. Also, insistence on sameness, strict adherence to routine, repetitive behaviors, and limited and intensive interests are characteristics of ASD (Lord et al., 2000). On the other hand, symptoms of ADHD are; hyperactivity, impairment of both attention and concentration and impulsivity (American Psychiatric Association, 2000). Besides, both ASD and ADHD groups show a lack of concern or inability to react to other people's emotions or feelings (Craig et al., 2015). Due to these stated features, patients with ASD and/or ADHD have difficulty understanding and interpreting other peoples' emotions and moods. As a result, for children and adolescents diagnosed with ASD and ADHD, social communication becomes a burden (Myers & Johnson, 2007). Therefore, it is crucial to perform an in-depth analysis of emotion recognition

CONTACT Mahiye Uluyagmur Ozturk muluyagmur@itu.edu.tr

processes and investigate the different individual deficiencies of ASD and ADHD.

In the current study, we aim to distinguish the participants with ASD, participants with ADHD and the control group by using their emotion recognition experiment data. An experimental environment was prepared where the participants wore an eye tracker and they were shown some emotive facial images as stimuli. Emotional stimuli are used in many studies to measure human perception (Kemp et al., 2004, Hayashi et al., 2017). In our work, the participants were asked to state the emotion in the presented images. The purpose of the experiment was to understand how participants reacted to these images showing response and response time (RT), and how their eye movements changed during the experiment.

The response of the RT and the eye tracker fixation data were recorded and used for the analysis. We presented results in two different ways. First of all, we made statistical analysis of the differences in the emotional recognition behaviors of participant groups by using one-way ANOVA. Secondly, we used the data obtained during the experiments to classify the participants with the machine learning methods.

The main question that we try to answer in this work is whether the emotion recognition ability and process is different between the clinical groups (ADHD and ASD) and the control group, and if so, this difference has potential use for differential diagnosis. People with ADHD and people with ASD are subjected to a series of clinical tests in hospitals. These tests are generally subjective, costly, time consuming and burdensome (Duda et al., 2016). On the other hand, early detection is essential in the success of interventions for both ASD and ADHD. With the support of more studies on larger population sizes and alternative types of inputs, approaches like ours can be used to facilitate the early diagnosis and start an early treatment process. The target purpose is to help psychiatrist and therapist working in ASD and ADHD on the automation of diagnostics of the mentioned participant groups using a computer-aided technique.

The principal contributions of this work are as follows:

- We show that using fixation and application log data collected during the emotion recognition experiment is crucial to distinguish ASD, ADHD and the control groups from each other.
- We report that responses, response time and pupil diameter measurements of the participant groups have statistically significant differences.
- We compare three classification algorithms for the diagnosis of ASD and ADHD and show that the Random forest method yields to the best results.

2 RELATED WORK

IN order to analyze the emotion recognition process, some instrumental measurements have been used in recent studies. One of them is the eye tracking fixation data, which includes informative data about the autonomic nervous system and cognitive behavior (Bal et al., 2010). By using an eye tracker, we can learn about the visual processing details of participants when they try to recognize the emotions. Thanks for the eye gaze information, it is possible to reveal the reason behind the impairments of people with emotion recognition (Van Der Geest et al., 2002, Dalton et al., 2005). In Kirchner et al., (2011) fixation data and Autism Diagnostic Interview-Revised (ADI-R) score were used to compare the performance of the autistic group and the control group in emotion recognition and face identity tasks. In both experiments, the autistic group performed worse than the control group. Similarly, in our emotion recognition experiments, the control group gave more correct responses to the emotional images than the participants with the disorders.

Eye gaze data consists of a pupil diameter measurement. One of the indicators that reflects autonomic nervous system activity is the pupil diameter or pupil size (Partala & Surakka, 2003, Pedrotti et al., 2014). In the cognitive system or during emotional processes, pupil size gives critical information about mental workload and cognitive functioning (Caffrey, 2008). As stated by Hvelplund, (2014), higher cognitive effort enlarges the pupil size, but pupil size enlarges less during lower cognitive load tasks. Attention and information processing are related to pupil responses of the people (Hess, 1975). In (Baltaci & Gokcay, 2016) only a pupil diameter feature was used to detect stress conditions of the participants. Moreover, the pupil diameter was used to classify the patients with mild cognitive impairments and control subjects (Lagun et al., 2011). In (Lin et al., 2003, Lin et al., 2009) studies of the eye gaze data acquisition were performed and it is used for the measurement of the attention of the operators. Specifically, eye fixation and pupil diameter features were found important for detecting the attention level of the operators.

Our study used the fixation, response and response time data describing the behavior of the participants while they recognized the emotions of the people in the pictures shown to them. This data was used to identify the diagnosis of the patients. We did not aim to infer the emotions of the participants in this study. Similar approaches to ours were followed in (Lagun et al., 2011, Rivera et al., 2012). The fixation, saccade features were used to classify the patients with mild cognitive impairments in (Lagun et al., 2011). In (Rivera et al., 2012) the eye tracking variables such as; fixation and saccade were employed to discriminate learners from non-learners among 6-8 month-old infants and the same model used to categorize the adults. Even though these studies were used for classification by using eye movement data, no emotion extraction was done.

Continuous attention is required in order to be able to interpret and respond to facial expressions during a conversation instantly. Since attention duration is an essential point for social interaction, RT or reaction time measurements would give critical information (Miyahara et al., 2007). In the study by Gold & Gold (1975), significant differences were observed regarding the reaction time between participants with ASD and the control children, even if participants did not focus on the facial stimuli.

Furthermore, there are several studies for ASD and ADHD classification by using artificial intelligence and machine learning algorithms. The majority of these studies used fMRI (functional magnetic resonance imaging) data type. Although fMRI is a powerful data type, it is difficult to process. We reached similar classification performances with the studies in which fMRI were used (Ecker et al., 2010). On the other hand, we outperformed some fMRI works (Colby et al., 2012, Iannaccone et al., 2015) by using fixation and application log data (APL) data, which are easy to process. In this manner, we showed the usefulness of the fixation and APL data for detecting the type of the disorder. The importance values of the features selected by the Random forest algorithm showed that the fixation features played a more active role in the classification process. This result has further strengthened our confidence in the importance of collecting participants' eye gaze data using an eye tracker during an emotion recognition experiment.

The remaining parts of the article are organized as follows: In Section 3, information of the participants, the stimuli, experimental procedures and data collection process are introduced. Classification methods that were used are described in Section 4. Section 5 presents data analysis results and classification results. Discussion and conclusion are in Section 6 and Section 7.

3 EXPERIMENTS

IN this section, we give information about the participants and the experiment conducted.

3.1 Participants

The experiments were conducted in Marmara University Medical Faculty Hospital Child and Adolescent Psychiatry Outpatient Department under the clinical supervision of an MD Professor, A. Rodopman Arman. Participants with ASD had atypical autism as the first diagnosis and ADHD as the second diagnosis. Thirty five participants with only ADHD, 18 participants with ASD and 15 control (typically developing) children underwent the prepared experiment. Unfortunately, the eye tracker measurements of some participants failed, due to calibration defects or the size of the eye tracker did not fit on some participants' faces. Therefore, we could not use data for those participants. Finally, 12 participants with complete data were selected for ASD, 12 participants with ADHD and 10 participants for the control group. In Table 1, for females and males, mean and standard deviation of participants' ages are presented. All participants had an IQ score of above 70. Also, those in the ASD group who were fluent in speech and able to read and write were included in the study. The criteria in Diagnostic Statistical Manual- IV-R (DSM-IV-R) (American Psychiatric Association, A. (2000) was used for the ASD and ADHD diagnosis. Turkish version (Gökler et al., 2004) of Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL) was utilized (Kaufman et al., 1997) for the diagnosis. Also, since the participants wore eye trackers, their visual acuity was examined by an ophthalmologist in the university hospital before the experiment. Experiments were conducted after the approval of the Marmara University Medical Faculty Ethical Advisory Board (Protocol code no: 09.2014.0194, reference: 70737436-050.06.04-140023995). Parental consent forms were read and signed by the parents of the participants.

Table 1. Demographics of the Participants.

	Female			Male		
	count	mean_age	std_age	count	mean_age	std_age
ADHD	5	9.20	1.17	7	9.21	0.99
ASD	3	10	0.82	9	11	1.89
Control	4	9.50	1.12	6	9.25	1.56

3.2 Experimental Design

The stimuli used in the experiments were presented to the participants by a web application called TrackEmo (Teker, 2015). It consists of 40 emotive human face images from the Cohn-Kanade database (Kanade et al., 2000).

The participants were shown two different scenes in the experimental phase of the TrackEmo. As seen in Figure 1a the first one is the empty scene, which represents the transition between choices scene and the next image scene. The second one is a human face scene (Figure 1b). The emotive (one of these emotions; angry, fear, happy, neutral and sad) face images are shown in this scene. The participants try to understand the emotion of these faces.

The stimuli consist of two phases; warming up for emotion recognition and the actual emotion recognition phase. The details are as follows:

Warm-up Phase: This phase was prepared in order to familiarize the participants with the main experiment. The images of the human faces were shown to the participants. The Cohn-Kanade database formed images of the face, ear, mouth, and nose regions, which are clear and contains some emotions. In this phase, the participants were asked the emotions of the five images.

a) b)

Figure 1. Parts of the TrackEmo User Interface: a) Empty Scene b) Human Face with an Emotional Expression.

Emotion Recognition Phase: This was the main part of the experiment. It was similar to the warm up phase, but in this phase, 40 images were shown to the participants. By looking at these images, whether participants were able to recognize the emotions or not and the duration of the process were measured. The images shown at each step were chosen randomly. However, the orders of the images were the same for all participants. There are sixteen angry, five sad, six fear, seven neutral, and six happy images. In order to increase the emotional empathy, negative emotions were used more often (Kirchner et al., 2011).

3.3 Procedure of the Experiment

The experiment was set up in a light-lit room. The stimuli were shown to the participants through a 17-inch LCD monitor. We used SMI Eye Tracking Glasses, which are worn as ordinary glasses. Before using the eye tracker, it was calibrated for each individual. The procedures carried out during the experimental procedure were as follows:

1. The doctors and psychiatrists did the participants' checks and tests.
2. The eligible participants and their families were informed about the experiments by the doctors, and if a participant accepted to participate in the experiments as a volunteer, his/her family signed the consent form.
3. The participant and family were put in a room where the experiment setup had been prepared.
4. The experimental setup and the experiment routine were explained to the participant.
5. Participant's age, gender, hunger, and tiredness information was recorded.
6. The participant sat in the prepared armchair.
7. Participant's ID was entered into the system.
8. Eye tracker was put on to the participant and calibration was done.
9. The experiment started with the emotion recognition warm-up phase, where the participants were asked to recognize the emotions on five images. Their response and response latency were recorded.
10. The next step was the emotion recognition phase, and it was the same as the former step, the only difference was 40 images were shown to the participant.

When the experiment was over, the eye tracker was taken off the participant, and he/she was thanked for their participation.

3.4 Dataset Acquisition

During the experiments, application log data and eye tracker data were gathered. We used these multimodal data for the analyses.

Application Log Data: Application log (APL) data consisted of response and RT of the participants. During the experimental phase, while participants were shown emotive face images, they were asked "What is the emotion of this woman or man?" Their answers were called as RC (response correct) data, and duration of the response was saved as RT. Radwan & Cataltepe, (2017) were used RC and RT data types. While we tried to diagnose the participants by using RC and other features, they aimed to estimate RC value of the participants with ASD.

Eye tracker data: According to Scinto et al., 1994 and Lagun et al., 2011 studies, investigating the eye fixation data is enough for cognitive researches instead of employing all eye movement data. We collected eye fixation data by using SMI Eye Tracking Glasses. It records fixation event data at 30 Hz sampling rate. Fixations are quick eye movements that show the points where a person has focused for a while. The fixation detection algorithm, which is used in the SMI Eye Tracking Glasses, figures out the sequential points in certain dispersion. Possible fixation points are checked by using a moving window, this window first spans a minimum number of points. Then in order to analyze the dispersion of the points, which are in the window, maximum and minimum x and y coordinates values of the points are subtracted from each other and difference values are summed. Namely, dispersion is calculated as follows:

$$D = [\max(x) - \min(x)] + [\max(y) - \min(y)] \qquad (1)$$

If the dispersion value is smaller than a determined maximum dispersion value, the window indicates a fixation, otherwise it does not and the window advances one point to the right. SMI Eye Tracking Glasses produces fixation duration, position X and position Y of fixation points on the stimulus, average pupil size X and Y coordinates in pixels, average pupil diameter in millimeters, dispersion X and dispersion Y coordinates of the fixation. These were employed as fixation features in the current work. The fixation points, which occurred when the participants were looking out of the screen, were excluded from the analysis. In this work, each participant was shown 40 images while they were wearing the eye tracking glasses. The eye tracker recorded the fixation event data while the participants were looking at the images. For each image, each participant had more than one fixation data point. Since each person had different eye gaze behavior, the number of fixation points that occurred while they were looking at the images was not consistent. In this work, we obtained 5990 fixation points from 12 participants with ASD, 4897 fixation points from 12 participants with ADHD, 3813 fixation points from 10 typically developing (control) participants. We referred to the fixation data as raw fixation (RF) data in the experiments.

In order to scale the feature values into the range [0, 1] the min-max normalization method was used. For feature x, it is calculated as follows:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \qquad (2)$$

where x_i represents instance i of the feature x and z_i is normalized value of the x_i. $\min(x)$, and $\max(x)$ stands for minimum value of the x feature and maximum value of the x feature, respectively. Min-max normalization values of the following features were calculated as; position X and position Y of the fixation points, average pupil size X and Y coordinates, and average pupil diameter. In order to make the data more stable, new features such as average and standard deviation of normalized features were generated. Since participants were demonstrated 40 emotive images, average and standard deviation value of the fixation data for each image was used as an instance for each participant. For example, if a participant had 15 fixation points for the second image, then an average and standard deviation of these fixation points were used as an instance for him/her. We called this data type as updated fixation data. In order to increase the feature size, we merged updated fixation data and APL data and also image id, emotion id, emotion level id and we called it an Eye tracker log (ET_log). ET_log consisted of 40 data points for each participant and 17 features (see Table 4 for the features).

4 METHOD

WE used Random forest (Breiman, 2001), Support Vector Machine (SVM) (Cortes & Vapnik, 1995) and Logistic Regression (LR) (Cox, 1958) classifiers on the raw fixation (RF) and ET_log data (Figure 2). Three different classification problems were considered; ASD vs. Control (ASD/Control), ADHD vs. Control (ADHD/Control), ADHD vs. ASD (ADHD/ASD).

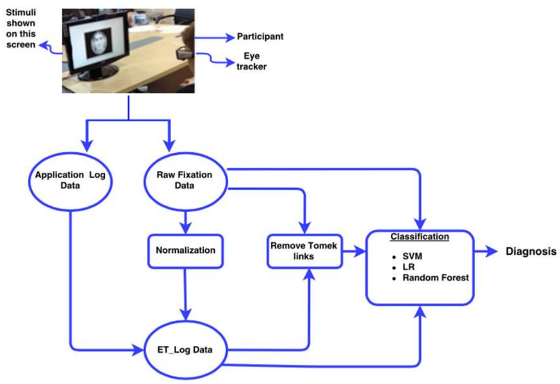

Figure 2. Experimental Setup and Workflow of the Presented Approach for Classifying the Participants

4.1 Random Forest

Random forest classifier consists of the combination of many decision tree classifiers on different subsamples of the training set. In the training phase, the decision tree algorithm learns simple decision rules by using the most distinctive features of the training data. According to these rules, data is split into two parts until the value of the maximum depth parameter is reached. In this work, we chose the maximum depth of the tree as 5 to avoid overfitting (Alpaydin, 2014). We used entropy as the split criteria to measure the quality of the splits. The split operation is processed by considering the entropy measurement of the features. An entropy of zero shows that instances could be separated from each other entirely by using the selected feature. Otherwise, it means that the selected feature is not good enough to split the data, so another feature is chosen to divide the data until the maximum depth of the tree or a leaf is reached. After decision rules are extracted for each decision tree, test participants are classified according to the majority voting score of all decision trees in the forest.

4.2 Support Vector Machines

We used SVM with a linear kernel in this work. SVM uses hyperplanes to separate two different groups from each other by drawing a line between them (Alpaydin, 2014). The hyperplane should be positioned in the space where training data points of two groups are far from each other. For example, in ASD/Control classification, the hyperplane tried to separate ASD (labeled as -1) and control (labeled as +1) data samples. When a new data point needs to be classified, if it falls on -1 side of the hyperplane it is classified as ASD, or if it falls on +1 side it is classified as the control. There is parameter C to regularize the misclassification rate of the SVM. If the value of C is large, SVM uses a smaller-margin hyperplane. Otherwise, if the value of C is small, SVM chooses a larger-margin hyperplane. The selection of C value is dependent on the dataset. According to the grid-search results, we decided the value of C as 1 for ASD/Control and ADHD/ASD classification problems; but for ADHD/Control classification problem the best results were obtained when C is assigned to 10. Also we evaluated the radial basis function (RBF) kernel (Buhmann, 2003). However, the linear kernel outperformed the RBF kernel for our dataset. Also, RBF required more computational cost. For these reasons, we chose the linear kernel for SVM.

4.3 Logistic Regression

Logistic regression (LR) is a simple binary classification algorithm. Since we performed binary classification for diagnosing the participants, we used this algorithm. It investigates the probability that a given sample belongs to class "A" or the probability that it belongs to class "B". For our study, if y is the diagnosis label of the participants, it consists of ADHD and the control if the problem is ADHD/Control classification. LR algorithm learns this function for the classification:

$$P(y|x) = \frac{1}{1+\exp(-w^T x)} \equiv \sigma(w^T x) \tag{3}$$

where $P(y|x)$ computes the posterior probability of elements of y, and x represents the feature set of the dataset (Bishop, 2007). This computation can be written as a logistic sigmoid function $\sigma(w^T x)$. Sigmoid function scales the values between [0, 1], thus we obtain a probability value. The w parameters of the LR are determined by the maximum likelihood method. A threshold value is selected for LR; we chose it as 0.5 for this work. The calculated probability value is compared with the threshold and then the class of data instances is determined.

4.4 Pre-processing

For the pre-processing step, we used the Tomek links method, which was developed by Ivan Tomek (Tomek, 1976) and was used in many research studies as a data cleaning technique (Elhassan et al., 2016, Wang et al., 2017). In the Tomek links method, one sample, x_i is chosen from class i, one sample x_j is chosen from class j. The Euclidean distance between x_i and x_j, $d(x_i, x_j)$ is measured. If there is no sample x_k that provides these cases, $d(x_i, x_k) < d(x_i, x_j)$ or $d(x_j, x_k) < d(x_i, x_j)$, then (x_i, x_j) is named as a the Tomek link, which represents a noisy instance. By using the Tomek links method, we removed noisy instances from the raw fixation and ET_log data (Figure 2).

4.5 Feature Ranking

Feature selection is a main issue to create decision trees. Generally, feature selection is performed by defining the importance of the features. One of the most used feature importance measurements are entropy (Quinlan 1986), Information Gain Ratio (Quinlan 1993) and Gini Index (Breiman et al., 1984). We used entropy as a feature importance measure in this work. For node m, N_m instances reach m, N_m^i, which belong to C_i. The probability of class C_i given x instance reaches node m calculated as follows:

$$\hat{P}(C_i|x,m) \equiv p_m^i = \frac{N_m^i}{N_m} \quad (4)$$

If p_m^i is 0, none of the samples arriving at node m are in class C_i. On the other hand if p_m^i is 1, it means all of the samples arriving at node m are in class C_i. Given K classes, the entropy impurity measure is defined as:

$$I_m = -\sum_{i=1}^{K} p_m^i log_2 p_m^i \quad (5)$$

For each feature, the feature importance is calculated as follows (Pedregosa et al., 2011):

$$w_m I_m - w_{left(m)} I_{left(m)} - w_{right(m)} I_{right(m)} \quad (6)$$

where n_m represents the importance of node m, w_m substitute the weighted number of instances in node m , I_m is entropy impurity calculated in equation (5), the children nodes of the node m are represented as subscript $left$ and $right$. Finally the feature importance value of feature f is calculated as follows:

$$feature\_importance_f = \frac{\sum_{m:node\ m\ splits\ on\ feature\ f} n_m}{\sum_{s \in all\ nodes} n_s} \quad (7)$$

In Figure 3 the feature importance values of the raw fixation data are presented. For ASD/Control, ADHD/Control, and ADHD/ASD classification problems, the average pupil size X coordinates in pixels, average pupil diameter in millimeters, position Y of fixation points on the stimulus and average pupil size Y coordinates in pixels are among the top three raw fixation data features. In Figure 4, the ET_log data features importance values for the Random forest algorithm are illustrated. The features of the ET_log data are shown in Table 4. Interestingly, the features of the APL data have lower feature importance value when compared to the fixation data features.

5 RESULTS AND ANALYSIS

IN this section we present data analysis and method analysis results.

5.1 Data Analysis

During the experiments, the participants were shown human face images with emotional expression and were asked "What is the emotion of this person?" For each participant group, we tried to understand whether their answer to the question agreed with that of the other participants who were in the same group by chance or not. We used Fleiss' kappa (Fleiss, 1971) to evaluate the degree of agreement between two or more participants' answers. Fleiss's kappa is a generalized version of the kappa statistical measurement (Fleiss, 1971). Kappa and weighted kappa measures limited only two raters to rate subjects. However, more than two raters able to rate the subjects in Fleiss' kappa.

In order to evaluate the kappa values, we used the interpretation methodology of (Landis & Koch, 1997). We measured the reliability of the agreement between the participants by using their response to the 40 emotive images in five different categories (angry, fear, happy, sad, neutral). The Fleiss' kappa value for the control group was 0.41, which can be interpreted as a moderate agreement. For the ADHD group the Fleiss' kappa value was 0.35, which showed fair agreement, for the ASD group 0.23 it pointed fair agreement again. Different from the ADHD and ASD groups, there was moderate agreement among participants in the control group. Although participants with ADHD had higher kappa values than ASD, both of them had a fair agreement degree. We generated a random dataset that included random responses to the images and measured the Fleiss' kappa value for that random group. The Fleiss' kappa for the random group was 0.0031, which is significantly lower than the Fleiss' kappa for all the classes.

The statistical differences were measured by one-way ANOVA, which is used to compare the means of two or more groups. In Table 2 and Table 3, we presented the statistical differences between groups in terms of RC and RT. According to the results, participants with ASD correctly recognized the fewer number of questions than the control group for all emotions. Also, RT of the ASD group was longer than the control group. Participants with ADHD perceived the happy emotion a little more accurately and sooner than the control group. Besides this, the RT of the ADHD group was shorter than the control group, with similar accuracy in angry emotions. ASD group performed better than ADHD group in recognition emotions of sadness and fear. However, they were always worse than ADHD group for RT. As a result, the RC behavior of the participant groups did not indicate statistically significant differences except for fear emotion. On the contrary, the RT was a distinguishing factor, especially for ASD group.

Table 2. Mean (M), Standard Deviation (SD) and F-test Value of the RC of the Participant Groups.

	RC					
	ASD	ADHD	Control			
	M (SD)	M (SD)	M (SD)	F test	p value	Post-hoc contrast*
Angry	5.25 (2.01)	6.08 (1.44)	6.08 (1.16)	-	-	-
Fear	0.75 (0.75)	0.50 (0.90)	1.42 (1.38)	3.63	0.028	ASD,ADHD<C
Happy	4.67 (1.61)	5.08 (0.79)	5.00 (0.60)	-	-	-
Neutral	1.92 (2.27)	3.17 (2.52)	3.17 (2.55)	-	-	-
Sad	3.08 (1.08)	2.75 (0.75)	3.25 (0.75)	-	-	-

*Notes: C=Control group; *For Post-hoc tests Bonferroni was used, p<0.05*

Table 3. Mean (M), Standard Deviation (SD) and F-test Value of the RT in Seconds of the Participant Groups.

	RT					
	ASD	ADHD	Control			
	M (SD)	M (SD)	M (SD)	F-test	p value	Post-hoc contrast*
Angry	5.49 (1.69)	4.45 (0.79)	4.52 (0.80)	7.12	<0.001	ADHD<C<ASD
Fear	7.10 (3.58)	5.05 (1.66)	5.54 (1.53)	4.69	<0.001	ADHD, C<ASD
Happy	4.19 (0.92)	3.50 (1.16)	3.36 (0.42)	5.52	0.005	ADHD, C<ASD
Neutral	6.71 (2.21)	4.92 (1.04)	4.34 (0.62)	16.18	<0.001	C<ADHD<ASD
Sad	7.16 (5.97)	5.16 (1.53)	4.27 (0.92)	5.60	0.004	ADHD, C<ASD

*Notes: C=Control group; *For Post-hoc tests Bonferroni was used, $p<0.05$*

The emotion recognition confusion matrix allows us to present how the participants confused the emotions during the experiment. With regards to Table 6, ADHD and the control groups have a similar pattern of correct answers. However, participants with ADHD responded with the questions as unknown (22%) more often than control (16%) and ASD (13%). For all participant groups the most confused emotion was fear. Principally, the ASD group answered the fear emotional image as happy and sad mostly. The correct recognition ratio of the sadness emotion was higher for ASD than ADHD. On the other hand, the most successfully recognized and the least confused emotion was happiness. The images that presented the neutral emotion were difficult for the participants. More frequently neutral images were responded as sad by participants with ASD. Participants with ADHD and the control group confused neutral images with sadness, but mostly they said they did not understand the emotion.

The pupil size of the participants was measured by the eye tracker while the participants were looking at different emotionally expressed images. The statistical differences between the participants' pupil diameter were measured with one-way ANOVA. In Table 5, F-test and p-value of the groups are presented. The pupil diameter has a statistically significant effect on distinguishing the control and ADHD groups. The alpha value was selected as 0.05, so the results were significant at the 5% significance level. For sad images, $p=0.076>0.05$, therefore the pupil diameter difference between the participant groups on sad images was not statistically significant.

The RT of the participants from three groups is shown in Figure 5. For some images, participants in different groups showed similar RT behavior. In general, the participants with ASD spent more time on the images. One-way ANOVA analysis indicated that there were statistically significant RT differences between the groups, $F(2,1437) = 29.831$, $p<0.0001$. As illustrated in Figure 5 RT of the participants with ASD is longer than the ADHD, $F(1,958) = 37.932$, $p<0.0001$. Also, the RT behavior of the control group is significantly different from the ASD group $F(1,958) = 31.562$, $p<0.0001$. On the other hand, the participants generally spent less time towards the end of the experiment, which can be interpreted as exhaustion at the end of the experiment.

5.2 Results

5.2.1 Evaluation Metrics

In the previous sections, we analyzed the effect of fixation and APL data on participants' emotional recognition process. In this section, we used these features as inputs to the Random forest, LR and SVM classification algorithms. For the evaluation of the classification process, leave-one-out cross-validation methodology (Alpaydin, 2014) was used.

All data instances of a participant were chosen as a test data, and the rest of the participants' data instances were used to train the model. This procedure was repeated for all participants. Thus, at each fold, diagnosis of a participant was predicted individually.

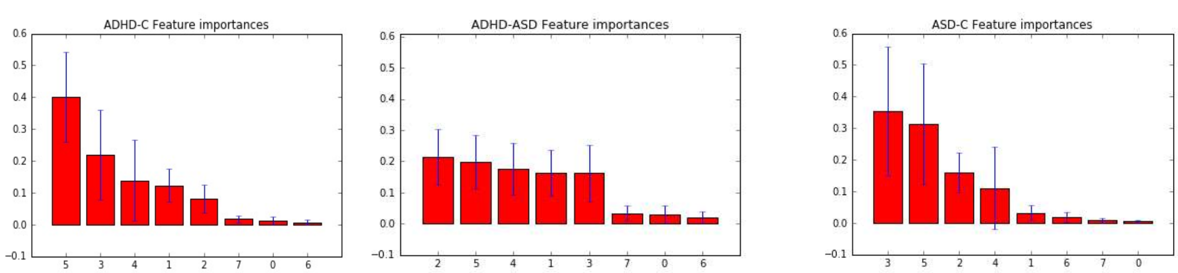

Figure 3. Raw Fixation Data Feature Importance Graphics for ASD/Control (ASD/C), ADHD/Control (ADHD/C), and ADHD/ASD Classification Problems. Y axis Shows Feature Importance Values. X Axis Shows Feature Ids, which are; 0: Fixation Duration, 1: Position X of Fixation Points on the Stimulus, 2: Position Y of Fixation Points on the Stimulus, 3: Average Pupil size X Coordinates in Pixels, 4: Average Pupil Size Y Coordinates in Pixels, 5: Average Pupil Diameter in Millimeters, 6: Dispersion X Coordinates of the Fixation, 7: Dispersion Y Coordinates of the Fixation. The Feature Importance of the Forest is shown by Red Bars and Blue Lines Indicate Feature Importance Variance between Trees of the Random Forest.

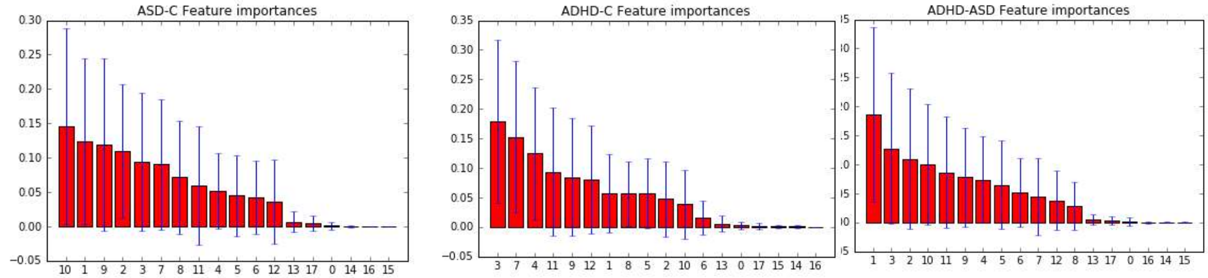

Figure 4. ET_log Data Feature Importance Graphics for ASD/Control (ASD/C), ADHD/Control (ADHD/C) and ADHD/ASD Classification Problems. Y Axis Shows Feature Importance Values. X Axis Shows Feature Ids, which are Presented in Table 4.

Table 4. The Feature Set of the ET_log Data.

Feature ID	Feature name	Feature ID	Feature name
0	number of fixation	9	average of dispersion X coordinates of the fixation
1	average of fixation duration	10	standard deviation of dispersion X coordinates of the fixation
2	standard deviation of fixation duration	11	average of dispersion Y coordinates of the fixation
3	average of normalized position X of fixation points on the stimulus	12	standard deviation of dispersion Y coordinates of the fixation
4	standard deviation of normalized position X of fixation points on the stimulus	13	image id
5	average of normalized position Y of fixation points on the stimulus	14	emotion id
6	standard deviation of normalized position Y of fixation points on the stimulus	15	emotion level
7	average of normalized pupil diameter in millimeters	16	RC
8	standard deviation of normalized pupil diameter in millimeters	17	RT

Table 5. Statistical Test Results for the Pupil Diameter.

	F-test	p-value	Post-hoc contrast*
Angry	9.562	<0.0001	ADHD<ASD<Control
Fear	3.543	0.031	ADHD<Control
Happy	4.022	0.019	ADHD<Control
Neutral	3.999	0.020	ADHD<Control
Sad	0.025	0.076	-

*Notes: *For Post-hoc tests Bonferroni was used, p<0.05*

We measured the classification performance of an algorithm using main evaluation metrics; accuracy, sensitivity and specificity (Alpaydin, 2014). Let TP, TN, FP, and FN show the number of instances that are classified as true positive, true negative, false positive, false negative respectively. In our problem, TP represents the number of correctly classified ASD individuals, and TN indicates the number of accurately detected control individuals in the ASD/Control classification. If a participant with ASD is misclassified FP occurs and if a control participant is incorrectly classified FN occurs. Similar procedures are valid for an ADHD/Control classification problem. On the other hand, for an ADHD/ASD classification problem TP shows correctly classified participants with ADHD and TN denotes correctly classified participants with ASD. Also, FP represents misclassified participants with ADHD and FN stands for misclassified ASD participants. The evaluation formulas are shown below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{9}$$

$$Specificity = \frac{TN}{TN+} \tag{10}$$

For the current work, the accuracy measurement represents the ratio of the correctly classified participants. The classifier produces an accuracy value for each participant and if this value is higher than 50%, we consider this participant is correctly classified. Otherwise, we decide that the diagnosis of the participant is not accurately detected. In our study, sensitivity describes the ratio of the participants with a disorder (ASD or ADHD) who are correctly identified as having the disorder and specificity represents the proportion of typically developing participants who are identified as not having the condition. But for ADHD/ASD problem sensitivity and specificity indicates correctly classified participants with ADHD and ASD, respectively.

Table 6. **Emotion Recognition Confusion Matrix for ASD, ADHD and the Control Groups. Rows Show Actual Emotions; Columns Show the Percentage of the Responses. The Last Column Shows the Percentage for the Unknown Responses.**

ASD actual/response	Angry	Fear	Happy	Neutral	Sad	Unknown
Angry	53	7	6	24	55	17
Fear	3	13	29	17	24	15
Happy	0	1	78	3	11	7
Neutral	4	5	18	27	37	10
Sad	2	7	3	12	62	15
ADHD actual/response	**Angry**	**Fear**	**Happy**	**Neutral**	**Sad**	**Unknown**
Angry	61	7	1	31	33	29
Fear	4	8	29	26	4	28
Happy	1	1	85	8	0	4
Neutral	4	0	10	45	15	26
Sad	3	3	2	13	55	23
Control actual/response	**Angry**	**Fear**	**Happy**	**Neutral**	**Sad**	**Unknown**
Angry	61	2	1	38	42	18
Fear	0	24	25	25	4	22
Happy	0	1	83	7	1	7
Neutral	0	1	8	45	20	25
Sad	2	3	2	20	65	8

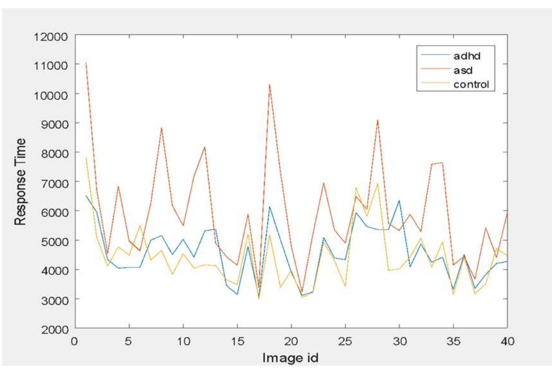

Figure 5. **Average RT Distribution for Each Shown Image of the Participant Groups.**

We implemented the data pre-processing operation and classification algorithms by using scikit-learn v0.19.1 (Pedregosa et al., 2011) machine learning library and our in-house Matlab and python scripts.

5.2.2 Classification Results

The average results over 50 runs of the Random forest, LR and SVM algorithms on the RF (raw fixation) and ET_log (Eye tracker log) feature sets are shown in Table 7, Table 8 and Table 9 for ASD/Control, ADHD/Control, and ADHD/ASD classification problems, respectively. In order to boost the overall performance, besides RF and ET_log features by themselves, RF + ET_log results were produced by taking the average of the classification results of these features (Alpaydın, 2014).

According to Table 7, the highest accuracy (86.36%) and sensitivity (100%) results for ASD/Control classification are achieved by using Random forest on the dataset with removing the Tomek links from the RF and ET_log (RF + ET_log). That means all participants with ASD correctly diagnosed as ASD in this classification. On the other hand, SVM achieved 80% specificity value on ET_log data. By removing the Tomek links from the datasets improved the performance of all classifiers except the sensitivity value for the SVM on the RF data.

As shown in Table 8, for ADHD/Control classification, Random forest outperformed the LR and SVM. The entire control group was correctly detected when the Random forest was used together with the Tomek link removal on the ET_log data. The highest sensitivity value of ADHD/Control was achieved on the RF + ET_log.

ADHD/ASD classification performances are shown in Table 9. The Tomek link did not affect the Random forest performance on the RF data, with and without Tomek links the highest specificity ratio was obtained as 83.33%. Participants with ASD and participants with ADHD have some similar characteristics regarding reactions to other people's emotions. Hence the ADHD/ASD classification problem was the hardest problem among the others. Therefore, we could only reach 70.83% accuracy rate.

When the classifier results were considered, the ASD and ADHD groups were successfully separated from the control group. The Random forest algorithm performed better than LR and SVM. The highest accuracy scores were 86.36% for ASD/Control, 81.82% for ADHD/Control and 70.83% for ADHD/ASD classifications by using the Random forest algorithm. Since ASD and ADHD groups have some similarities such as social perception deficits and emotion recognition deficits (Yerys et al., 2009, Taurines et al., 2012), they could not be distinguished from each other as well as the control group. Also, using Tomek links for outlier instance removal and the combination of the results of the feature sets increased the classification performance. When we compared the performances of the SVM and LR, we found out that, LR achieved higher sensitivity, specificity and accuracy results for each feature type. Besides, removal of the Tomek links from the ET_log data resulted in a specificity value of 100% for the ADHD/Control classification when the Random forest was used. Similarly, when the Tomek links were eliminated, the results of RF + ET_log labeled all participants with ASD correctly during ASD/Control classification. It indicates that if the Tomek links are eliminated and the RF and ET_log results are fused, the classification performance of the ASD/control would improve.

6 DISCUSSION

IN the current study we concentrated on finding the differences of the participant groups for the emotion recognition behavior. We presented a classification framework for ASD and ADHD diagnoses by using the emotion recognition behavior data.

6.1 Contribution of RC and RT

The feature selection algorithm used by Random forest found that the feature rank of the RC feature was lower than the raw fixation and RT features. This result was consistent with the fact that RC values generally did not have a statistical difference between the participants except for the fear emotion (see Table 2). On the other hand, the fact that the RT value was statistically significant (see Table 3) led to this feature being ahead of other APL features during the feature selection operation. Supporting our conclusion, Bal et al., (2010) found that though RC had no statistically significant difference between participants with ASD and the control group, they indicated that participants with ASD were slower than the control group in emotion recognition. Also, in Berggren et al., (2016) ADHD, ASD and the control groups were analyzed in terms of their ability of facial affect recognition. According to their experimental results, ASD group gave more incorrect answers, and their RT value was longer than the control group. The RT of the ADHD group was shorter than the ASD group and was similar to the control group. Therefore, we conclude that our findings are consistent with previous work.

6.2 Contribution of the Pupil Diameter

Through the eye tracker that participants wore during the experiment, the differences between the pupil diameter data of the participant groups were observed. The average normalized pupil diameters of the participants with ADHD were less than the participants with ASD and the control group while they were looking at angry emotions ($p<0.0001$). The pupil size of the participants with ASD was smaller than the control group (see Figure 6). This result is consistent with findings of Martineau et al., (2011) in which the authors found smaller pupil diameters in participants with ASD while they were shown neutral faces, avatars and objects.

6.3 Feature Ranking

Random forest algorithm ranks the features according to their importance score. As seen in Figure 3, when the Random forest algorithm used raw fixation data for the classification, the pupil diameter features have a higher rank. The other data type ET_log data consists of the normalized fixation data and the application log data. Random forest algorithm used fixation features of the ET_log data as the most distinguishing features (see Figure 4). The RC and RT features have lower feature importance value when compared to the fixation data features. One reason for this situation is that ASD and ADHD groups had a fair agreement (according to the Fleiss' kappa value, see Section 5.1) in their responses to the images within themselves. However, the participants of the control group had a moderate agreement. Moreover, the RC feature could not discriminate the participant groups statistically. Therefore, the RC was not a distinguishing factor among the other features for the participant groups except the fear emotion. But the RT feature indicated a statistically significant difference between the participant groups, thus the RT feature had a higher feature importance rank than RC (see Section 4.2).

The statistical test results for the pupil diameter in Table 5 confirms that the pupil diameter of the participant with ADHD and the control group has a statistically distinguishing effect. In Lagun et al., (2011) they performed classification of control subjects from mild cognitive impairment subjects by using the eye tracker fixation and saccade data. They claimed that, although the pupil diameter did not increase the classification accuracy, fixation duration improved their performance. In our study, both of them are the distinguishing features.

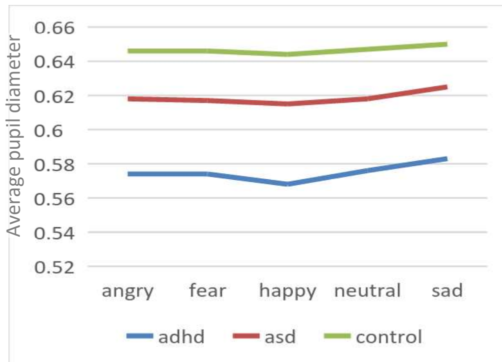

Figure 6. Average Normalized Pupil Diameter (mm) of the Participant Groups for when they were Shown Different Emotion Types.

6.4 Limitations of the Study

Eye tracker mobile glasses are worn like regular glasses; however, the size of the eye tracker did not always fit on some participants' faces. Also, some participants with ASD would not have worn the mobile eye tracker and could not follow the instructions of the experiment. Therefore, we could not measure these participants' data and did not include them in our study, which reduced the dataset size. If we could gather more data, we could have obtained more accurate results. Instead of using wearable eye trackers, the use of remote eye trackers could be an option for further studies.

This study included ASD and ADHD participants that were above an IQ of 70, and that could both read and write. User interfaces are other than what we used (TrackEmo), allowing emotion detection responses of participants without the ability to read, would allow inclusion of more participants. Also, the vision criteria we used also might have imposed limitations on the participants that we could include and hence the results that we arrived.

7 CONCLUSION

PURPOSE of this work was (1) investigating the differences in the emotion recognition process between ADHD, ASD and control groups; and (2) to perform statistical analysis and classification of participants based on the eye tracker and application log data.

Table 7. Classification Results for ASD/Control(C). The Abbreviations used in the Table are the following: Sensit.: Sensitivity, Specif.: Specificity, Acc.: Accuracy, RF: Raw Fixation, ET_log: Eye Tracker Log Data, RF + ET_log: Average Combination of RF and ET_log, LR: Logistic Regression, SVM: Support Vector Machine.

ASD/C	Data type								
	RF			ET_log			RF + ET_log		
Classifiers	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
Random Forest	91.67	60.00	77.27	50.00	30.00	40.91	91.67	40.00	68.18
Random Forest_Tomeklink	91.67	60.00	77.27	66.67	50.00	59.09	100.00	70.00	86.36
LR	91.67	60.00	77.27	50.00	40.00	45.45	58.33	60.00	59.09
LR_Tomeklink	91.67	60.00	77.27	50.00	50.00	50.00	83.33	70.00	77.27
SVM	83.33	30.00	59.09	33.33	80.00	54.55	41.67	40.00	40.91
SVM_Tomeklink	75.00	50.00	63.64	41.67	60.00	50.00	50.00	50.00	50.00

Table 8. Classification Results for ADHD/Control(C). The Abbreviations used in the Table are the following: Sensit.: Sensitivity, Specif.: Specificity, Acc.: Accuracy, RF: Raw Fixation, ET_log: Eye Tracker Log Data, RF + ET_log: Average Combination of RF and ET_log, LR: Logistic Regression, SVM: Support Vector Machine.

ADHD/C	Data type								
	RF			ET_log			RF + ET_log		
Classifiers	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
Random Forest	83.33	60.00	72.73	58.33	70.00	63.64	91.67	40.00	68.18
Random Forest_Tomeklink	83.33	60.00	72.73	66.67	100.00	81.82	83.33	80.00	81.82
LR	58.33	60.00	59.09	50.00	50.00	50.00	66.67	60.00	63.64
LR_Tomeklink	58.33	60.00	59.09	58.33	70.00	63.64	66.67	60.00	63.64
SVM	58.33	40.00	50.00	50.00	60.00	54.55	58.33	40.00	50.00
SVM_Tomeklink	41.67	40.00	40.91	50.00	60.00	54.55	58.33	40.00	50.00

Table 9. Classification Results for ADHD/ASD. The Abbreviations used in the Table are the following: Sensit.: Sensitivity, Specif.: Specificity, Acc.: Accuracy, RF: Raw Fixation, ET_log: Eye Tracker Log Data, RF + ET_log: Average Combination of RF and ET_log, LR: Logistic Regression, SVM: Support Vector Machine.

ADHD/ASD	Data type								
	RF			ET_log			RF + ET_log		
Classifiers	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
Random Forest	41.67	83.33	62.50	66.67	66.67	66.67	75.00	66.67	70.83
Random Forest_Tomeklink	41.67	83.33	62.50	66.67	66.67	66.67	66.67	66.67	66.67
LR	41.67	50.00	45.83	58.33	66.67	62.50	50.00	58.33	54.17
LR_Tomeklink	50.00	58.33	54.17	58.33	66.67	62.50	50.00	58.33	54.17
SVM	41.67	50.00	45.83	66.67	41.67	45.45	50.00	41.67	50.00
SVM_Tomeklink	41.67	41.67	41.67	75.00	50.00	62.50	66.67	41.67	54.17

We collected data from the participants with ASD, participants with ADHD and the control group by using the SMI Eye Tracking Glasses and the TrackEmo software program. In order to find distinguishing factors for the participant groups the obtained data were analyzed by using the ANOVA statistical analysis method. Besides we used three classification algorithms to be able to classify the participant groups.

The experiments in this study indicated that the usage of APL and fixation data were promising for distinguishing the ASD, ADHD and the control groups from each other. The eye tracker pupil diameter and eye gaze proved to be very informative about the emotion recognition behavior of the participants. We presented the statistical significance analysis of different factors, such as pupil diameter, RC, and RT. We used various machine learning techniques, such as outlier removal using the Tomek links, feature relevance and Random forest, Logistic Regression, SVM classifiers to classify the participants into different categories. We achieved the best accuracy results by using the Random forest algorithm. Having balanced the value for specificity and sensitivity is important. Therefore, distributions of specificity and sensitivity values (see Table 7, Table 8, and Table 9) became balanced especially for ASD/Control and

ADHD/Control classification when results of the RF and ET_log were combined.

For the dataset, we collected certain conclusions on different types of inputs and emotions, namely:

- Verification that pupil diameter is an important feature for classification of ASD, ADHD and the control groups.
- Recognition of fear emotion is a distinguishing factor between the control and other groups.
- For all emotions, response time of the participants with ASD is the highest among the other groups.
- The confusion rate of the fear emotion with sad emotion is high for participants with ASD.
- Recognition of happiness is not a distinguishing factor for ASD, ADHD and control groups.
- Participants with ASD frequently responded to neutral images as sad.
- Participants showed similar response time behavior to the same type of emotive images.
- According to the Fleiss' kappa agreement measurement, the participants of the control group gave more consistent responses to emotional images than the other groups.

With further studies on larger datasets, we believe that our efforts will lead to methods and tools that help psychiatrists with their clinical diagnosis and treatment planning and provide benefit to the psychologists and teachers during the process of individualized education for each patient.

8 ACKNOWLEDGEMENT

WE would like to thank the ITU Computer Engineering Department Kullab for usage of the eye tracker, Garayli, Esra Siler, Emel Aksu, Idil Oguz for their help in data acquisition.

9 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

10 REFERENCES

E. Alpaydin. (2014). *Introduction to machine learning*. MIT press.

American Psychiatric Association, A. (2000). *Diagnostic and statistical manual of mental disorder* (4th ed.). Washington DC: American Psychiatric Association.

E. Bal, Harden, E., Lamb, D., Van Hecke, A. V., Denver, J. W., & Porges, S. W. (2010). Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of autism and developmental disorders*, 40(3), 358-370.

S. Baltaci & Gokcay, D. (2016). Stress Detection in Human–Computer Interaction: Fusion of Pupil Dilation and Facial Temperature Features. *International Journal of Human–Computer Interaction*, 32(12), 956-966.

S. Berggren, Engström, A. C., & Bölte, S. (2016). Facial affect recognition in autism, ADHD and typical development. *Cognitive neuropsychiatry*, 21 (3), 213-227.

C. Bishop. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st edn. 2006. corr. 2nd printing edn. Springer, New York.

L. Breiman. (2001). Random forests. *Machine learning*, 45(1), 5-32.

M. D. Buhmann. (2003). Radial basis functions: theory and implementations (Vol. 12). *Cambridge university press*.

C. Caffrey. (2008). Using pupillometric, fixation-based and subjective measures to measure the processing effort experienced when viewing subtitled TV anime with pop-up gloss. *Copenhagen studies in language*, (36), 125-144.

J. B. Colby, Rudie, J. D., Brown, J. A., Douglas, P. K., Cohen, M. S., & Shehzad, Z. (2012). Insights into multimodal imaging classification of ADHD. *Frontiers in systems neuroscience*, 6.

C. Cortes, & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

D. R. Cox. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 215-242.

F. Craig, Lamanna, A. L., Margari, F., Matera, E., Simone, M., & Margari, L. (2015). Overlap between autism spectrum disorders and attention deficit hyperactivity disorder: searching for distinctive/common clinical features. *Autism Research*, 8(3), 328-337.

K. M. Dalton, Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., ... & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature neuroscience*, 8(4), 519-526.

E. Delaherche, Chetouani, M., Bigouret, F., Xavier, J., Plaza, M., & Cohen, D. (2013). Assessment of the communicative and coordination skills of children with autism spectrum disorders and typically developing children using social signal processing. *Research in Autism Spectrum Disorders*, 7(6), 741-756.

M. Duda, Ma, R., Haber, N., & Wall, D. P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. Translational psychiatry, 6(2), e732.

C. Ecker, Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E. M., & MRC AIMS Consortium. (2010). Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. *Neuroimage*, 49(1), 44-56.

T. Elhassan, Aljurf, M., Al-Mohanna, F., & Shoukri, M. (2016). Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method. *Journal of Informatics and Data Mining*.

J. L. Fleiss. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5), 378.

M. S. Gold, & Gold, J. R. (1975). Autism and attention: Theoretical considerations and a pilot study using set reaction time. *Child Psychiatry & Human Development*, 6(2), 68-80.

B. Gökler, Ünal, F., Pehlivantürk, B., Kültür, E. Ç., Akdemir, D., & Taner, Y. (2004). Okul Çaği Çocuklari İçin Duygulanim Bozukluklari ve Şizofreni Görüşme Çizelgesi-Şimdi ve Yaşam Boyu Şekli-Türkçe Uyarlamasinin Geçerlik ve Güvenirliği. *Çocuk ve Gençlik Ruh Sağliği Dergisi*.

T. Hayashi, Mizuno-Matsumoto, Y., Kohri, S., Nitta, Y., & Tonoike, M. (2017). Event-related theta and alpha oscillations under emotional stimuli: an MEG study. *Intelligent Automation & Soft Computing*, 23(1), 175-182.

E. Hess. (1975). The tell-tale eye. New York*, NY: Van Nostrand*.

K. T. Hvelplund. (2014). Eye tracking and the translation process: reflections on the analysis and interpretation of eye-tracking data. MonTI. *Monografías de Traducción e Interpretación*, 201-223.

R. Iannaccone, Hauser, T. U., Ball, J., Brandeis, D., Walitza, S., & Brem, S. (2015). Classifying adolescent attention-deficit/hyperactivity disorder (ADHD) based on functional and structural imaging. *European child & adolescent psychiatry*, 1-11.

T. Kanade, Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In Automatic Face and Gesture Recognition, 2000. *Proceedings. Fourth IEEE International Conference on* (pp. 46-53). IEEE.

J. Kaufman, Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., ... & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(7), 980-988.

A. H. Kemp, Silberstein, R. B., Armstrong, S. M., & Nathan, P. J. (2004). Gender differences in the cortical electrophysiological processing of visual emotional stimuli. *NeuroImage*, *21*(2), 632-646.

J. C. Kirchner, Hatri, A., Heekeren, H. R., & Dziobek, I. (2011). Autistic symptomatology, face processing abilities, and eye fixation patterns. *Journal of autism and developmental disorders*, 41(2), 158-167.

S. Kuusikko, Haapsamo, H., Jansson-Verkasalo, E., Hurtig, T., Mattila, M. L., Ebeling, H., ... & Moilanen, I. (2009). ER in children and adolescents with autism spectrum disorders. *Journal of autism and developmental disorders*, 39(6), 938-945.

D. Lagun, Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein, E. (2011). Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of neuroscience methods*, 201(1), 196-203.

J. R. Landis, Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Y. Lin, Zhang, W. J., & Watson, L. G. (2003). Using eye movement parameters for evaluating human–machine interface frameworks under normal control operation and fault detection situations. International Journal of Human-Computer Studies, 59(6), 837-873.

Y. Lin, Zhang, W. J., Wu, C., Yang, G., & Dy, J. (2009). A fuzzy logics clustering approach to computing human attention allocation using eyegaze movement cue. International Journal of Human-Computer Studies, 67(5), 455-463.

C. Lord, Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.

J. Martineau, Hernandez, N., Hiebel, L., Roché, L., Metzger, A., & Bonnet-Brilhault, F. (2011). Can pupil size and pupil responses during visual scanning contribute to the diagnosis of autism spectrum disorder in children? *Journal of psychiatric research*, 45(8), 1077-1082

M. Miyahara, Bray, A., Tsujii, M., Fujita, C., & Sugiyama, T. (2007). Reaction time of facial affect recognition in Asperger's disorder for cartoon and real, static and moving faces. *Child Psychiatry & Human Development*, 38(2), 121-134.

S. M. Myers, & Johnson, C. P. (2007). Management of children with autism spectrum disorders. *Pediatrics*, 120(5), 1162-1182.

T. Partala, & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59(1), 185-198.

F. Pedregosa, Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825-2830.

M. Pedrotti, Mirzaei, M. A., Tedesco, A., Chardonnet, J. R., Mérienne, F., Benedetto, S., & Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction*, 30(3), 220-236.

A. M. Radwan & Cataltepe, Z. (2017). Improving Performance Prediction on Education Data with Noise and Class Imbalance. *Intelligent Automation & Soft Computing*, 1-8.

S. Rivera, Best, C., Yim, H., Walther, D., Sloutsky, V., & Martinez, A. (2012). Automatic selection of eye tracking variables uncovers similar mechanisms for visual categorization in adults and infants. Journal of Vision, 12(9), 1258-1258.

L. F. Scinto, Daffner, K. R., Castro, L., Weintraub, S., Vavrik, M., & Mesulam, M. M. (1994). Impairment of spatially directed attention in patients with probable Alzheimer's disease as measured by eye movements. *Archives of Neurology*, 51(7), 682-688.

R. Taurines, Schwenck, C., Westerwald, E., Sachse, M., Siniatchkin, M., & Freitag, C. (2012). ADHD and autism: differential diagnosis or overlapping traits? A selective review. *ADHD Attention Deficit and Hyperactivity Disorders*, 4(3), 115-139.

U. Teker, (2015). Analysis of ER performance with sensor and log data. Istanbul Technical University Graduation Project.

I. Tomek. (1976). Two modifications of CNN. IEEE Trans. Systems, *Man and Cybernetics*, 6, 769-772.

B. E. Yerys, Wallace, G. L., Sokoloff, J. L., Shook, D. A., James, J. D., & Kenworthy, L. (2009). Attention deficit/hyperactivity disorder symptoms moderate cognition and behavior in children with autism spectrum disorders. *Autism Research*, 2(6), 322-333.

J. N. Van Der Geest, Kemner, C., Verbaten, M. N., & Van Engeland, H. (2002). Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study. *Journal of Child Psychology and Psychiatry,* 43(5),669

Q. Wang, Xin, J., Wu, J., & Zheng, N. (2017, March). SVM Classification of Microaneurysms with Imbalanced Dataset Based on Borderline-SMOTE and Data Cleaning Techniques. *In Ninth International Conference on Machine Vision (pp. 103411S-103411S). International Society for Optics and Photonics.*

11 NOTES ON CONTRIBUTORS

Mahiye Uluyagmur Ozturk is a PhD candidate in Computer Engineering at Istanbul Technical University, also she received her MSc degree from Istanbul Technical University. She currently works as a research and teaching assistant at Computer Engineering Department. Her research interests include machine learning, feature selection, statistical methods for large datasets, data mining, human-machine interaction.

Ayse Rodopman Arman, Prof., MD is graduated from Marmara University School of Medicine, Istanbul, Turkey (1992). She has finished her residency program in Child and Adolescent Psychiatry Department, Marmara University School of Medicine and currently works as an academician and Director of the same department in Istanbul, Turkey. Her main interests are childhood neurodevelopmental disorders.

Gresa Çarkaxhiu Bulut was born in 1984, Kosovo. She completed her medical education in Istanbul University Cerrahpaşa Faculty of Medicine (2008) and residency on Marmara University Faculty of Medicine on Child and Adolescent Psychiatry (2013) in Turkey. In 2011, she worked as a research fellow in Harvard University McLean Hospital. She is maintaining her research and clinical work in Sakarya University Faculty of Medicine as an assistant professor, since 2017.

Onur Tuğçe Poyraz Fındık is a child and adolescent psychiatrist at the Marmara University Faculty of Medicine, Istanbul, Turkey. She graduated at the Cerrahpaşa Faculty of Medicine at the Istanbul University, and completed her specialist training at the Marmara University Faculty of Medicine in Istanbul. Her fields of research are neurodevelopmental disorders and forensic psychiatry.

Herdem Aslan Genç is a child and adolescent psychiatrist. She is currently working at a university hospital as an assistant professor in İstanbul, Turkey. Her main areas of interests include early diagnosis and intervention of neuropsychiatric disorders of early childhood and psychopathologies of adolescence.

Yanki Yazgan, M.D. has taught on neurodevelopment and psychopathology at the Marmara University School of Medicine, Istanbul, and holds an adjunct faculty position at Yale Child Study Center, New Haven. Dr Yazgan has published on multidisciplinary/multicontext assessment of and interventions for the neurodevelopmental disorders (NDD) including ASD and ADHD.

Umut Teker is a Software Spcialist at Anadolu Hayat Emeklilik which is an Insurance Company. He is bachelor of Science in Computer Engineering from Istanbul Technical University. His research Interests are machine learning, big data, Internet of Things and image Processing.

Zehra Cataltepe is a professor at the Istanbul Technical University and also the co-founder of tazi. io, which is a startup company with the purpose of developing online machine learning products. She received her M.Sc. and Ph.D. degrees from Caltech in Computer Science. Her research interests include machine learning algorithms and software, feature selection, anomaly detection, online machine learning, machine learning for human learning, learning on heterogeneous, networked and time series data, and big data.