ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

EMOTION RECOGNITION PROCESS ANALYSIS BY USING EYE TRACKER, SENSOR AND APPLICATION LOG DATA

Ph.D. THESIS

Mahiye ÖZTÜRK

Department of Computer Engineering

Computer Engineering Programme

JANUARY 2019

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

EMOTION RECOGNITION PROCESS ANALYSIS BY USING EYE TRACKER, SENSOR AND APPLICATION LOG DATA

Ph.D. THESIS

Mahiye ÖZTÜRK
(504122507)

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Zehra ÇATALTEPE

JANUARY 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

GÖZ İZLEME CİHAZI, SENSÖR VE UYGULAMA VERİLERİ İLE İNSANLARDA DUYGU TANIMA ANALİZİ

DOKTORA TEZİ

Mahiye ÖZTÜRK
(504122507)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Zehra ÇATALTEPE

OCAK 2019

Mahiye ÖZTÜRK, a Ph.D. student of ITU Graduate School of Science Engineering and Technology 504122507 successfully defended the thesis entitled "EMOTION RECOGNITION PROCESS ANALYSIS BY USING EYE TRACKER, SENSOR AND APPLICATION LOG DATA", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Zehra ÇATALTEPE** Istanbul Technical University	....
Jury Members :	**Assoc. Prof. Dr. Hatice KÖSE** Istanbul Technical University	....
	Prof. Dr. Mustafa Ersel KAMAŞAK Istanbul Technical University	....
	Prof. Dr. Mehmet Barış KORKMAZ Istanbul University	....
	Prof. Dr. Banu DİRİ Yıldız Technical University	....

Date of Submission : 19 December 2018
Date of Defense : 11 January 2019

To my little baby,

FOREWORD

First of all I would like to express my sincere gratitude and deep appreciation to my advisor and guide Prof. Dr. Zehra Çataltepe for her constant encouragement and valuable suggestions. She is one of the most special people I have ever met. She is an example of excellence as a researcher, mentor, instructor, mother and role model. This thesis would not have been possible unless her support.

I would like to thank my thesis committee members Prof. Dr. Barış Korkmaz and Dr. Hatice Köse for their feedback and discussion which indeed improved this thesis. I express my special thanks to Prof. Dr. Eşref Adalı for his guidance and advice throughout my academic life and thanks to him for encouragement and for seeing the potential in me.

This thesis has been supported by ITU Scientific Research Projects (Project Code: 39651). I would like to thank the funding institution for their support.

I would like to thank Marmara University Psychiatry Department members Prof. Dr. Ayşe Rodopman Arman, Prof. Dr. Yankı Yazgan, Dr. Gresa Çarkaxhiu Bulut, Dr. Tuğçe Onur Poyraz Fındık, Dr. Sultan Seval Yılmaz, Dr. Herdem Aslan Genç for their collaborations. Thanks to the students from ITU: Leyla, Umut, Esra, Emel, İdil for their supports during the experiments. I also thank the children who participated to the experiments. I am very grateful to all of you. Also, thanks to ITU Computer Engineering Department Kullab for allowing us to use the mobile eye tracker for my experiments. I would like to show my gratitude to my colleagues Hakan, Gözde, İlknur, Müge, Ahmet, Kübra, Dilara, Çağatay, Resul, Emrullah for caring and supporting me during this long journey.

I am indebted to my family for their endless love and trust in me. Especially, thanks to my mother Nurhan and my sister Zehra for their priceless emotional support and confidence. I would like to thank my father Ahmet for instilling in me the spirit of hardworking, discipline, and tolerance. Also, I am sincerely grateful to my brother Fatih for opening my mind about going after my dreams, I will forever be grateful to him. My special thanks go to my grandparent for their encouragement about starting my graduate education. Finally, thanks to my loving husband Akın, for his love, patience and understanding. He always shows how proud he is of me.

The last word goes for my little baby who will be born in February 2019 inşallah. The presence of her inside me has given me great inspiration for completing this thesis, thus I would like to dedicate it to her.

January 2019 | Mahiye ÖZTÜRK (M.Sc.)

TABLE OF CONTENTS

Page

ABBREVIATIONS

APL	**:**	Application Log
CNN	**:**	Convolutional Neural Network
LSTM	**:**	Long-Short Term Memory
RC	**:**	Response Correct
RT	**:**	Response Time
Sensit.	**:**	Sensitivity
Specif.	**:**	Specificity
Acc.	**:**	Accuracy

LIST OF TABLES

LIST OF FIGURES

Page

EMOTION RECOGNITION PROCESS ANALYSIS BY USING EYE TRACKER, SENSOR AND APPLICATION LOG DATA

SUMMARY

One of the essential components of social interaction is recognizing emotions. Interpreting facial expressions correctly enables easier communication among people. In addition to providing an understanding of other people's intentions accurately and reacting to them appropriately, emotion recognition processes also carry clues about one's overall emotional well-being. 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 neurodevelopmental disorders. Emotion recognition behavior and performance may vary between people with major neurodevelopmental disorders such as ASD, ADHD and the control groups. It is crucial to identify these differences for early diagnosis and individual treatment purposes.

This thesis presents, an analysis of emotion recognition behavior of ASD, ADHD and the control groups. ASD is a complex neurodevelopmental disorder that usually surfaces during the first year of life. Some characteristics of ASD are as follows: Difficulties in social communication and interactions, problems in conducting and sustaining a relationship, difficulty of initiating and keeping 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. On the other hand, symptoms of ADHD are; hyperactivity, impairment of both attention and concentration and impulsivity. Besides, both ASD and ADHD groups show a lack of concern or inability to react to other people's emotions or feelings. 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. Therefore, it is crucial to perform an in-depth analysis of emotion recognition processes and investigate the different individual deficiencies of ASD and ADHD.

The research question motivating our research can be stated as follows: Do the emotion recognition ability and process differ between the clinical groups (ADHD and ASD) and the control group? If so, this difference has potential usage 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. On the other hand, early detection is essential in the success of interventions for both ASD and ADHD. To address such issues, we intend to develop a research design that can help distinguish the participant groups by using emotion recognition data.

In this work, we collaborate with psychiatrists from Marmara University Medical Faculty Hospital Child and Adolescent Psychiatry Outpatient Department. The criteria in Diagnostic Statistical Manual- IV-R (DSM-IV-R) (American Psychiatric Association, A. (2000)) is used for the ASD and ADHD diagnosis. The experiments that we designed 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 before the experiments. Thirty five participants with 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. 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.

Within the scope of this thesis, we aim to distinguish the participants with ASD, participants with ADHD and the control group by using data collected during a set of emotion recognition experiment. We propose an experimental environment where the participants wear an eye tracker and they are shown some emotive facial images as stimuli. Emotional stimuli are used in many studies to measure human perception. In our work, the participants are asked to state the emotion in the presented images. The purpose of the experiment is to understand how participants react to these images, measured by their eye movement and also what their emotion detection responses are for each image. The response, the response time and the eye tracker fixation data are recorded and used for analysis. We present results under two different categories. First, we perform statistical analysis of the differences in the emotional recognition behaviors of participant groups by using one-way ANOVA test. Second, we employ the data obtained during the experiments to classify the participants' diagnosis by using machine learning methods. As machine learning techniques we use not only feature based classifiers such as Random forest, Logistic Regression, SVM classifiers also deep learning techniques such as CNN (Convolutional Neural Network) and LSTM (Long-Short Term Memory).

In order to clean the data from noise, Tomek links removing method is employed with the feature based methods. When we investigate the important features of the fixation and application log data combination, we conclude that fixation features have a higher importance than the application log data which includes response and response time. This result is confirmation of that the eye movement data is more informative than the application log data.

Deep learning techniques enable the production of models with multiple layers of processing to learn the representations of data with multiple abstraction levels. In the established neural network model, the CNN layer is used to expose the abstract representation of the features and LSTM method is used to learn the model of the data. Different time steps options and different integration of eye movement features are employed as inputs to the generated models.

The principal contributions of this work are as follows:

- We report that response, response time and pupil diameter measurements of the participant groups have statistically significant differences.
- We show that using eye movement data and application log data collected during the emotion recognition experiment is crucial to classify ASD, ADHD and the control groups.
- We compare the feature based classifiers and deep learning classification algorithms.
- We propose a deep learning framework incorporating (CNN) and (LSTM) for participant diagnosis classification task based on the emotion recognition behaviors of the participants.

With the support of more studies on larger population sizes and alternative types of inputs, approaches like ours can be used to facilitate early diagnosis and hence enable early treatment processes. 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.

GÖZ İZLEME CİHAZI, SENSÖR VE UYGULAMA VERİLERİ İLE İNSANLARDA DUYGU TANIMA ANALİZİ

ÖZET

Dikkat eksikliği ve hiperaktivite bozukluğu (DEHB) (İng. ADHD – Attention Deficit and Hyperactivity Disorder) olan ve Otizm Spektrum Bozukluğu (OSB) (İng. ASD – Autism Spectrum Disorders) olan çocuklar duygu tanımada yaşıtlarına göre farklılıklar gösterebilmektedir. Bu farklılıklar günlük ve akademik yaşamlarında zorluk yaşamalarına neden olmaktadır. Bu çalışmanın amacı, DEHB'li ve OSB'li çocukların duygu tanıma süreçleri ile tipik gelişen çocukların duygu tanıma süreçlerini karşılaştırmaktır. Eğer gruplar arasında önemli farklar olduğu tespit edilirse, bu farklılıklar ileride teşhiste ayırt edici olarak ya da eğitim amacı ile kullanılabilir.

Çalışmada DEHB'li, OSB'li ve tipik gelişim gösteren 8-12 yaş aralığında katılımcılar yer almaktadır. DEHB'li ve OSB'li katılımcılar Marmara Eğitim ve Araştırma Hastanesi ile Göztepe Eğitim ve Araştırma Hastanesi'nden uzman psikiyatrist doktorların seçtiği, aileleri çalışmaya katılmayı uygun gören çocuklardan ve kendisi ya da ailesinin onayı alınmış tipik gelişen çocuklardan oluşmaktadır. Deneylerde bakmaları gereken bir ekran olduğu için ve bu esnada göz izleme cihazı takmaları gerektiğinden, katılımcıların deneylerden önce göz doktoru tarafından göz muayeneleri yapılmıştır, göz kusuru bulunan kişiler katılımcı olarak alınmamıştır. OSB'li bireyler arasından orta ya da hafif derecede otizm gösteren katılımcılar alınmıştır.

Bu çalışmada DEHB'li, OSB'li ve tipik gelişen çocuklarda duygu tanıma yeteneklerini tespit edebilmek için bir deney ortamı hazırlanmıştır. Deney esnasında katılımcılara mutlu, üzgün, korkmuş, kızgın ve duygusuz olmak üzere 5 farklı tipte duygu ifadesi sergilemiş, tıp doktorları tarafından da onaylanmış Cohn-Kanade veri kümesinden insan resimleri gösterilmiştir ve bu resimlerdeki duygunun ne olduğu katılımcılara sorulmuştur. Resimlerdeki duygu yoğunluğu farklı seviyelerdedir. Bazı resimlerdeki duygular açıkça belli iken bazı resimlerdeki duygular belli belirsizdir. Böylece katılımcıların aynı duygunun farklı yoğunluktaki hallerini tanıyıp tanıyamadıkları test edilmiştir. Deney sırasında katılımcılara 40 adet resim gösterilmiştir. Gösterilecek resmin hangi duygu ve seviyede olacağına rastgele olarak karar verilmiştir ve resimler her katılımcıya aynı sırada gösterilmiştir. Katılımcıların duyguları tanıma yetenekleri yanı sıra, kendilerine yöneltilen "resimdeki duygu nedir" sorusuna ne kadar sürede cevap verdikleri de ölçülmüştür. Ayrıca deney esnasında katılımcılara göz izleme cihazı takılmış ve bu cihaz sayesinde katılımcıların deney esnasında baktıkları yerler tespit edilmiştir. Göz izleme cihazı, bir kullanıcının göz hareketlerine bağlı olarak nereye baktığının anlaşılmasını sağlayan bir alettir. Kullanıcı bu cihazı normal bir gözlük gibi takar.

Bu tezde, duygu tanıma deneyleri esnasında toplanan veriler ile katılımcı gruplarının birbirinden ayırt edilip edilemeyeceği konusu üzerinde durulmuştur. Bu anlamda yapılan çalışma Türkiye'de bir ilk ve kullanılan yöntemler ve hazırlanan veri kümesi

açısından dünyada ilk olma özelliği taşımaktadır. Daha önceki çalışmalarda genelde hazır veri setleri üzerinde çalışılmıştır. Yeni bir deney düzeneği oluşturarak veri toplamak bu çalışmanın yenilikçi kısımlarından biridir.

Elde edilen veriler ile öncelikle hangi duygu tiplerinin katılımcılar üzerinde ayırt edici bir etki oluşturup oluşturmadığına bakılmıştır. Bu analizler için katılımcıların resimlere verdikleri cevaplar ve cevap verme süreleri temel alınmıştır. Cevap ve cevaplama süresi verilerinin istatistiksel olarak anlamlı bir şekilde grupları ayırt edip edemedikleri ANOVA istatiksel analiz metodu kullanılarak tespit edilmiştir. Elde edilen sonuçlara göre korku duygu ifadesine sahip resimlerin normal gelişim gösteren çocuklar tarafından tanınma oranları OSB ve DEHB'li çocuklara göre istatistiksel olarak daha fazladır. Öte yandan mutlu, üzgün, kızgın, nötr duygu ifadeli resimler tüm gruplar tarafından aynı oranda tanınabilmiştir, dolayısıyla ayırt edici olmamışlardır. Cevaplama süresi verisinin ise tüm duygular için katılımcı grupları arasında ayırt edici özelliğe sahip olduğu gözlenmiştir. OSB'li çocuklar, diğer çocuklardan hep daha uzun sürede cevap vermiştir. Sonuç olarak cevaplama süresinin grupları ayırt etmede önemli bir öznitelik olduğu ortaya çıkmıştır.

İstatistiksel analizlerden sonra makine öğrenmesi yöntemleri kullanılarak katılımcıların sınıflandırılması yapılmıştır. Öncelikle sınıflandırma işlemi sadece cevaplar ve cevaplama süresi öznitelikleri kullanılarak yapılmıştır. ANOVA testlerinden elde edilen sonuçlara göre katılımcıların resimlere verdikleri cevaplar, katılımcı grupları için ayırt edici olamamıştır. Bu durum bize, cevap ve cevaplama süresini kullanarak hazırladığımız sınıflandırıcılardan gelen sonuçların, kullandığımız tek-kayıt çıkışlı çapraz doğrulama (İng. leave-one record- out) doğrulama tekniği sayesinde yüksek çıktığını, aslında yeterli olmadığını göstermiştir. Dolayısıyla dördüncü ve beşinci bölümdeki sonuçlar tek-katılımcı çıkışlı çapraz doğrulama (İng. leave-one participant-out) doğrulama tekniği kullanılarak sınanmış ve daha güvenilir sonuçlar elde edilmiştir.

Dördüncü bölümde öznitelik tabanlı sınıflandırıcılardan yararlanılmıştır. Ancak cevap ve cevaplama süresi verilerine ek olarak göz hareketi özniteliği olan sabitleme (İng. fixation) verisi de sınıflandırma işleminde kullanılmıştır. Sabitleme verisi göz izleme cihazının ürettiği, katılımcının odaklandığı noktanın koordinatlarını, baktığı bir noktaya bakma süresini, göz bebeği çapı gibi bilgileri içerir. Bu çalışmada, katılımcının deney düzeneğine baktığı andaki sabitleme verileri kullanılmıştır. Böylece verideki gürültünün azaltılması ve analizlerden daha doğru sonuçlar alınması hedeflenmiştir. Her resimde her katılımcının kaç tane sabitleme verisi ürettiği tespit edilebilmektedir. Her katılımcı farklı resimler üzerinde farklı miktarda sabitleme verisi üretebilmektedir. Sabitleme verisi, her bir resim için tek bir değerden oluşan cevap ve cevaplama süresi öznitelikleri ile birleştirilmek istendiğinde, bir kişinin her resim üzerinde oluşturduğu sabitleme verisinin ortalaması alınmıştır. Oluşan bu yeni veriye ET_log denmiştir.

Bilindiği gibi Rastgele Orman(İng. Random Forest) algoritması en iyi ve en gerekli öznitelikleri kullanarak sınıflandırma yapar. Bu nedenle çalışmanın bu kısmında öznitelik seçme yöntemleri kullanmak, sınıflandırma performansını anlamlı bir şekilde etkilememiştir. Ancak veride gürültü temizleme yöntemi olan Tomek link atma metodunu kullanmak ET_log verisi ile alınan sınıflandırma sonuçlarını iyileştirmiştir. Bunun yanında ham sabitleme verisi ile alınan sonuçlar üzerinde bir etkisi olmamıştır.

Bu nedenle işlemi bir adım öteye götürerek grupların sınıflandırılmasında derin öğrenme tekniklerinden yararlanılmıştır.

Derin öğrenme teknikleri, çoklu soyutlama seviyesine sahip verilerin temsillerini öğrenebilmek için, çoklu işlem katmanlarına sahip modellerin üretilmesini sağlar. Biz bu çalışmada derin öğrenme tekniklerinden Evrişimsel Sinir Ağları (İng. CNN - Convolution Neural Network) ve Uzun Kısa Vadeli Hafıza Ağları (İng. LSTM - Long Short Term Memory) algoritmalarına yoğunlaştık. Kurulan yapay sinir ağı modelinde, CNN katmanı, verideki gizli öznitelikleri diğer bir ifadeyle verinin soyut temsilini ortaya çıkarmak için kullanıldı. LSTM yöntemi ise veriyi sınıflandırmak için kullanıldı. Ürettilen modellerde performans karşılaştırması yapabilmek için, hem sadece LSTM katmanının kullanıldığı sinir ağları inşa edildi hem de CNN ve LSTM katmanlarının art arda olduğu sinir ağları oluşturuldu.

Tez çalışmasının bu kısmında sabitleme göz hareketi verisine ek olarak sıçrama (İng. saccade) göz hareketlerinden de faydalanılmıştır. Böylece, veri kümesindeki öznitelik miktarı artırılmıştır. Üç farklı model oluşturularak katılımcı gruplarının sınıflandırma işlemi yapılmıştır. İlk modelde sadece sabitleme verisi kullanarak sınıflandırma yapılırken, ikinci modelde sabitleme ve sıçrama öznitelikleri bir araya getirilip tek bir veri kümesine dönüştürülmüştür. Üçüncü ve son modelde ise paralel iki katman oluşturulmuş; ilk katmanda sabitleme verisi girdi olarak yapay sinir ağına verilip, paralel ikinci katmanda ise sıçrama verisi sisteme girdi olarak verilmiştir. Bu şekilde, sabitleme ve sıçrama göz hareketlerinin ayrı ayrı model performansına katkıları gözlemlenebilmiştir.

Önerilen yöntemlerin gürbüzlüğünü göstermek için orijinal veriye gürültü ekleyerek yeni bir veri kümesi oluşturulmuştur ve geliştirilen yöntemler bu veri kümesi ile de test edilmiştir. Elde edilen sonuçlara göre LSTM ve CNN yöntemleri ardışık katmanlar şeklinde birlikte kullanarak bir model oluşturmak başarılı sonuçlar alınmasını sağlamıştır.

Daha büyük popülasyon büyüklükleri ve alternatif girdiler hakkında daha fazla çalışmanın desteklenmesi ile, bizimki gibi yaklaşımlar erken tanıyı kolaylaştırmak ve böylece erken tedavi süreçlerini mümkün kılmak için kullanılabilir. Amaç, OSB ve DEHB üzerinde çalışan psikiyatrist ve terapistlere, bilgisayar destekli bir teknik kullanarak söz konusu katılımcı gruplarının tanılama sistemlerine yardımcı olmaktır.

1. INTRODUCTION

Disorders such as Learning Disabilities, Autism Spectrum Disorder (ASD), and Attention Deficit Hyperactivity Disorder (ADHD) have negative impact on human beings all over the world. People with these disorders have difficulty in academic performance and in daily life. ADHD is a common neurodevelopmental psychiatric disorder among school children [2, 3]. The core symptoms of the ADHD are inattention, hyperactivity and impulsivity. According to American Psychiatric Association [2] ASD is characterized by social impairments in interpersonal communication. Ability to recognize emotions has been considered one of the prevalent difficulties of people with ASD. Related to their deficiency of establishing eye contact, they have difficulties emotion recognition and therefore communication.

The study of emotional states is helpful for interpreting human actions. To interpret the action of the person that we talk with or the consequences of our actions we can make use of his/her emotional status. For example after we tell a joke if the person is smiling then we can interpret that our joke is funny for him. If the person is frowning then we can interpret that we have offended the person. During our lifetime we keep learning how to display and how to interpret other peoples emotions throughout our conversations with them.

We can recognize emotions by looking at the face of the person. Facial statement, mouth shape and eyebrow shape determine critical visual information about the emotional status of a person. The voice tone also reflects emotion status [4]. So we can use voice as an input while recognizing emotions. Besides sight and audio features physiological measurements give some information about emotions. For example if the person gets excited his heart rate and body temperature increase [5]. There have been studies that use gait and posture for emotion recognition [6]. Responses of a person while he is interacting with a computer program can also be used for emotion recognition. For example features obtained from the persons interaction with the program (application log data) can be used to determine emotion of the person. The interactions could be used to directly measure emotion (What is your emotion?) or understanding of emotion (What is the person in the picture is feeling?) or they could be used to indirectly determine the emotion of the person (for example repeated key strokes could indicate an angry player in a game.). Whether a person understands or fails to understand certain emotions can also be an indicator of a certain condition such as ASD [2].

1.1 Significance of the Thesis

As far as we know, this is the first work in Turkey with Turkish people that deal with measurement of emotion recognition ability of ASD, ADHD and typical development (TD) groups by using eye tracker and application log data. This study is also unique in the world because of the use of machine learning methods on these data for the emotion recognition task.

It is very important to be able to present detailed information about emotion recognition abilities of the participants to the specialists. The information we aim to provide within the scope of this thesis includes not only the statistical analysis of the raw data, but also the details of the machine learning methods and their outcomes. Furthermore we will investigate whether participants have difficulties in recognizing emotions at different levels of strength.

This work can be used as a preliminary study for improving emotion recognition ability of individuals using methods such as active learning [7] which aims to reduce the number of trials required for teaching.

1.2 Purpose of Thesis

The purpose of the thesis is comparison of the emotion recognition process of different groups of people. For the use case considered, children with ASD or ADHD and typical developing children are compared according to their emotion recognition ability and process. Understanding how an individual can not recognize an emotion compared to the other individuals who can, allow for characterization of these differences. If significant differences between emotion recognition process of these groups are discovered, in the future these could be used as a distinguishing feature for diagnosis.

1.3 Background and Related Work

1.3.1 ASD, ADHD and emotion recognition

Emotion recognition is a process which requires voice, face expression as inputs; especially eyes are very informative for emotion recognition. As it is known, people with ASD do not have tendency to look at eyes while they communicate or interact with other people [8, 9]. In [9] study, investigation results of the looking behavior to the emotional faces showed that ASD group has lower fixation percentage on the face region (mouth, nose, eye) than the control group. However, any considerable significant difference between groups in terms of percentage of fixation on eye region is not reported. [8] analyzes the gaze patterns of the people with autism spectrum disorders on the emotional face images and they measure the response time and emotion recognition ability of them. As a result, they conclude, participants with ASD (mean age= 32.71) look frequently outside the eyes on the face. By contrast with ASD participants, typical development participants look eyes as it is determined in [10] study. It means a very critical input of the emotion recognition process is missing for people with ASD. Therefore emotion recognition is a challenge for them.

In research of [11], 86 children with ASD with an average age of 10.65 and 114 typically developing children with an average age of 10.32 are tested to match emotion expressions with correct emotion names. The correctness of their answers and the response time are analyzed and as a result of the research, the children with ASD have remarkably lower rates of correct answers according to the typically developing children.

Age of the participant is critical on the emotion recognition of the children with ASD. According to [12], older ASD participants have higher performance than younger ASD subjects on recognition of the emotions.

Evaluating the accuracy of emotion recognition and latency to recognize emotions in children with ADHD is one of the aims of the [13]. Also in this work, the relationship between emotion recognition and social behaviors has been investigated. The stimuli of the work consist of videos which show five images of one emotion according to emotion levels; the first image is a neutral image and while moving toward the end of the video emotion level increases. The participants give just one answer to all levels of an image. Unlike this, we have asked the participants, different level of images as separate questions. Thus we obtain more detailed results about the emotion recognition ability of the participants.

1.3.2 Applications for people with ASD and ADHD

FEFA is educational software that is able to recognize the emotions of the users and train them. The system was developed by [14] and a final release can be obtained from the following reference: Center of Neurodevelopmental Disorders at Karolinska (2015). FEFA recognizes the weak side of users concerning emotions and serves them a suitable educational program. Inside FEFA's test and educational modules, two types are shown: pictures of whole face and pictures of the only eye region. According to the selected part, pictures covering the six primary emotions (happiness, sadness, fear, confusion, disgust and neutral) are shown to the users and asked them if they verify these emotions. At the end of the test section, the user's response latency and correct answer ratio are calculated. Inside FEFA application only users' answers and response latency are accumulated, however, in our research besides users' answers and response latency, eye tracker data is used.

In a mobile application called dmTEA, the behavior of children with ASD is measured using various games and the results are delivered to related teachers [15]. As a source of data, only the reaction of the children to mobile games is collected. However in our research sensors data are also used as a source of data.

1.3.3 Machine learning methods for diagnosis of people with ASD and ADHD

There are several studies for autism and ADHD classification, also for analyzing their measurement results by using machine learning algorithms. [16] implements SVM classification algorithm to classify adolescents with ADHD and controls, the classification is done according to functional and structural brain patterns while participants are working on a Flanker/NoGo task. Instead of using eye tracker, fMRI data is obtained from the participants.

[17] presents an affective state prediction of college students (aged between 20-30) by using physiological measures such as SCR (Skin Conductance Response), EMG (Electromyography), respiration, EEG (Electroencephalography). In order to predict the affective state, they employ decision rules, k-nearest neighbour and decomposition tree data mining models on the feature extracted physiological measurement data.

There is another work done by [18] which use machine learning methods to predict the emotional condition of the six individual with autism spectrum disorders. They measure the electrodermal activity (EDA), electrocardiogram (ECG), facial EMG, temperature, heart sound, photoplethysmogram, bioimpedance and use these sensor data in support vector machine (SVM) model.

[19] proposes a machine learning method to classify children with autism spectrum disorder and typically developing children. They have investigated the upper-limb movement as a classification factor for distinguishing children with autism spectrum disorder aged 2-4 and normals. Their classification accuracy is 96.7%. According to these indications they have considered motor signatures may be an identical feature for the ASD patients.

Although recent autism behavioral diagnosis methods are robust and valid, these methods are very time consuming and restrictive [20]. The Autism Diagnostic Observation Schedule-Generic (ADOS) [21], which is very popular and reliable instrument to diagnose ASD, is divided into four modules and individuals take place in a module according to their language and developmental abilities. It takes 30-60 minutes to carry out a module in ADOS and there are 29 items in Module 1 of ADOS. In this work, their goal is to minimize the number of examined items of Module 1 to accelerate the diagnosis of the disease. Therefore they have selected just eight items of Module 1 in their experiments and obtained approximately %100 accuracy level by using machine learning algorithms. As a data source, they practice on Autism Genetic Resource Exchange (AGRE) and Boston Autism Consortium (AC). 612 individuals with autism and 15 non-spectrum individuals are classified by using machine learning algorithm according to their Module 1 of ADOS scores. 16 classification algorithms have run and the best accuracy performance has received from alternating decision tree (ADTree) algorithm.

1.3.4 Eye tracking

Experimental psychology and clinical neuroscience researches have showed that there is an important connection between eye gaze and mental disorders [22–24]. In [24] eyes are able to give information on mental disorders; therefore they conclude that eye tracking is very critical for health monitoring. Regarding mental health tracing smooth pursuit movements and saccadic features are extremely valuable.

[25] reviews the literature for whether eye movement should be used in the assessment of the Disorders of Consciousness (DOC) which has types such as Minimally Conscious State (MCS), Vegetative State (VS), and coma. Their aim is testing the performance of the computerized eye tracking methods for diagnostic classification of patients who have these disorders. It is emphasized that there exist a close relationship between eye movements and organic dysfunction in the brain, hence eye tracking may have been used to assess the severe DOC. Furthermore, they have asserted, eye tracking studies will advance the accuracy of clinical appraisal in disorders of consciousness DOC.

The objective of [26] is to discover the gaze behavior of high-functioning children with ASD. To achieve this goal the participants are shown the famous male, female faces and animal faces. They analyze whether the participants looked at eyes or other parts of the face with considering fixation count and fixation duration of the participants.

The contribution of [27] presents a comparison between the identification of emotions from natural human faces and synthetic facial expressions in terms of images being dynamic or static. They have reported that there are no differences between the identification of dynamic and static natural human faces. However, they have released that dynamic synthetic facial expressions were recognized more correctly than static facial expressions.

1.4 Main Contribution of the Thesis

Main contribution of this thesis are as follows:

- We show that using eye movement data 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 propose a deep learning framework for the task of the participant classification which can learn the emotion recognition behavior of the participants.
- We employ two deep learning techniques, CNN and LSTM. CNN is a powerful method to extract hidden and relevant features of the data. Besides this, we utilize the long short-term memory characterization of LSTM algorithm.

1.5 Organization of the Thesis

The remainder of the thesis is organized as follows:

Chapter 2 presents the demographic information of the participants and experimental design. Here, we introduce age, gender and diagnostic groups of the participants. We propose the eye tracking device as an instrument for our experiments to be able to capture the eye gaze of the participants. The way that we collect the data and the alignment method between the eye tracker video and the fixation points are described in this section. Finally, the results of some initial data analysis and ANOVA statistical analysis are demonstrated.

Chapter 3 constructs the workflow for classification of the application log data. In order to prevent the class imbalance problem, we prefer to use random resampling technique. We perform ReliefF feature selection algorithm to extract relevant features of the application log data. As the validation technique, we employ leave-one record-out. Finally, we present classification results of SVM, decision tree, random forest, AdaBoost, KNN algorithms.

Results of the literature review on ASD and ADHD classification by using feature-based machine learning algorithms and emotion recognition performances of these participant groups are presented in Chapter 4 and 5. Chapter 4 demonstrates multimodal classification results on noisy eye tracker and application log data of children with autism and ADHD. We make use of fixation features as the eye tracker data; response correct (RC) and response time (RT) features as the application log data. The RC and RT features have lower feature importance value when compared to the fixation data features.

Chapter 5 introduces LSTM and CNN based analysis results in eye movement data for diagnosis of participants with ASD and ADHD. In addition to fixation eye movement data we use saccadic eye movements in this chapter. To the extent of our knowledge, this is the first study that use eye movement data with LSTM and CNN algorithms to classify ASD and ADHD patients.

Chapter 6 concludes this thesis and presents directions for future work.

2. DATA ACQUISITION

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

2.1 Participants

The experiments were conducted in Marmara University Medical Faculty Hospital Child and Adolescent Psychiatry Outpatient Department. 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 2.1, 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 [28] of Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL) was utilized [29] 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 2.1 : 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	6	11	1.89
CONTROL	4	9.50	1.12	9	9.25	1.56

Figure 2.1 : Emotion levels for angry emotion.

2.2 Experimental Design

The stimuli used in the experiments are presented to the participants by a web application called TrackEmo. It includes 40 emotive human face images from Cohn-Kanade database [1]. Cohn-Kanade database consists of video files which are formed of emotion expressive face images. In these images, emotion level starts from neutral and gradually increase the intensity of the emotion and finally reaches to a peak expression. For our work, the videos are converted into image sequences. Four different emotional levels of emotive expressive images are chosen from the image sequences. A sample of emotion levels of an angry emotion expressive video is shown in Figure 2.1; the first level represents neutral form of the emotion and emotion level increases on the other images gradually.

2.2.1 Phases of the experiments

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 is prepared in order to familiarize the participants with the main experiment. The images of the human faces are shown to the participants. The Cohn-Kanade database includes images of the face, ear, mouth, and nose regions, which are clear and contains some emotions. In this phase, the participants are asked the emotions of the five images.

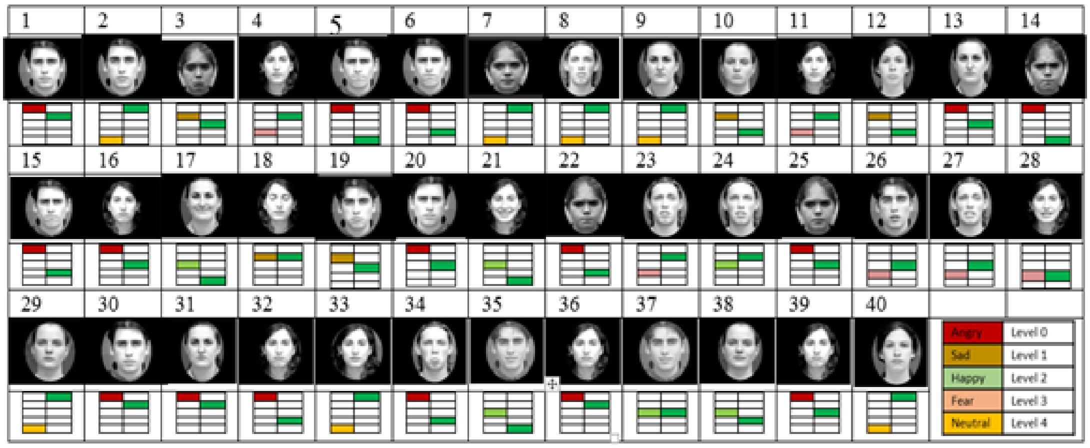


Figure 2.2 : The images from Cohn-Kanade [1] database that are used in emotion recognition experiment. In the table first column represents the emotion and the second column is emotion intensity level.

Emotion Recognition Phase: This is the main part of the experiment. It is similar to the warm-up phase, but in this phase, 40 images are shown to the participants. Response and response time of the participants are measured. The images shown at each step are chosen randomly. However, the orders of the images are 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 are used more often [30].

2.3 Procedure of the Experiment

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

1. The doctors and psychiatrists do the participants' checks and tests.

2. The eligible participants and their families are informed about the experiments by the doctors, and if a participant accept to participate in the experiments as a volunteer, his/her family signs the consent form.

3. The participant and family are put in a room where the experiment set-up have been prepared.

4. The experimental set-up and the experiment routine are explained to the participant.

5. Participant's age, gender, hunger, and tiredness state are recorded.

6. The participant sits in the prepared armchair.

7. Participant's ID is entered into the system.

8. Eye tracker is put on to the participant and calibration is done.

9. The experiment starts with the emotion recognition warm-up phase, where the participants are asked to recognize the emotions on five images. Their response and response latency are recorded.

10. The next step is the emotion recognition phase, and it is the same as the former step, the only difference is 40 images are shown to the participant.

When the experiment is over, the eye tracking glasses are taken off the participant, and he/she is thanked for their participation.

2.4 Dataset Acquisition

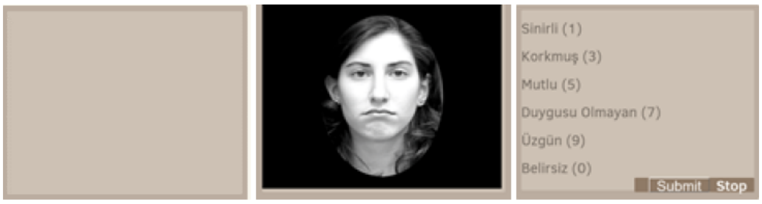
During the experiments, application log data and eye tracker data are gathered.

2.4.1 Application log data

Application log (APL) data consists of response and response time (RT) of the participants. During the experimental phase, while participants are shown emotive face images, they are asked "What is the emotion of this woman or man?" Their answers are called as RC (response correct) data, and the duration of the response is saved as RT. [31] employs RC and RT feature types and the aim of [31] study to estimate RC value of the participants with ASD. However, we try to diagnose the participants by using RC and other features.

2.4.2 Data alignment for eye tracker video and fixation points

The participants are shown three different scenes in the experiment phase of the TrackEmo program. As seen in Figure 2.3 a) the first one is the empty scene, that represents the transition between choices scene and the next image scene. The second one is a human face scene (Figure 2.3 b) in which participants see an expressive human



a) b) c)

Figure 2.3 : Parts of the TrackEmo user interface: a) empty scene b) human face with an emotional expression c) emotion choices.

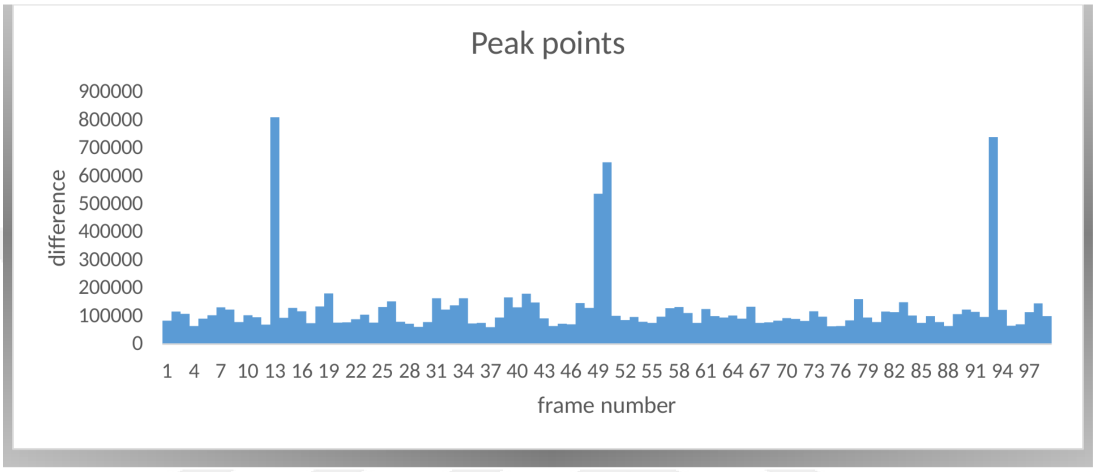


Figure 2.4 : Histogram of the RGB values of the consecutive frames and peak points.

face with an emotion (one of angry, fear, happy, neutral and sad) and try to capture the emotion. Once the participant confirms, the experimenter shows the emotion choices which is the third scene (Figure 2.3 c) and he/she selects one of the shown emotions.

The most important point of this work is to detect the fixation event when participants see the emotional face images and obtain the fixation features such as pupil diameter, dispersion, pupil coordinate. Therefore we need to figure out in which frame emotional face scene or choices scene appears. In order to do that, the video is divided into frames and RGB values of each frame is calculated. Then the absolute difference of the RGB values of two consecutive frames are computed. Thereby if sequential two frames have very different RGB values, as it is shown as peak points in the Figure 2.4, we infer new emotional face image appears on the screen. In Figure 2.4, the 13th frame has a peak, because in this frame an emotional image appears after an empty scene. Also, the 51st frame has a peak since there is a transition from an image scene to a choice scene here. We manually identify the first emotional face scene and then we assume that the next peak point is the choice scene and the next peak point is again image scene and so on.

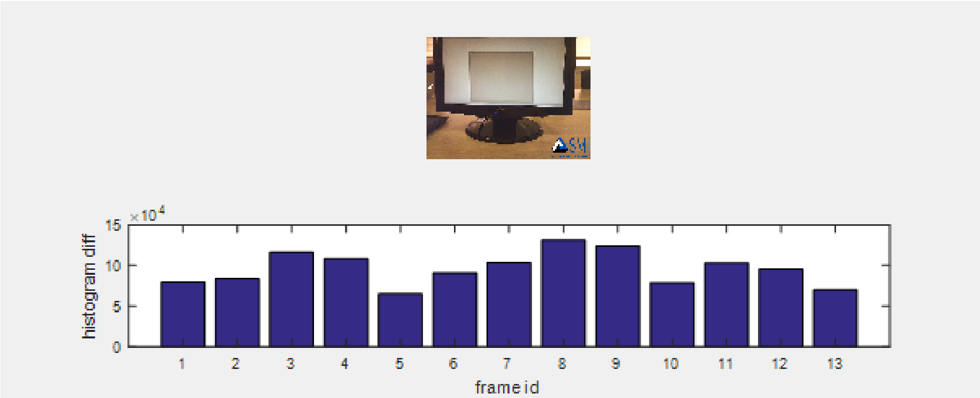


Figure 2.5 : Histogram of the first thirteen frames.

Figure 2.6 : 14th frame has a peak, because emotional image appears on the screen.

Because of choices scene and empty scene have very similar RGB values, we do not take into account the transition between them.

As it is shown in the Figure 2.5 histogram difference values are between $5x10^4$ and $15x10^4$ along empty scene. When the image is seen on the screen (see and Figure 2.6), histogram difference value peaks and reach the $10x10^5$, it means 10 times of the previous value.

According to Figure 2.6 and Figure 2.7, it is proven that the participant looked at the emotional images between 13th and 51st frames, after replying the question "What is the emotion of this person?" choice screen appears. As a result, exact representation time of the frame is determined and the part of the fixation information that is in this time interval is utilized. Thus, pupil diameter, pupil size, dispersion information is achieved during emotional image is on the screen. The main purpose is to ascertain the physiological changes that occurs on the participants body.

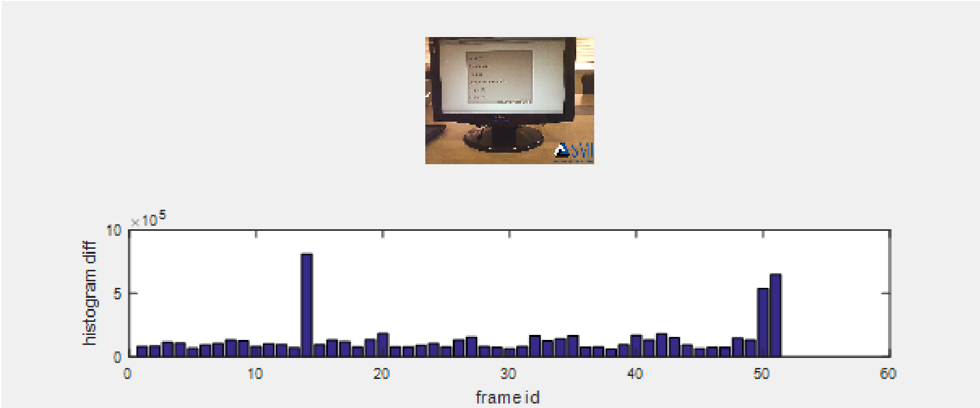


Figure 2.7 : Since choice scene is shown, 51st frame has a peak point.

2.4.3 Eye tracker data

According to [32,33] 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. The dispersion is calculated as follows:

$$D = [max(x) - min(x)] + [max(y) - min(y)] \tag{2.1}$$

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 produce 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 features are employed as fixation features in the current work. The fixation points, which occur when a participant is looking out of the screen, are excluded from the analysis. In this work, each participant is shown 40 images while they are wearing the eye tracking glasses. The eye tracker records the fixation data while the participant is looking at the images. For each image, each participant creates more than one fixation data point. Since each person has different eye gaze behavior, the number of fixation points that occurred while they are looking at the images is 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 refer to the fixation data as raw fixation (RF) data in the experiments from now on.

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

$$Z_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{2.2}$$

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 are calculated; position X and position Y of the fixation points, average pupil size X and Y coordinates, and average pupil diameter.

Since the participants are demonstrated 40 emotive images, average and standard deviation value of the fixation data for each image are used as features for the participants. For example, if a participant has 15 fixation points for the second image, the average and the standard deviation of these fixation points are utilized as features for him/her. We call this data type as updated fixation data. In order to increase the feature size, we merge updated fixation data and APL data and we call it Eye tracker log (ET_log). The features of the ET_log are presented in Table 4.1.

2.5 Data Analysis

2.5.1 Fleiss' Kappa

During the experiments, the participants are shown human face images with emotional expression and are asked the emotion of the person shown. For each participant

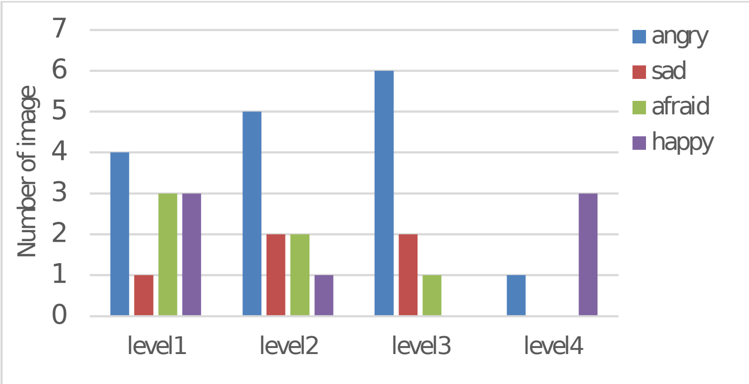


Figure 2.8 : Number of images per emotion and emotion level during the experiment phase.

group, we try to understand whether their answer to the question agree with that of the other participants who are in the same group by chance or not. Therefore, we employ Fleiss' kappa [34] to evaluate the degree of agreement between two or more participants' answers. Fleiss's kappa is a generalized version of the kappa statistical measurement. Kappa and weighted kappa measures are 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 prefer the interpretation methodology of [35]. We measure 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 is 0.41, which can be interpreted as a moderate agreement. For the ADHD group, the Fleiss' kappa value is 0.35, which shows fair agreement; for the ASD group, it is 0.23 that points fair agreement too. As distinct from the ADHD and ASD groups, there exists a moderate agreement among participants in the control group. Although participants with ADHD have higher kappa values than ASD, both of them have a fair agreement degree. We generate a random dataset that includes random responses to the images and we observe the Fleiss' kappa value for that random group. The Fleiss' kappa for the random group is 0.0031, which is significantly lower than the Fleiss' kappa for all the classes.

2.5.2 Emotion level based response correct and response time distribution of the participant groups

The number of images per emotion and emotion level during the experiment phase is presented in Figure 2.8. Also, for each emotion and emotion levels, response correct ratio of the participants is shown in Figure 2.9. For anger, five level 3 images are shown to all groups. Totally 60 angry images with level 3 are shown per human groups and the participants with ADHD respond correctly 34 of them. The response correct ratio of the participants with ADHD for angry level 3 emotion images is calculated as 34*100/60. However, for all emotion levels, the participants with ADHD and the control group show similar behavior on angry emotion recognition, both groups answer correctly to 73 out of 192 angry emotional images.

The images shown to the participants in different emotion and emotion levels are represented in Figure 2.2. As seen, the emotions in the images in level 1 are rather uncertain. Therefore, the ratio of the correctly answered questions is very low in level 1. On the other hand, as the level of emotion increased, the rates of recognition of emotions increased for all three groups. The response correct ratio for angry images is the highest for the control group, except for the level 1 images. ADHD and ASD groups follow the control group, respectively. The correct recognition ratio of the happy emotion is generally high for all groups. In support of this finding, error rate of the happy emotion is the least compared to the other feelings in [36]. For fear recognition, there is a statistically significant difference (p=0.028) between the groups, correct response ratio of the control group is much more than the other groups. The statistical differences is measured by one-way ANOVA test which is used to compare the means of two or more groups. We can not find any statistical difference between groups by regarding the response correct feature for the other emotions.

According to the average results in Table 2.2 and Table 2.3 the participants with ASD correctly recognize fewer number of questions than the control group for all emotions. Also, answer time of the ASD group is longer than the control group. Participants with ADHD perceive the happy emotion a little more accurately and sooner than the control group. Besides this, the response time of the ADHD group is shorter than the control group, with similar accuracy in angry emotions. The ASD group performs better than the ADHD group on emotions of sadness and fear, however they are always worse than ADHD group for response time.

Figure 2.10 illustrates the emotion recognition confusion matrix of the participants. The rows show actual emotions and the columns demonstrate the percentage of the responses. The last column displays the percentage for the unknown responses. As

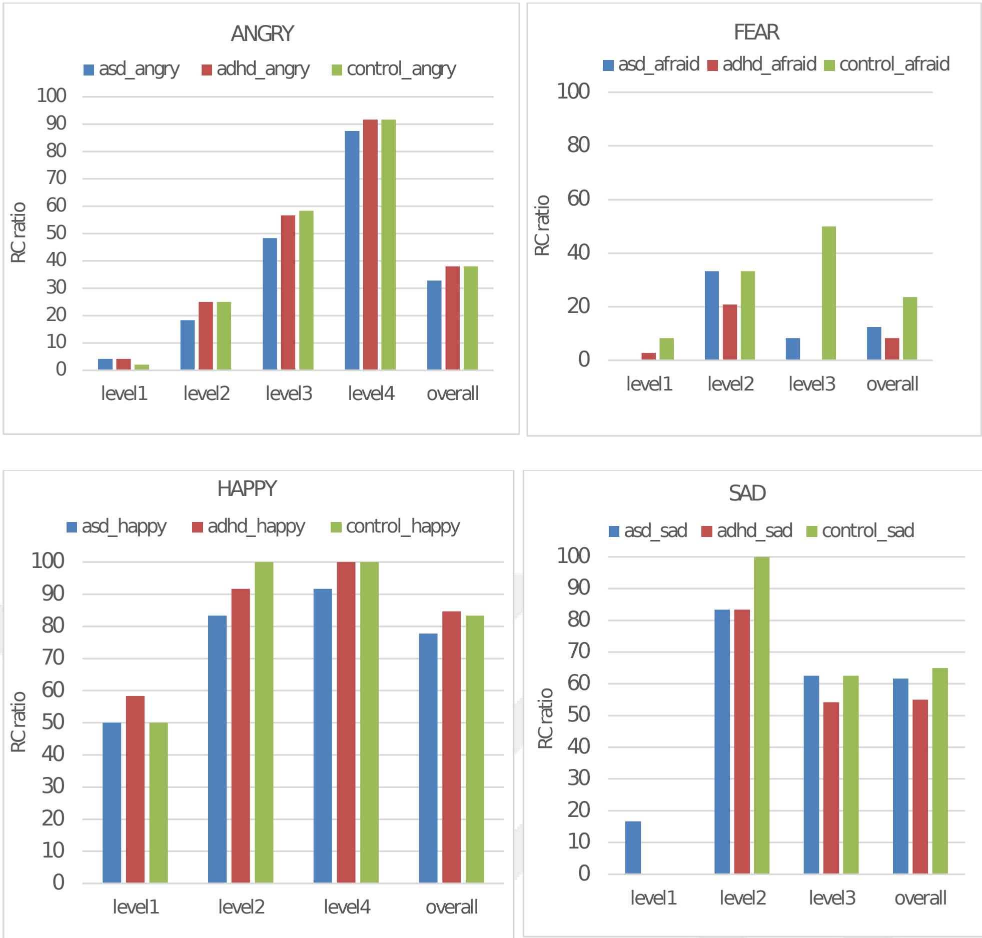


Figure 2.9 : Response correct distribution of the participants for emotions and emotion levels.

Table 2.2 : Mean of the number of correctly recognized emotions (M), Standard Deviation of the number of correctly recognized emotions (SD) and F-test value of the Response Correct (RC) of the participant groups.

	Response Correct					
	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:*For Post-hoc tests Bonferroni was used, $p<0.05$.

Table 2.3 : Mean of the number of correctly recognized emotions (M), Standard Deviation of the number of correctly recognized emotions (SD) and F-test value of the Response Time (RT) of the participant groups.

	Response Time					
	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:*For Post-hoc tests Bonferroni was used, p<0.05.

it is seen in Figure 2.10 the ADHD and the control groups have similar pattern of correct answers. However, the participants with ADHD respond unknown (22%) more often than the control group (16%) or the ASD group (13%). The most confused emotion is fear among the others for all participant groups. Particularly, the ASD group answers the fear emotional images as happy and sad mostly. The correctly recognition ratio of the sad emotion is higher for ASD than ADHD. On the other hand, the most correctly recognized and the least confused emotion is happiness. The images that present neutral emotion are another difficulty for the participants. More frequently the neutral images are responded as sad by participants with ASD. Participants with ADHD and the control group confuse neutral images with sad, but mostly they say they can not understand the emotion.

The response time of the participants from the three groups is shown in Figure 2.11. For some images, participants in different groups present similar response time behavior. For example, all participant groups spend a lot of time on image id 18 which contains a sad level 1 emotion. In general, the participants with ASD spend more time on the images. One-way ANOVA test indicates that, there are statistically significant response time differences between the groups, $F(2,1437)=29.831$, $p<0.0001$. As illustrated in Figure 2.11, response time of the participants with ASD is longer than the ADHD, $F(1,958)=37.932$, $p<0.0001$. Also, the response time behavior of the control

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 2.10 : Emotion recognition confusion matrix for ASD, ADHD and the control groups.

group is significantly different from the ASD group F(1,958)=31.562, p<0.0001. On the other hand, all participants generally spend less time towards the end of the experiment, which can be interpreted as exhaustion at the end of the experiment.

In Figure 2.12, we group (cluster) images according to the average response time. If participants' response time are similar for two images, they will be placed in close clusters and hence will be closer in the dendrogram. Figure 2.12 shows that images can be divided into four groups. The red coloured part of the dendrogram consists of sad images with ids 12, 19, 18 and a neutral image with id 8. The images with id 1 and 28 are grouped in the same cluster, these images show negative emotions, anger and fear, respectively.

2.5.3 Emotion based average pupil diameter (mm) distribution of the participant groups

We measure the pupil size of the participants by the eye tracker. The statistical differences between pupil diameter have been measured with one-way ANOVA test. In Table 2.4, F-test and p value of the groups are presented. The alpha value is selected

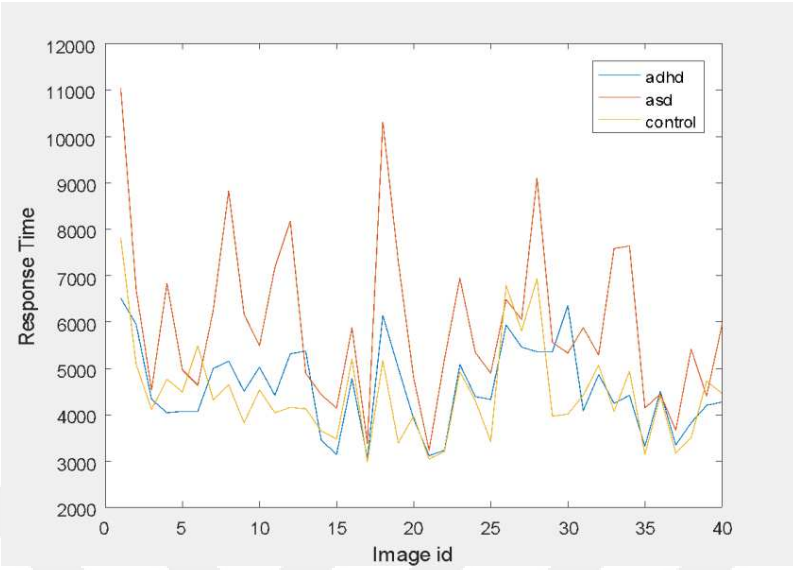


Figure 2.11 : Average response time (ms) distribution for each shown image of the participant groups.

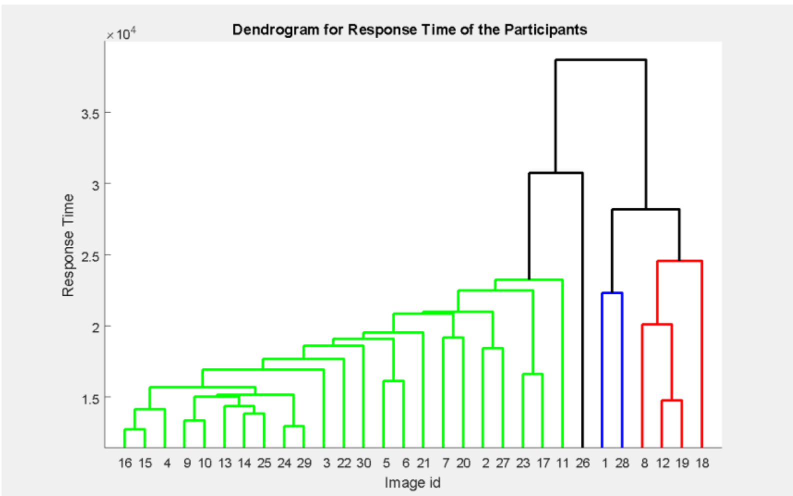


Figure 2.12 : Dendrogram representation of the participants' average response time (ms) (see Figure 2.2 for images corresponding to each image id).

Table 2.4 : Statistical test results for pupil diameter.

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

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

as 0.05, so the results are significant at the 5% significance level. For sad images, p=0.076>0.05, therefore the pupil diameter difference between participant groups on sad images is not significant.

3. APPLICATION LOG DATA BASED CLASSIFICATION

There are several studies for autism and ADHD classification problem by using machine learning algorithms. Different type of datasets are utilized as datasets during the classification process such as physical signals, answers to the questions and so on. SVM classification algorithm is implemented in [16] to classify adolescents with ADHD and the control group, classification is done according to functional and structural brain patterns while the participants get involved in a Flanker/NoGo task. Another study [19] conducts an experimental setup in which the upper-limb movement of participants is investigated as a classification factor for distinguishing children with autism spectrum disorder aged 2-4 and normals.

The workflow that we follow for the analysis of the first part of the collected data is shown in Figure 3.1. The first step is gathering the application log data from the participants. Application log data is formed of responses and response latencies to images displayed to the participants.

The main idea of this work is to investigate whether diagnosis of a participant can be differentiated from the others or not based on the application data. The diagnosis of each participant is used as an instance label. By using machine learning algorithms, we try to predict the diagnosis of each participant.

Due to some technical problems we have imbalance number of samples for each participant group. For solving this problem random sampling algorithm [37] is used as resampling method. The ADHD group has the maximum number of instances, therefore we decide to increase the instances number of other groups to 30.

3.1 Resampling

As we mentioned before, there are three types of participant group in our dataset; these are the ADHD group, the ASD group and the control group. Unfortunately, because of some technical problems, we could not collect equal number of data from

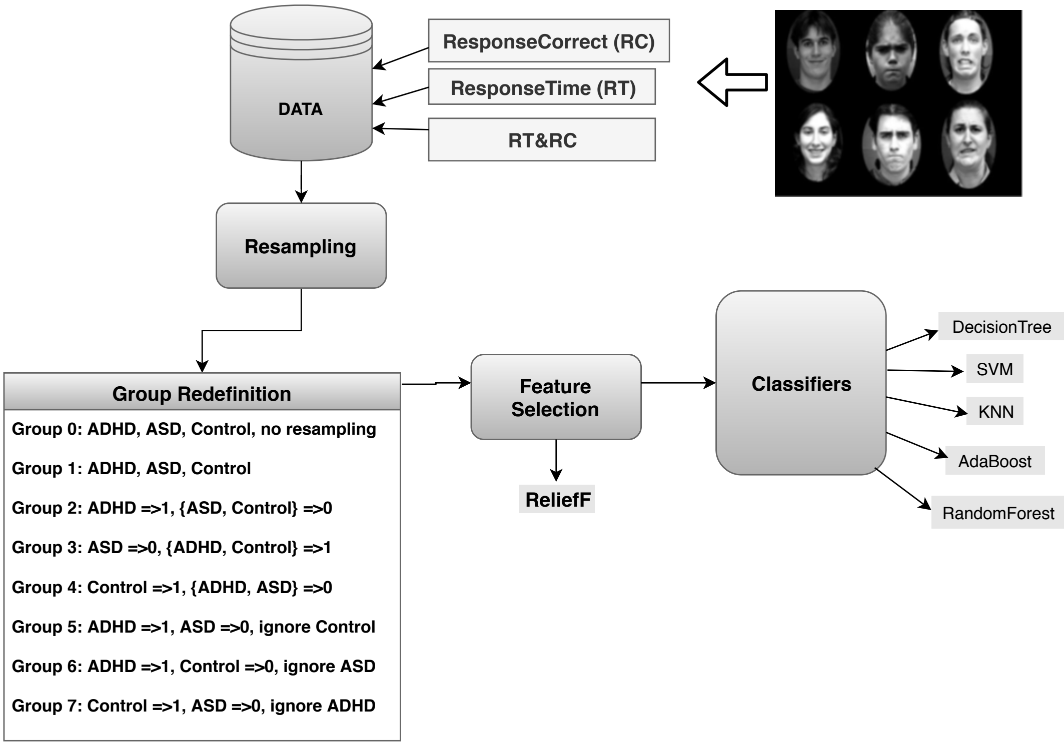


Figure 3.1 : Workflow of the system.

each participant group. Therefore, the dataset has an unbalanced distribution. Due to the fact that each group has different number of participants, the need to equalize the participants' number for each groups by resampling process emerged. Hence, in order to resample the records, we employ random sampling algorithm [37]. The ADHD group has the maximum number of instances, so the number of other groups' instances is also increased to 30. With the help of this method, the instances of the ASD group are selected randomly and added to the same group. The same process is applied to the control group. Finally, all groups have the same number of participants.

3.2 Mutual Information

Mutual information of two random variables measures the correlation between these two random variables. Mutual information and entropy notion have a relationship. In information theory, entropy identifies the quantity of information [38]. In other words entropy measures the uncertainty of a random variable [39]. By using Equation 3.1 entropy of a random variable is measured.

$$H(X) = -\sum_i p(X_i)\log(p(X_i)) \tag{3.1}$$

Mutual information of two random variables is formulated as:

$$I(X;Y) = \sum_{y\varepsilon Y} \sum_{x\varepsilon X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{3.2}$$

In Equation 3.2, $p(x)$ and $p(y)$ are probability density functions and $p(x,y)$ is joint probability distribution function of X and Y random variables.

Conditional entropy, $H(X|Y)$, which measures the uncertainty about X random variable when Y random variable is known, and mutual information can be formulated as follows:

$$H(X) = H(X|Y) + I(X;Y) \tag{3.3}$$

$$I(X;Y) = H(X) - H(X|Y) \tag{3.4}$$

$H(X)$ is a measure of uncertainty information of X. If $H(X)$ is subtracted from $H(X|Y)$, the remaining results will be mutual information between X and Y random variables [40]. Mutual information is a symmetric measure, mutual information between X and Y is the same with the mutual information between Y and X, the order of the random variable is not important (see Equation 3.8).

$$I(X;Y) = I(Y;X) \tag{3.5}$$

$$I(X;Y) = H(X) - H(X|Y) \tag{3.6}$$

$$I(Y;X) = H(Y) - H(Y|X) \tag{3.7}$$

$$H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{3.8}$$

3.3 Symmetric Uncertainty (SU)

Symmetric uncertainty is a correlation measure and it can be used to find out correlation between two features. Symmetrical uncertainty between feature and class value is measured [41]. Symmetric uncertainty is calculated by this formula:

$$SU(X,Y) = \frac{2 * I(X;Y)}{H(X) + H(Y)} \tag{3.9}$$

Table 3.1 : SU values of the different redefined groups on RC and RT data.

Redefined Groups	RC	RT
all groups	0.200	0.254
ADHD_others	0.247	0.310
ASD_others	0.146	0.064
Control_others	0.262	0.315
ADHD_ASD	**0.336**	**0.449**
ADHD_Control	0.143	0.072
ASD_Control	0.142	0.077

In this formula $I(X;Y)$ is mutual information and $H(X)$ and $H(Y)$ are entropy of random variables X and Y respectively [42].

In symmetric uncertainty features' values are normalized within range 0-1 [43]. If $SU(X,Y)$ value is 0, it means there is no correlation between random variables X and Y. Conversely, if $SU(X,Y)$ value is 1, knowledge of one of two random variables gives complete information about each other.

In this work, we compute symmetric uncertainty between each feature vector (image) and diagnosis label of the participants. Our purpose is to figure out which features are more related to the diagnosis of participants. The class redefinition process is applied for symmetric uncertainty. On the redefined class data groups, we execute SU and get the average SU values for all features as shown in Table 3.1. This table shows, symmetric uncertainty values are the highest for the ADHD and ASD groups.

3.4 Classification Problem Redefinition

After resampling the data, group labels of the participants are redefined. Since symmetric uncertainty values between response correct, diagnosis and response time, diagnostics are very low, group redefinition has been applied to application log data. In group redefinition method, number of the diagnosis type is decreased to two classes (except group1) with the help of these combinations:

- Group0: ADHD, ASD, Control, no resampling
- Group1: ADHD, ASD, Control
- Group2: ADHD => 1, {ASD, Control} => 0

- Group3: ASD => 0, {ADHD, Control} => 1
- Group4: Control=> 1, {ADHD, ASD} => 0
- Group5: ADHD => 1, ASD => 0, ignore Control
- Group6: ADHD => 1, Control => 0, ignore ASD
- Group7: Control => 1, ASD => 0, ignore ADHD

Class0 is not redefined, it has already existed. All the other seven classes are generated for redefinition process. For Class1 definition, all groups are used, just as the same as the previous experiments. For Group2 definition, ASD and the control groups' records are combined and labeled as "0" class and is compared to ADHD group records. In Group3, ADHD and the control groups instances are merged and took label "0" and then they are compared to ASD group that has label "1". ADHD and ASD groups are labeled as "0" and the control group is labeled as "1" in Group4 description. The highest symmetric uncertainty values are taken in case of only ADHD and ASD data records are used for Group5 definition. Group6 definition includes ADHD and the control groups. The control and ASD groups are in Group7. Referring to the generated classes, it has been observed that ASD group and ADHD group can be separated from each other easier than the others have.

3.5 Feature Selection

Feature selection process chooses a small subset of features that includes relevant and informative features. A relevant feature is highly correlated with the class label. By contrast, class and irrelevant features are not directly interrelated, but these features affect the learning process in a negative way. On the other hand, redundant features do not have any contribution to the learning process [44].

In this work, ReliefF [45] feature selection algorithm is used to select most informative features on the dataset. ReliefF algorithm is an extended version of the Relief feature selection algorithm, which was proposed by Kira and Rendel [46]. They employ Manhattan (L1) norm instead of Euclidean (L2) norm while calculating distance between instances. Moreover, they prefer absolute differences between instances than square of them. The weight of feature i is calculated as follows:

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \tag{3.10}$$

where x is a feature vector and $nearHit_i$ denotes the nearest feature vector which is in the same class with x feature vector. On the other hand, $nearMiss_i$ represents again the closest feature vector to x but it is in the different classes from x. According to Equation 3.10, if distance between x and $nearHit$ feature vector is high, weight of the feature i decreases. Beside this, if x is not close to $nearMiss$ feature vector, weight of the feature i increase too.

3.6 Experimental Results and Discussion

For this work, firstly we analyze the collected application log data regarding the number of participants that can be used for analysis. In addition to response correct (RC) and response latency (response time (RT)) data, we combine RC and RT and obtain RC & RT data. There are three types of groups in this work, but we could not collect an equal number of participants for each group. Therefore we need to resample the data for missing participants. Then, we apply the group redefinition because of the correlation between the response correct and the diagnosis and also the correlation between the response latency and the diagnosis are very low. Besides these, in order to improve the classification accuracy, ReliefF feature selection algorithm is used to choose the most informative and relative features.

3.6.1 Hypotheses

We state five hypotheses related to our study:

- Hypothesis1: We can differentiate participant groups using response correct data.
- Hypothesis2: We can differentiate participant groups using response time (latency) data
- Hypothesis3: Combining response correct and response time data will be useful to differentiate the participants groups.
- Hypothesis4: Redefining participant groups will increase the classification accuracy.

- Hypothesis5: Feature selection on the RC, RT and RC & RT data results in better classification.

3.6.2 Results

After the experimens finalized we analyze the collected application log data. Application log data is evaluated by both correlation analysis and machine learning methods. As correlation analysis, symmetric uncertainty method is employed. In addition to this, classification algorithms are used as machine learning techniques.

For each class, five classification algorithms are utilized; these were Decision Tree (DT) [47], Random Forest (RF) [48], Support Vector Machine (SVM) [49], k-nearest neighbour (K-NN) [50] and AdaBoost [51].

In Table 3.2, the best performance of the classification algorithms with and without feature selection process, for each group definition and on each data type is demonstrated. For example, the most accurate result is 80% for control_others group definition, on response correct data. It is achieved by selecting 15 features and implementing KNN algorithm. On the contrary, if feature selection process is not applied accuracy decreases to 68.89%. Another row shows the highest performance for asd_others group is 91.11% on the response time data by using Adaboost and ReliefF algorithms.

When we consider our hypotheses and machine learning results (see Table 3.2), the following conclusions are obtained (the accuracy results are leave-one-out accuracies, and hence they are test accuracy results):

- These participant groups can be differentiated by doing classification on response correct (RC) data:
  - Control and others by 80% accuracy
  - ASD and others by 75.56% accuracy
  - ADHD and ASD by 71.67% accuracy
- These participant groups can be differentiated by doing classification on response time (RT) data:
  - ASD and others by 91.11% accuracy

Table 3.2 : Accuracy values obtained using RC, RT and both features, with and without feature selection and different classifiers.

Group Definition	Data type	acc No_ReliefF	Best Classifier No_ReliefF	acc ReliefF	Best Classifier and setup for ReliefF
NoResampling	RC	49.18	Adaboost, RF	50.82	SVM, Relief_30
all groups	RC	43.33	DT	53.33	DT, Relief_30
adhd_others	RC	67.78	DT	67.78	Adaboost, Relief_15
asd_others	RC	77.78	Adaboost	75.56	Adaboost, Relief_30
control_others	RC	68.89	Adaboost	80.00	KNN, Relief_15
adhd_asd	RC	76.67	SVM	71.67	Adaboost, Relief_15
adhd_control	RC	60.00	Adaboost, RF	63.33	Adaboost, Relief_15
asd_control	RC	68.33	DT	63.33	DT, Relief_15
NoResampling	RT	72.13	DT	50.82	KNN, Relief_15
all groups	RT	72.22	DT	46.67	SVM, Relief_30
adhd_others	RT	62.22	DT,Adaboost	67.78	KNN, Relief_30
asd_others	RT	85.56	Adaboost	91.11	Adaboost, Relief_30
control_others	RT	84.44	DT	83.33	DT, Relief_30
adhd_asd	RT	90.00	Adaboost	90.00	Adaboost, Relief_30; DT,Relief_15
adhd_control	RT	73.33	DT	80.00	DT, Relief_15
asd_control	RT	65.00	KNN, Adaboost	78.33	RF,Relief_15
NoResampling	RC&RT	63.93	DT	50.82	SVM, KNN, Adaboost, Relief_30
all groups	RC&RT	61.11	DT	46.67	Adaboost, Relief_15
adhd_others	RC&RT	65.56	DT	68.89	Adaboost, Relief_15
asd_others	RC&RT	87.78	DT	71.11	KNN, Relief_30
control_others	RC&RT	82.22	DT	83.33	Adaboost, Relief_15
adhd_asd	RC&RT	90.00	Adaboost	81.67	DT, Relief_30
adhd_control	RC&RT	76.67	DT	75.00	Adaboost, Relief_30
asd_control	RC&RT	80.00	Adaboost	70.00	Adaboost, Relief_15

- ADHD and ASD by 90% accuracy
- Control and others by 84.4% accuracy
- ADHD and Control by 80% accuracy
- ASD and Control by 78.3% accuracy
- All groups with resampling by 72.22% accuracy
- All groups but no resampling by 72.13% accuracy

- These participant groups can be differentiated by doing classification on response correct and response time (RC & RT):
  - ADHD and ASD by 90% accuracy
  - ASD and others by 87.78% accuracy
  - Control and others by 83.3% accuracy
  - ASD and Control by 80% accuracy
  - ADHD and Control by 76.67% accuracy
  - Redefining participants groups increased the classification accuracy. Except RT data for all other data groups (RC and RC & RT) classification performance of Class0 and Class1 are the worst, because they are not redefined groups. It means, if we do not redefine the groups, accuracy of the classification will decrease. For this reason our hypothesis is confirmed.
  - Table 3.2 shows us ReliefF feature selection method on the RC, RT and RC & RT data to produce better classification in general.

3.7 Summary

As a result of this study, by means of data analysis on the collected log data, we achieve to separate participants with ADHD, ASD and control groups from each other. Selecting relevant and informative features increases the classification accuracy. Also classification performance benefited from group redefinition. The best performance is obtained for the asd others group on response time data. Namely we are able to distinguish participants with ASD from the other group which includes participants with ADHD and control. Also by 90% accuracy we differentiate participants with

ADHD from participants with ASD by using Adaboost algorithm with and without feature selection process on response time data. In addition to this, on RC & RT data ASD and ADHD participants can be separated with 80% accuracy.

4. MULTIMODAL CLASSIFICATION ON NOISY EYE TRACKER AND APPLICATION LOG DATA

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 the 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 use classification algorithms with the Tomek links noise removing method.

4.1 Background and Related Work

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 [12]. Emotion recognition processes carry clues about one's overall emotional well-being. Interpreting the facial expressions makes it easier to communicate with other people [52]. 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 neurodevelopmental disorders.

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 [53]. 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 [54, 55]. In our work, the participants are asked the emotion in the presented images. The purpose of the experiment is to understand how participants react to these images, and how their eye movements changed during the experiment.

The response of the participants, RT and the eye tracker fixation data are recorded and used for the analysis. We use the data obtained during the experiments to classify the participants with the machine learning methods.

The main question that we try to answer 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 [56]. 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.

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 [36]. 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 [57, 58]. In [30] fixation data and Autism Diagnostic Interview-Revised (ADI-R) score are 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 perform worse than the control group. Similarly, in our emotion recognition experiments, the control group give 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 [59, 60]. In the cognitive system or during emotional processes, pupil size gives critical information about mental workload and cognitive functioning [61]. As stated by [62], 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 [63]. In [28] only a pupil diameter feature is 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 [33]. In [64, 65] studies of the eye gaze data acquisition are performed and it is used for the measurement of the attention of the operators. Specifically, eye fixation and pupil diameter features are found important for detecting the attention level of the operators.

Our study uses the fixation, response and response time data describing the behavior of the participants while the participants are try to recognize the emotions of the people in the pictures shown to them. This data is used to identify the diagnosis of the patients. We do not aim to infer the emotions of the participants in this study. Similar approaches to ours are followed in [33, 66]. The fixation and saccade features are used to classify the patients with mild cognitive impairments in [33]. In [66] the eye tracking variables such as; fixation and saccade are 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 are used for classification by using eye movement data, no emotion extraction is done.

There are several studies for ASD and ADHD classification by using artificial intelligence and machine learning algorithms. The majority of these studies use fMRI (functional magnetic resonance imaging) data type. Although fMRI is a powerful data type, it is difficult to process. We achieve similar classification performances with the studies in which fMRI are used [67]. On the other hand, we outperform some fMRI works [**?**, 68] by using fixation and application log data (APL) data, which are easy to process. In this manner, we demonstrate 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 shows that the fixation features play 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.

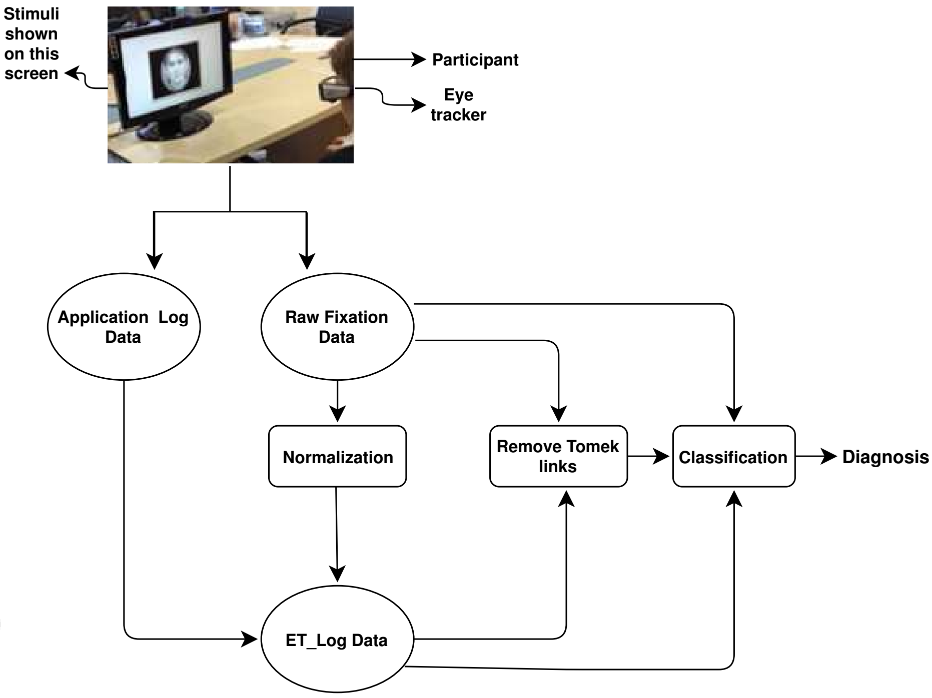


Figure 4.1 : Experimental setup and workflow of the presented approach for classifying the participants.

4.2 Methodology

We employ Random forest [47], Support Vector Machine (SVM) [49] and Logistic Regression (LR) [69] classifiers on the raw fixation (RF) and ET_log data (Figure 4.1). Three different classification problems are considered; ASD vs. Control (ASD/Control), ADHD vs. Control (ADHD/Control), ADHD vs. ASD (ADHD/ASD).

4.2.1 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 [70]. We use 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, the test participants are classified according to the majority voting score of all decision trees in the forest.

4.2.2 Support Vector Machines (SVM)

We use 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 [70]. 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 tries 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 have decided the value of C as 1 for ASD/Control and ADHD/ASD classification problems; but for ADHD/Control classification problem the best results are obtained when C is assigned to 10. Also we evaluate the radial basis function (RBF) kernel [71]. However, the linear kernel outperforms the RBF kernel for our dataset. Also, RBF requires more computational cost. For these reasons, we chose the linear kernel for SVM.

4.2.3 Logistic Regression (LR)

Logistic regression (LR) is a simple binary classification algorithm. Since we perform binary classification for diagnosing the participants, we use 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 takes *ADHD* and the *control* label names if the problem is ADHD/Control classification problem. LR algorithm learns this function for the classification:

$$P(y|x) = \frac{1}{1+\exp(-w^T x)} \equiv \sigma(w^T x) \tag{4.1}$$

where $P(y|x)$ computes the posterior probability of elements of y, and x represents the feature set of the dataset [70]. 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.

4.2.4 Pre-processing

For the pre-processing step, we use the Tomek links method, which is developed by Ivan Tomek [72] and is used in many research studies as a data cleaning technique [73, 74]. 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 remove noisy instances from the raw fixation and ET_log data (see Figure 4.1).

4.2.5 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 [75], Information Gain Ratio [75] and Gini Index [76]. We use entropy as a feature importance measure in this work. For node m, N_m instances reach node m and N_m^i represents instances which belong to class i (C_i). The probability of class C_i given x instance reaches node m calculated as follows:

$$P(C_i|x,m) \equiv p_m^i = \frac{N_m^i}{N_m} \tag{4.2}$$

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:

$$I_m = -\sum_{i=1}^{K} p_m^i log(p_m^i) \tag{4.3}$$

For each feature, the feature importance is calculated as follows [77]:

$$n_m = w_m I_m - w_{left(m)} I_{left(m)} - w_{right(m)} I_{right(m)} \tag{4.4}$$

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 4.3, 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:

$$feature\_importance_f = \frac{\sum_{m:nodemsplitsonfeaturef} n_m}{\sum_{s\varepsilon allnodes} n_s} \tag{4.5}$$

Figure 4.2 presents the feature importance values of the raw fixation data for each group. In this figure,Y axis shows the feature importance values and 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 values are shown with Red Bars and blue lines indicate the feature importance variance between trees of the Random forest. 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 4.3, the ET_log data features importance values for the Random forest algorithm are illustrated where Y axis shows the feature importance values and X axis shows feature ids which are presented in Table 4.1. Interestingly, the features of the APL data have lower feature importance value when compared to the fixation data features.

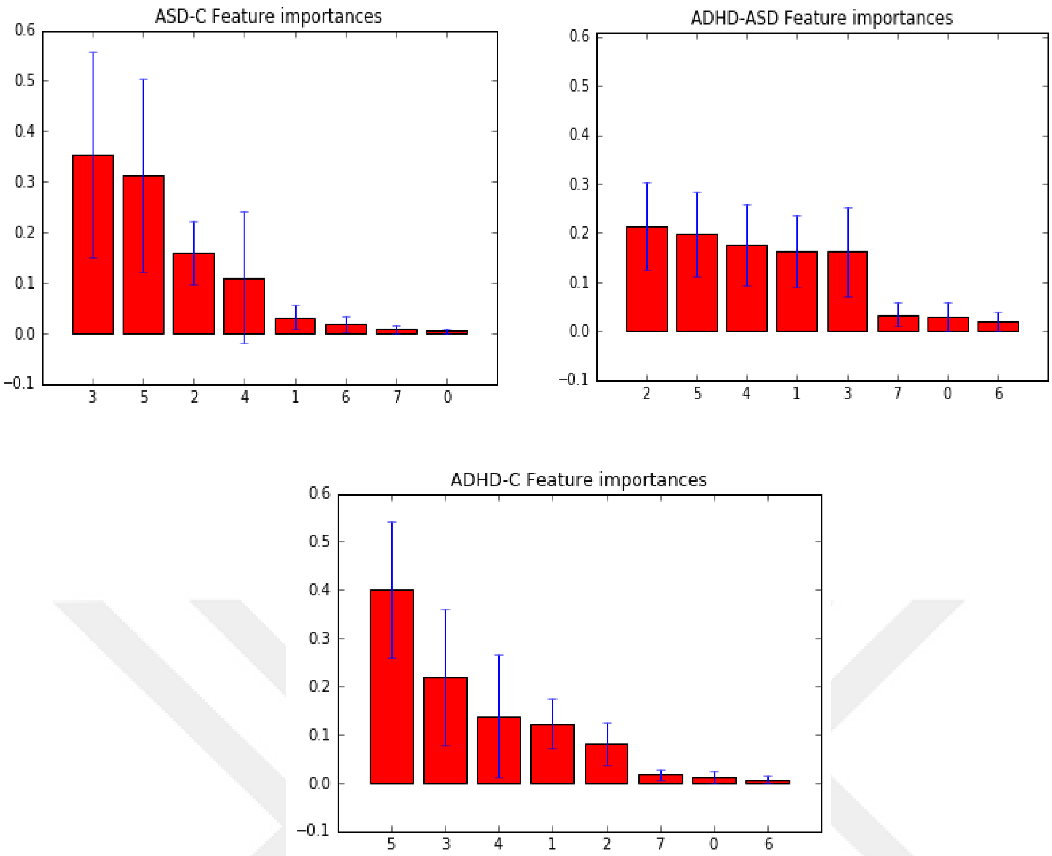


Figure 4.2 : Raw fixation data feature importance graphics. Y axis shows the feature importance values and X axis shows feature ids.

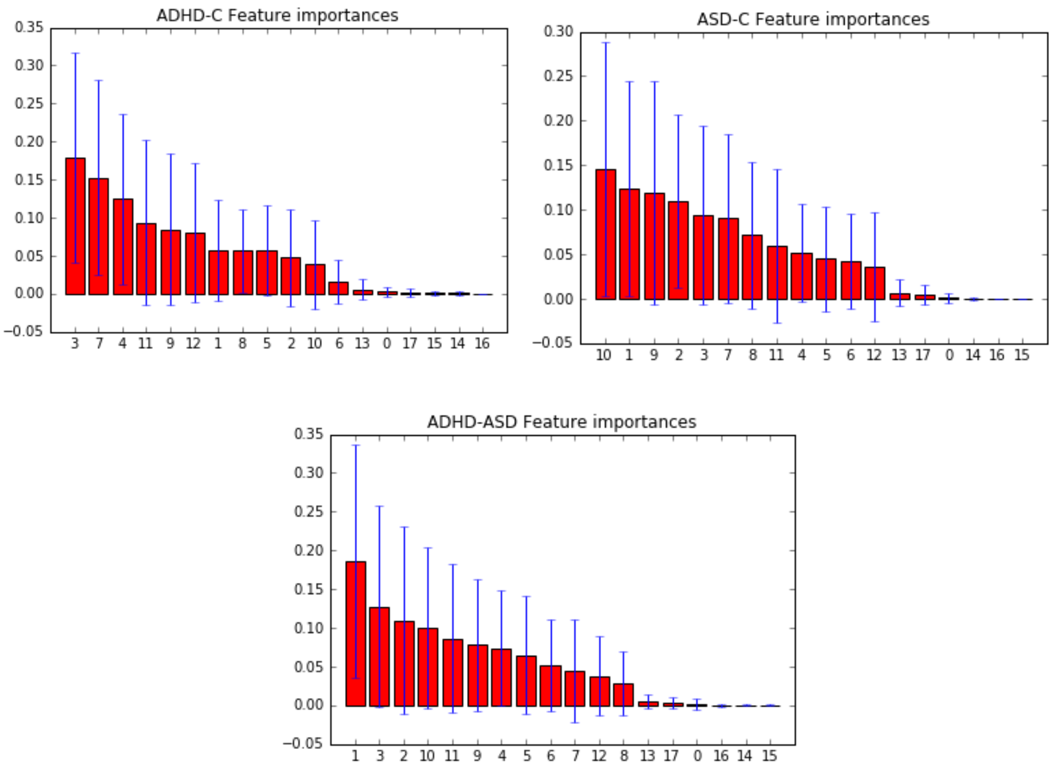


Figure 4.3 : ET_log data feature importance graphics. Y axis shows the feature importance values and X axis shows feature ids.

Table 4.1 : 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

4.3 Results

4.3.1 Evaluation metrics

In the previous sections, we have analyzed the effect of fixation and APL data on participants' emotional recognition process. In this section, we employ 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 [70] is used. All data instances of a participant are chosen as a test data, and the rest of the participants' data instances are used to train the model. This procedure is repeated for all participants. Thus, at each fold, diagnosis of a participant is predicted individually.

We measure the classification performance of an algorithm using main evaluation metrics; accuracy, sensitivity and specificity [70]. 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:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.6}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4.8}$$

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.

We have implemented the data pre-processing operation and classification algorithms by using scikit-learn v0.19.1 [77] machine learning library and our in-house Matlab and python scripts.

4.3.2 Results and discussion

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 4.2, Table 4.3 and Table 4.4 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 are produced by taking the average of the classification results of these features [70].

Table 4.2 : Classification results for ASD/Control (C).

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 4.3 : Classification results for ADHD/Control (C).

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 4.4 : Classification Results for ADHD/ASD.

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

According to Table 4.2, 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 algorithm accomplishs 80% specificity value on ET_log data. Removing the Tomek links from the datasets improves the performance of all classifiers except the sensitivity value which is produced by SVM on the RF data.

As shown in Table 4.3, for ADHD/Control classification, Random forest outperforms the LR and SVM. The entire control group is correctly detected when the Random forest is used together with the Tomek link removal on the ET_log data. The highest sensitivity value of ADHD/Control is achieved on the RF + ET_log.

ADHD/ASD classification performances are shown in Table 4.4. Removing Tomek links does not affect the Random forest performance on the RF data, with and without Tomek links the highest specificity ratio is 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 is the hardest problem among the others. Therefore, we can only reach 70.83% accuracy rate.

When the classifier results are considered, the ASD and ADHD groups are successfully separated from the control group. The Random forest algorithm performs better than LR and SVM. The highest accuracy scores are 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 [78, 79], they can not be distinguished from each other as well as the control group. Nevertheless, using Tomek links for outlier instance removal and the combination of the results of the feature sets increase the classification performance. When we compare the performances of the SVM and LR, we conclude that, LR achieves higher sensitivity, specificity and accuracy results for each feature type. For ADHD/Control classification we obtain 100% sensitivity value if Random forest is used with Tomek links removing technique on the ET_log dataset. Similarly, when the Tomek links are eliminated from the dataset, usage of RF + ET_log identify 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.

5. LSTM AND CNN BASED ANALYSIS OF EYE MOVEMENT DATA FOR DIAGNOSIS OF ASD AND ADHD

This chapter focuses on employing deep learning techniques for the classification process of the participants with ADHD and ASD. Eye tracker fixation data is used as a time series data and we give it as input to the produced LSTM and CNN models.

5.1 Background

Deficits in social interaction and social attitude are critical symptoms of children and adults with ASD or ADHD, increasing the importance of emotion recognition for people with these major neurodevelopmental disorders.

Emotional recognition behaviors of ASD and ADHD may be different from normal. Since ASDs have difficulties in establishing eye contact and they have visual inattention, they cannot get the necessary clues about other people's emotions from their eyes and face. Inattention and impulsivity in ADHD patients might cause not be able to perceive other people's emotions correctly. Especially, since they have visual inattention, they may have difficulties in receiving the information necessary for the communication [13]. Emotion recognition behavior of the normal and ADHDs; normal and ASDs can be used as a diagnostic criterion for ASDs and ADHDs [13, 30, 80, 81]. According to the literature [13] eye movements can be used as an instrument to clarify the nature of the emotion recognition process in human. Eye movements can be collected with eye tracker technology from the children with ADHD and ASD, and thus the behavior of these children during emotional recognition tasks can be examined.

Eye tracking technology has begun to be widely used in the fields such as psychology, neuroscience, marketing which require the exploration of the human visual system [82]. The effects of psychological diseases and environmental conditions on the human oculomotor system can be observed by these systems. In the present study, by using mobile eye tracking glasses, we record the participants' eye movements during an emotion recognition experiment. Since eye movements provide information about the psychological impairments of the patients [83], we investigate whether eye movements are discriminative between ADHD, ASD and control groups. There are two essential eye movements: fixation and saccade. Fixation requires focusing on the point being looked at for a while; hence the primary interest of the person can be determined. In the saccade movement, the places to be looked at quickly change and the eye movements are in the form of jumping.

In this work, we aim to contribute to diagnosis systems of ASD and ADHD diseases by using eye movements. Machine learning methods enable individuals with emotion recognition disorders to be classified at low cost. Besides diagnosis of the expert clinicians, these methods are able to present significant quantitative data to the experts as a result of classification [84]. This makes possible to have a measurable score for each participant.

The latest studies on ASD and ADHD classification tasks employ different machine learning algorithms. Generally, if the data type is f-MRI, support vector machine (SVM) is used [67, 85, 86]. Clinical data classification is implemented by using decision tree, Random Forest, SVM, Logistic Regression, Naïve Bayes [20, 56, 87]. However, when using temporal information such as eye movements, these hand-crafted feature extraction and classification techniques are insufficient. For this reason, we benefit from deep learning algorithms such as CNN and LSTM. Since LSTM considers long time dependencies when running on time series data, it succeeds in temporal data [88, 89]. We use CNN due to its considerable performance in the representation of the intrinsic features from the data [90, 91]. For this study, we create models by adding the CNN layer before the LSTM layer to extract important features from the data sets. In the scope of this section, as the first model we use only fixation features, as the second model in order to increase the dataset size we combine fixation and saccade features. Besides them in order to improve the representation of the eye movements changes, as the third model, we fuse the results after running fixation and saccade features on parallel CNN-LSTM layers. Finally, we present that combining fixation and saccade features and using them as a dataset improve the classification performance. We test our models on a dataset that we collect from the experiments and on an another dataset that we produce by adding noise to original dataset.

The contributions of this chapter are as follows:

- We use multiple types of representative features for the emotion recognition abilities of the participants.

- To the best of our knowledge, this is the first study that use eye movement data with LSTM and CNN algorithms to classify ASD and ADHD patients.

5.2 Related Work

5.2.1 Machine learning for ASD/ADHD diagnosis

Table 5.1 shows the datasets, methods and performance results of the studies that performed machine learning methods to diagnose ASD and ADHD. The majority of the studies tries to distinguish these clinical groups (ASD and ADHD) from the control group using fMRI data. Another research area that has been focused on is trying to decrease the number of psychological tests conducted to diagnose these diseases. In these studies, machine learning algorithms are used to determine important questions of the psychological tests. The commonly used algorithm is SVM.

Table 5.1 : Machine learning studies for ASD/ADHD/Control diagnosis prediction. (SN: Sensitivity, SP: Specificity, Acc: Accuracy).

Ref	Task	Methodology	Performance
[20]	Decreasing the time for application of ADOS (Autism Diagnostic Observation Schedule) Module 1. • **Participants in training data:** 612 Autism, 612 Non-spectrums • **Participants in test data:** AC (Autism Consortium): 110 Autism, 4 Non-spectrum, SSC (Simons Simplex Collection): 336 Autism, 0 Non-spectrum	Classification of individuals with autism and individuals without autism. • **Features:** ADOS data • **Classifiers:** Many machine learning algorithms but ADTree outperformed the others.	• Acc > 99% • SN > 99% • SP > 99%

Table 5.1 (continued): Machine learning studies for ASD/ADHD/Control diagnosis prediction. (SN: Sensitivity, SP: Specificity, Acc: Accuracy).

Ref	Task	Methodology	Performance
[16]	Flanker/NoGo task • **Participants:** 18 participants with ADHD and 18 control group • **Age:** 12-16 years	Classify adolescent with ADHD and control group. • **Features:** Structural Magnetic resonance imaging data • **Classifier:** SVM • **Validation:** Leave-one-out	• Acc=77.78% • SN=77.78% • SP=77.78%
[19]	Reach-to-drop task • **Participants:** 15 children with ASD, 15 normal children • **Age:** 2-4 years	Kinematic Analysis • **Features:** Upper-limb movement data • **Classifier:** SVM • **Validation:** Leave-one-out	• Acc=96.7%
[87]	Selecting a subset of ADOS behaviors to classify children with ASD and control. • **Number of participants:** 4540 individuals	Classification of individuals with autism and individuals without autism. • **Features:** ADOS data • **Classifiers:** Eight machine learning algorithms • **Validation:** 10-fold cross validation	• Acc = 98.27% for module 2 • Acc = 97.66% for module 3
[56]	Selecting a subset of SRS (Social Responsiveness Scale) behaviors to classify children with ASD and children with ADHD. • **Participants:** ASD (n = 2775) ADHD (n = 150).	Classification of individuals with ASD and individuals with ADHD. • **Features:** 65-item SRS • **Classifiers:** Six machine learning algorithms • **Validation:** 10-fold cross-validation	• Acc = 96.5%

Table 5.1 (continued): Machine learning studies for ASD/ADHD/Control diagnosis prediction. (SN: Sensitivity, SP: Specificity, Acc: Accuracy).

Ref	Task	Methodology	Performance
[86]	• **Participant group1:** 15 adolescent and adults with autism, 15 control participants • **Mean age:** 21.66 • **Participant group2:** 22 adults with autism, 22 control participants • **Ages:** 18-42 years	Classification of individuals with ASD and control • **Features:** fMRI data. • **Classifiers:** SVM	• Acc= 95.9%
[67]	• **Participants:** 22 adults with autism, 22 control participants • **Ages:** 18-42 years	Whole-brain classification • **Features:** MRI scans • **Classifier:** SVM • **Validation:** Leave-one-out	• SP = 86% • SN = 88%
[68]	ADHD-200 competition data, • **Participants:** 491 control individuals, 285 individuals with ADHD.	Classification of participants with ADHD versus control group. • **Features:** fMRI data • **Classifier:** SVM-RFE • **Validation:** 10-fold cross-validation	• Acc = 55% • SP = 80% • SN = 33%
[85]	• **Participants:** 1,264 verbal individuals with ASD and 462 control • **Age:** below and above 10	Targeting best-estimate clinical diagnosis of ASD vs. non-ASD • **Features:** Autism Diagnostic Interview-Revised and SRS scores. • **Classifier:** SVM • **Validation:** multi-level cross-validation	Below age 10: • SN = 89.2% • SP = 59.0% Above age 10: • SN = 86.7% • SP= 53.4%

Table 5.1 (continued): Machine learning studies for ASD/ADHD/Control diagnosis prediction. (SN: Sensitivity, SP: Specificity, Acc: Accuracy).

Ref	Task	Methodology	Performance
[92]	ADHD-200 MRI dataset, • **Participants:** 67 control, 67 participants with ADHDI (ADHD inattentive type), 67 participants with ADHDC (ADHD combined type) • **Age:** 7-14 years	Classification of participants with ADHD versus control group. • **Features:** fMRI • **Classifier:** linear extreme learning machine (ELM) and SVM with linear and radial basis function (RBF) kernels	• Acc = 92.85%

5.2.2 Long short term memory based diagnosis

LSTM is one of the most robust deep learning frameworks on time series classification studies. In [84], they classify individuals with ASD and typical controls by using LSTM on resting-state fMRI time series data. They use an LSTM method that produces an output at the end of each LSTM cell. Thereby at each time step, they obtaine a clearer analysis of the rsfMRI signal, instead of using only one final output for the sequence. We use a similar method, a dense layer with a single node is connected to the output of each repeating LSTM cell. [84], reached 68.5% accuracy. To validate the performance, they used 10-fold cross-validation.

LSTM is a subtype of RNNs. [93] develops a generic predictive model by using RNN algorithm to anticipate the diagnoses, medication categories and the next visit to the hospital of the patients. Just as we try to solve the problem of disease detection by taking advantage of the temporal feature of our eye tracker data, they try to solve a multi-label prediction problem by using temporal dynamic behavior for a time sequence.

RNN has become very popular methods for learning from sequential data. LSTM, which is a kind of RNN and has extra hidden units in addition to the RNN, ensures that the information in the data is stored in the neural network memory for a longer period of time. In [94], they take the studies on EHR dataset one step further by using the LSTM model, and they outperform [93].

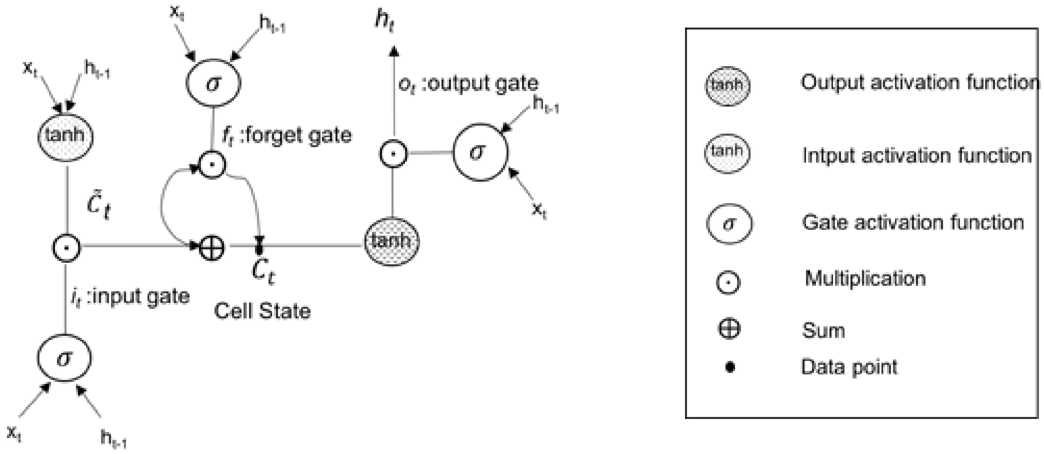


Figure 5.1 : A memory block of LSTM.

On the other hand, in [95], they focus on the diagnosing dyslexia disorder. As a dataset, they examine the eye gaze data of the participants with dyslexia. According to their results, the performance of the feature based machine learning methods is better than the sequential approach LSTM.

In [96] study LSTM is used, because FCN (Fully Convolutional Networks) is not successful enough to learn the temporal feature in the diagnosis of small bowel disease. Therefore, we use CNN algorithm only for extracting important features from the data, and then the LSTM model performs the learning process.

5.3 Methodology

Before describing the proposed models, the fundamental theories of LSTM and CNN methods will be explained for preliminary information.

5.3.1 Cell structure of the LSTM

In this section, we introduce the cell structure of the LSTM model. LSTM architecture includes recurrently connected memory blocks. Each memory block consists of a memory cell which lets information to flow along it. Besides, to select and control the cell state, LSTM includes input, forget and output gates inside the memory blocks.

Figure 5.1 shows a memory block of an LSTM. Each gate takes a current input fixation data x_t and the previous hidden vector $h_{(t-1)}$ as inputs. At each time step, cell state of the memory block is updated according to the gate update formulas given below:

$$
\begin{aligned}
f_t &= \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \\
i_t &= \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \\
\widetilde{C}_t &= (W_{\widetilde{C}x}x_t + W_{\widetilde{C}h}h_{t-1} + b_{\widetilde{C}}) \\
C_t &= f_t * C_{t-1} + i_t * \widetilde{C}_t \\
o_t &= \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \\
h_t &= o_t * tanh(C_t)
\end{aligned}
\tag{5.1}
$$

where f_t, i_t, o_t are forget gate, input gate, output gate at t time step respectively. Also, σ represents a sigmoid function, it produces values between 0 and 1. The weights of the current inputs are shown as W_{fx}, W_{ix}, $W_{\widetilde{C}x}$, W_{ox} and the weights of the current hidden states are as follows, W_{fh}, W_{ih}, $W_{\widetilde{C}h}$, W_{oh}. Besides, b_f, b_i, $b_{\widetilde{C}}$, b_o are bias values.

The most important ability of the LSTM is long-term memory. Thereby, this algorithm can make an accurate estimation of sequential data. LSTM needs to throw away some redundant information from the cell state and the forget gate takes this duty. The forget gate gives the input x_t and previous hidden vector h_{t-1} to a sigmoid function and produces an output between 0-1. If the output is 0, all information in the previous cell state should be removed; if the output is 1, all information in the previous cell state should be kept and transferred to the next layers; and if the output is a value between 0-1, some of the information is kept and some of it is removed.

The input gate, which is a sigmoid layer, and a hyperbolic tangent layer, $\widetilde{C}_t$, determine the new value of the next cell state. Some information from the previous cells should be forgotten, and only necessary parts of the new input data should be used to update the new cell state. Thereby, information to be forgotten is not transferred to the next layers, only determined information is transferred to the subsequent layers.

The last one, output gate, defines what parts of the cell state will be used as an output by running a sigmoid function. Then, cell state is passed to a hyperbolic tangent function to scale the values between -1 and 1. In the end, by pointwise multiplication operation of the output and cell state, the output of the LSTM memory block, h_t, occurs. The value of the h_t is both sent to the next LSTM memory block and can be used as an output for that LSTM memory block. Thus, LSTM model carries information through the long time.

5.3.2 Structure of the CNN

Convolutional Neural Network (CNN) is used for training from scratch, transfer learning and feature extraction processes. We used CNN for feature extraction operation in the current work. CNN is a structure composed of multiple convolution layers, non-linear activation functions in these layers and pooling layers [90, 97]. In convolution operation, a filter is applied to the matrix-shaped data. In conventional feed forward neural networks, all neurons in a layer are tied to all the neurons of the next layer. Those kind of layers are called fully connected layers. In addition to fully connected layers, convolution is applied to the input layer to produce the output for CNN model. Thereby, all regions in the input layer are connected to neurons in the next layer. Each layer uses different filters, and finally, outputs are combined. The element-wise non-linear functions such as ReLU, sigmoid, tangent hyperbolic are used. After convolution layers, pooling layers are applied. The pooling layers subsample the input data. In the pooling layers max, min or average pooling operation is implemented to the output of each filter. As a result of the pooling process, a fixed size matrix is obtained, and the size of the output is reduced considerably. Throughout the training, CNN automatically learns the features. Thus, the feature representations of the data are learned. On the contrast, the feature-based classification algorithms use handcrafting gaze based features [98]. Therefore, usually, CNN outperforms the other handcrafted feature-based algorithms [98].

5.4 The Proposed Models

5.4.1 Identification of the time step

To determine the time step for the LSTM, we take advantage of the number of eye movements that the participants produce when looking at the images shown. The eye tracker measured that each participant produce different numbers of fixation and saccade eye movements data during the experiment. According to the histograms in Figure 5.2, when the participants look at one of the shown images, usually they focus on 10-20 points on the image for a while. First, we demonstrate the number of instances with fixation count values less than 60 in Figure 5.2 a). As you can see, most of the samples have around 10 fixation points. To get a more detailed view of this area, in Figure 5.2 b) we propose the histogram of instances whose fixation count is less than 20. It is observed that fixation count values are concentrated between 5 and 8. Going a step further, the distribution of samples with a fixation number less than 15 is examined in Figure 5.2 c) and it appears that the participants generally do not have less than 5 fixations. As a result, we decide to carry out the analyses by selecting time step values as 5, 10 and 15.

5.4.2 Truncation and padding of the dataset

Due to the nature of the LSTM method, the data to be used as input must be in a specific format. In this work, we choose the time step as a predetermined value. For our data set, the fixation and the saccade data on each picture of each user are used as the time step, and this value is set to 15, 10 and 5 according to the histograms in Figure 5.2.

However, the number of fixation and saccade samples may be different for participants, in which case we need to reduce or increase the number of samples in order to equalize them for everyone. Padding and truncation operations solve this problem. We present a new padding and truncation methods. When the data count is missing, we pad the dataset by using the mean value of the existing instances. For the truncation procedure, we develop a more intelligent method. The Euclidean distance is calculated between successive fixation/saccade points. Two points with little difference are averaged and used as a single sample instead of two of them. This process continues until the target number of the time step is reached. Thus, the diversity in the data is not degraded since truncation is performed by discarding the data sample with the closest value.

Since the eye movement data is in a temporal structure, the temporal dependency of the samples on each other is considered. We aim to determine the temporal relationship between the eye movement instances that we use as time series. According to this relationship we classify the participants. At each time step, the fixation points occurred at that time step are used for model training. We convert our data into a tensor-shaped so that our model learns the temporal relations between instances. For example, when t=1 we observe the relationship between all training participants' first fixation data points that occurred on the 40 images during the experiments. For all other time

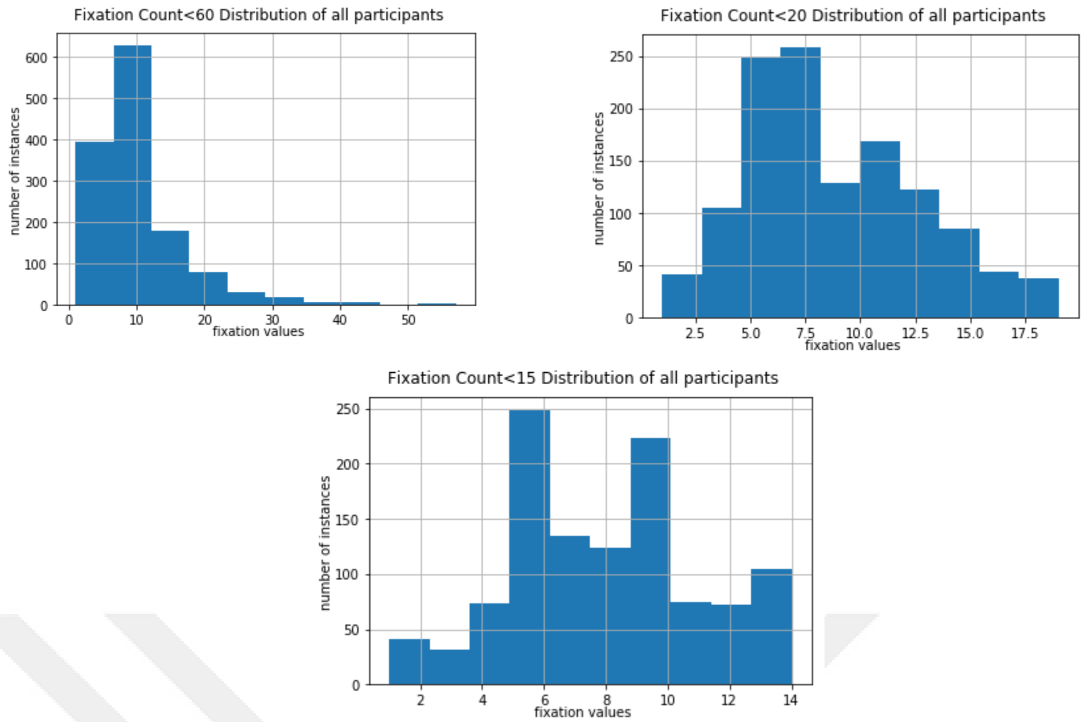


Figure 5.2 : Fixation distributions for all participants.

steps, we apply this strategy to our model to learn the data. Thus, we tackle that eye movement in time series are useful in distinguishing the participants.

5.4.3 CNN-LSTM model

The primary goal of this work detects the diagnosis of the participants. Since CNN is a robust algorithm to determine and extract abstract and hidden features of the data, we use it to obtain these hidden features before feeding the data into the LSTM layer. In the convolutional part, there is one convolutional layer and it has 30 convolutional filters which represent the dimensionality of the output space. Besides, there exists kernel parameter that specifies the length of the 1D convolution window and we chose it as 3 in this work. As an activation function ReLU is used in the convolutional layer. It calculates below function:

$$f(x) = max(0,x) \tag{5.2}$$

If the input value is less than zero, the output of the ReLU is zero. Otherwise, the output is equal to the input value. Due to its non-saturating form, ReLU converges faster than

Table 5.2 : CNN parameters.

#filter	30
Filter size	3x3
Activation function	ReLU
Pool size	3

tanh function [90]. Therefore, it is commonly used by deep learning networks for the hidden layers.

In the third model, we use max pooling layer with the CNN layer to prevent overfitting. Also, max pooling provides downsampling of an input representation. Furthermore, max pooling generalizes the outputs of the CNN layer, especially in image processing it discovers the features invariant to scale and orientation changes. Therefore, it acts as a transformer which has the ability to reach from low level feature to higher level feature. The pool size is selected 3 in this study, so maxpooling layer pools the interpretation of CNN layer into 3 blocks.

To interpret and learn the model of the data we use the LSTM algorithm which is a type of Recurrent Neural Network (RNN). LSTM method can use its reasoning about previous events in the data to inform later ones. The strength of the LSTM algorithm is to persist the information in it. Here, LSTM layer includes 15 consecutive cell units.

We investigate the diagnosis detection for the participants with ADHD or ASD using eye tracking data collected during an emotion recognition experiment. In order to solve this problem, we transform the fixation and saccade data into a proper format so that time series classification methods can be applied. This section describes the LSTM model for participants' diagnosis problem. We explain the LSTM RNNs model with input and output gates [88]. And also it uses forget gates [89] that provides the cell memory to be forgotten or appropriately regularized. Thereby, cell state is updated before using as an input. Besides, in order to improve the quality of the extracted features, we add a CNN layer to our network [99]. Then these improved features are given to the LSTM layer. As in the previous studies [94], we use a fully connected layer as an output layer. Since we implement binary classification problems such as ASD/Control, ADHD/Control, and ASD/ADHD, the sigmoid function is used as an activation function and binary cross-entropy is used as the loss function. Besides, Adam optimizer method is employed for the optimization step.

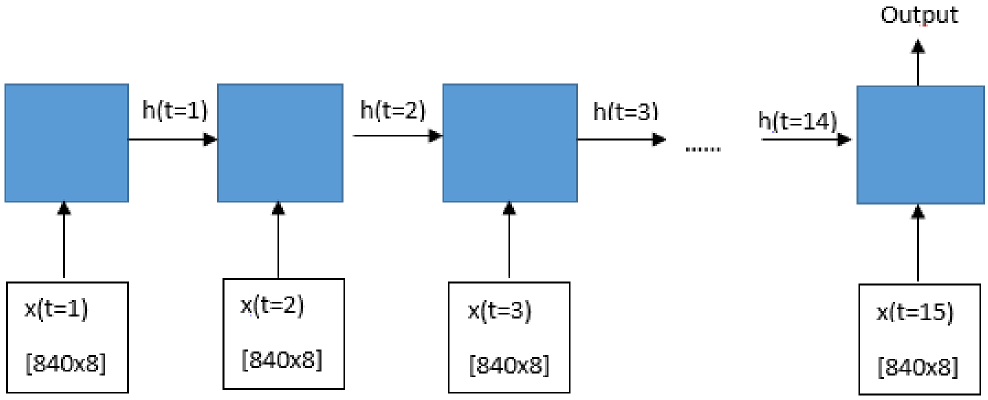


Figure 5.3 : The architecture of the LSTM model. Each blue box is an LSTM memory block that receives input x(t) and the previous output h(t-1) at each timestep t.

Figure 5.3 shows the way that we employ the LSTM for diagnosis prediction. When we use only fixation data, each memory block takes 840x8 vector (8 is feature size) as an input vector at each time step. Since we use 15 LSTM units, the input vector is transformed into an 840x15 vector.

The likelihood function of our model can be written as follows:

$$l(w|X) = \prod_{n}^{N} (y^n)^{r^n} (1-y^n)^{(1-r^n)} \tag{5.3}$$

where w where is weight of the inputs X. N stands for the size of the sequence. y^n is the sigmoid function which is equal to $\frac{1}{1+exp(-wx^n)}$, r^n is the actual label, $r^n \varepsilon \{0,1\}$. This likelihood function must be maximized, therefore log likelihood of this function is calculated. We have binary cross-entropy:

$$loss(l(w|X)) = -\frac{1}{|N|} \sum_{n}^{|N|} r^n log(y^n) + (1-r^n) log(1-y^n) \tag{5.4}$$

Since the eye movement data is in a temporal structure, the temporal dependency of the samples on each other is considered in this work. We use eye movements as time series data. We aim to determine the temporal relationship between the eye movement instances. According to this relationship, we classify the participants. In order to do that, we employ pre-determined time step values such as 15, 10 and 5 . If we assume that there are 15 fixation or saccade data points for each of the 40 shown images, the fixation points formed at that time step are used for model training. We convert our data into a tensor-shaped so that our model learns temporal relations between the instances.

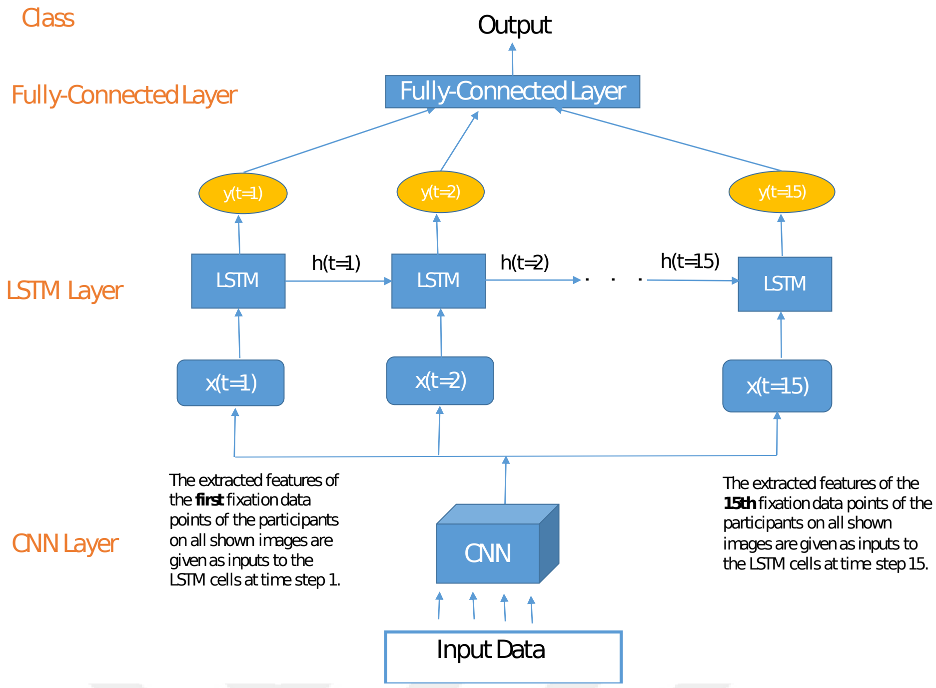


Figure 5.4 : The architecture of the CNN-LSTM model.

For example, when t=1 we observe the relationship between all training participants' first fixation data points that occurred on 40 images during the experiments. And for all other time steps (t = 1..15) we apply the same procedure to make our model learn the data. The output of the first cell is transferred to the next LSTM cell. This continues until the last time step is reached. At first, the training set is a 12600x8 vector where 12600=21*40*15, since there are 21 participants in the training set for ASD/Control classification problem, each of the participants are shown 40 images and we assume on each image 15 fixation points are occurred by each participant. 8 is feature size of the fixation data. After LSTM performs on the data, training dataset becomes 12600x15, since as LSTM cell size we choose 15, it means our model represents existing features in a more larger dimension. Additionally, test data has only one participant, therefore the test data is a 600x8 vector where 600=40*15. Thanks to the forget gate and the output gate in the LSTM cell structure through all time steps, some of the learned data is forgotten and some of the learned data is transferred to the other cells. For example, if ASD/ADHD classification is performed, LSTM tries to preserve uncommon features of the ASD and ADHD groups, and endeavors to forget common characteristics of them.

Table 5.3 : Model parameters.

algorithmName	LSTM, CNN_LSTM, CNN_maxpool_LSTM
timeStep	5ts, 10ts, 15ts
dataType	f: fixation, fs: fixation+saccade, fsl: fixation_layer+saccade_layer

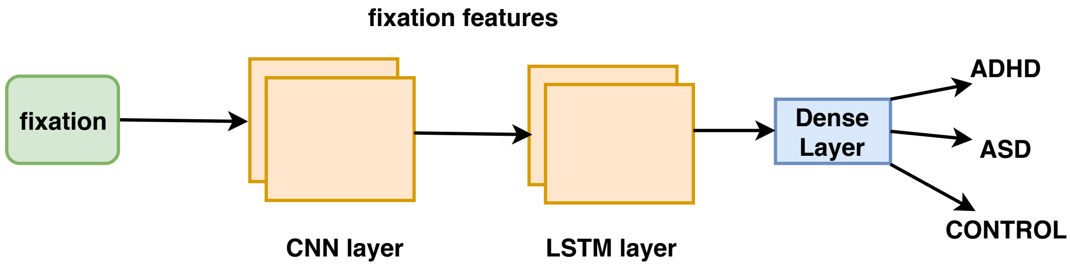


Figure 5.5 : CNN-LSTM model with only fixation features.

5.4.4 CNN-LSTM architecture with a single pipeline

As a baseline model we design a simple model that is shown in Figure 5.5 and Figure 5.6. The CNN layer is used in the front end of the LSTM layer with a sigmoid dense layer on the output. The first layer is CNN layer produces representative features and they are sent to the LSTM layer. In Figure 5.5 as an input to the model only fixation features are given. To increase the feature size, fixation and saccade features are combined and used as one dataset in Figure 5.6.

We propose 3 different models by using CNN and LSTM algorithms. A template for naming our models is as follows: *algorithmName_timeStep_dataType*. These parameters take the values demonstrated in Table 5.3. The first model includes only an LSTM layer (model naming parameter for algorithm: LSTM), the second one utilized a CNN layer before the LSTM layer for extract hidden representation of the data (model naming parameter for algorithm: CNN_LSTM). Finally, in addition to CNN and LSTM layers, the third model contains max pooling down sampling strategy in CNN (model naming parameter for algorithm: CNN_maxpool_LSTM). Besides these, all models have a flatten layer and a dense layer.

5.4.5 CNN-LSTM architecture with two parallel pipelines (Fusion of the layers)

We design a CNN-LSTM model with two parallel pipelines. This model takes the fixation and the saccade data as inputs. The number of LSTM cells, CNN filter size and used time steps in each pipeline are equal, but the size of the features are different.

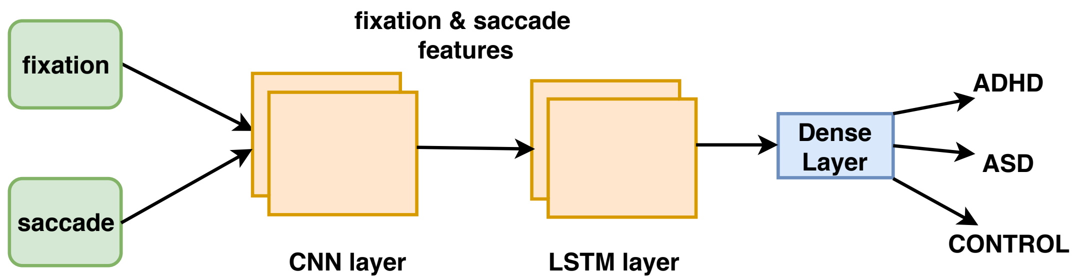


Figure 5.6 : CNN-LSTM model with fixation and saccade features.

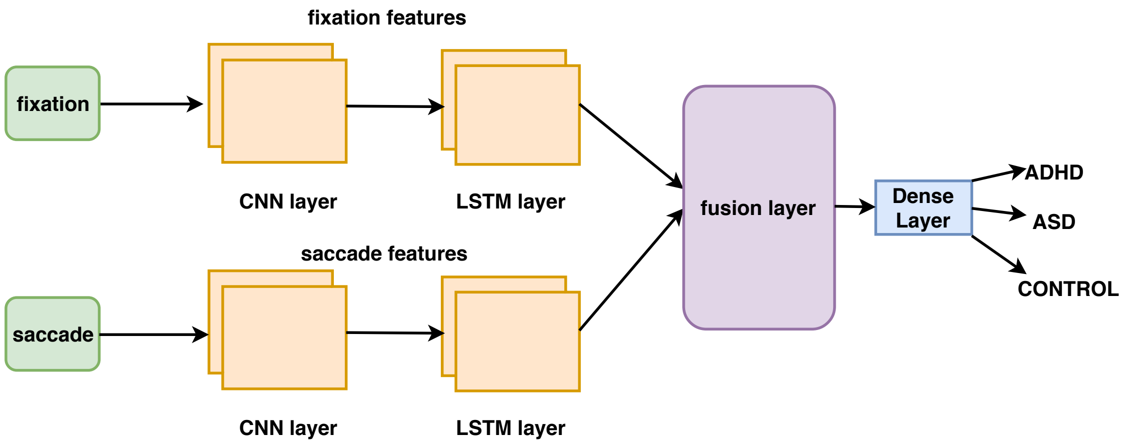


Figure 5.7 : Fusing two CNN-LSTM parallel layers.

After temporal relationships between the features are revealed, we merge the results from both LSTM layers.

By proposing this model, we aim to investigate whether the saccade features and the fixation features have different nature. If so we should have come up with different results when we combine fixation and saccade features into a single dataset and give it to the single pipeline and when we give them separately two parallel pipelines and fuse the representation of each pipeline outputs.

5.5 Results

5.5.1 Noisy dataset

In order to measure the robustness of our models, we generate a new dataset by adding Gaussian noise to the fixation and saccade datasets. We produce a noise vector for each feature vectors of the dataset, by taking into account the standard deviation value of each feature of the dataset and then multiplied it by a random scalar drawn from the standard normal distribution. Then, we sum up the generated noise vector with the feature vector of the original dataset and finally obtain our noisy dataset.

5.5.2 Experimental results and discussion

We use Keras library for training the LSTM and CNN models. Leave-one-out cross-validation technique [70] is conducted for the classifier evaluation. All data instances of a participant are chosen as test data, and the rest of the participants' data instances are used to train the model. This procedure is repeated for all participants. Thus, at each fold, diagnosis of a participant is predicted individually. We conduct ASD/Control, ADHD/Control and ADHD/ASD classification experiments. In the first two experiments, we try to distinguish the participants in the control group from the disordered participants. The most challenging problem is to discriminate the participants with ASD from the participants with ADHD, since ASD and ADHD have many common characteristics.

In order to evaluate the performance of the classifiers, we utilize the advantage of the sensitivity (Equation 4.7), specificity (Equation 4.8) and accuracy (Equation 4.6) measurements. For our study, sensitivity shows the probability of a participant being disordered when s/he has a disorder. If we need to explain better, sensitivity is the ability to characterize a participant with ASD or ADHD as a disordered individual. On the other hand, specificity presents the probability of being a typical development participant when s/he is a normal individual.

We present the classification results of the proposed models in Table 5.4, Table 5.5, Table 5.6, Table 5.7, Table 5.8, Table 5.9.

According to Table 5.4 and Table 5.5, sensitivity values are higher than specificity values on both noisy dataset and original dataset for ASD/Control classification. This situation indicates that it is easier to detect participants with ASD than the control group. The control participants are more commonly misclassified than the participants with ASD. On the other hand, since the fixation data contains very strong features even if it is used without any other feature type, the classification results are satisfying.

It is desirable for us to have high sensitivity and specificity values in the classification process. Because both participants with ASD and the control participants should be

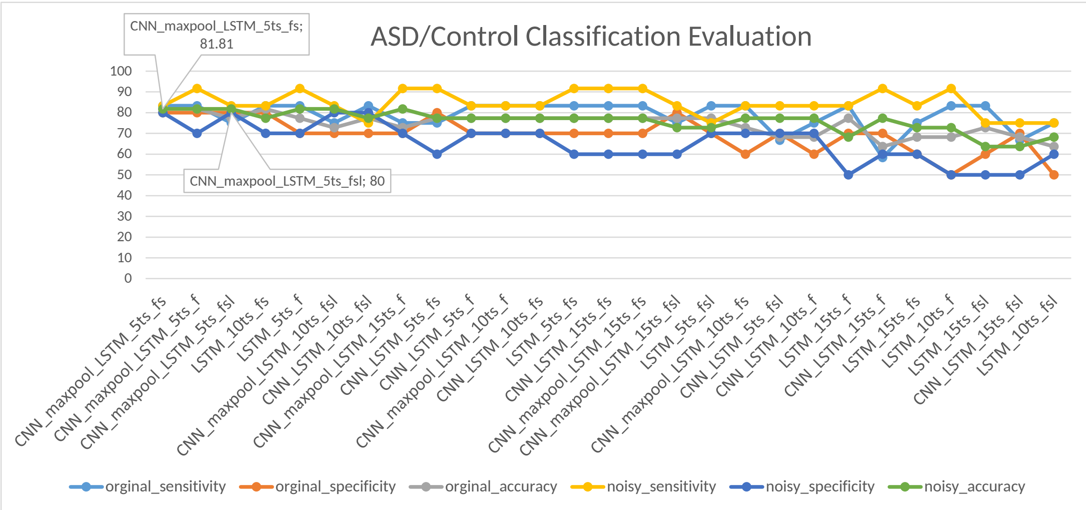


Figure 5.8 : The evaluation of ASD/Control classification.

classified correctly. According to the Figure 5.8, high sensitivity and high specificity values are observed when we use a single pipeline which includes a combination of CNN and LSTM layers with applying the maxpool process after the CNN layer. Also time step should be chosen as 5. Using the fixation features and the saccade features together improves the performance.

When the results obtained on both the noisy and the original datasets are examined, combining fixation and saccade features and giving them into a single CNN-LSTM pipeline as an input yields high classification performance. On the other hand, when we use two parallel CNN-LSTM pipelines and provide the fixation features to the first pipeline and saccade features to the other pipeline, we achieve considerably good results in terms of sensitivity, specificity and accuracy.

Table 5.4 : The classification results for ASD/Control on the original dataset.

	fixation			fixation+saccade			fixation_layer+saccade_layer		
Classifier/time step	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
LSTM_5ts	83.33	70.00	77.27	83.33	70.00	77.27	83.33	70.00	77.27
CNN_LSTM_5ts	83.33	70.00	77.27	75.00	80.00	77.27	66.67	70.00	68.18
CNN_maxpool_LSTM_5ts	83.33	80.00	81.82	**83.33**	**80.00**	**81.82**	75.00	80.00	77.27
LSTM_10ts	83.33	50.00	68.18	83.33	80.00	81.82	75.00	50.00	63.64
CNN_LSTM_10ts	75.00	60.00	68.18	83.33	70.00	77.27	83.33	70.00	77.27
CNN_maxpool_LSTM_10ts	83.33	70.00	77.27	83.33	60.00	72.73	75.00	70.00	72.73
LSTM_15ts	83.33	70.00	77.27	75.00	60.00	68.18	83.33	60.00	72.73
CNN_LSTM_15ts	58.33	70.00	63.64	83.33	70.00	77.27	66.67	70.00	68.18
CNN_maxpool_LSTM_15ts	75.00	70.00	72.73	83.33	70.00	77.27	75.00	80.00	77.27

Table 5.5 : The classification results for ASD/Control on the noisy dataset.

	fixation			fixation+saccade			fixation_layer+saccade_layer		
Classifier/time step	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
LSTM_5ts	91.67	70.00	81.82	91.67	60.00	77.27	75.00	70.00	72.72
CNN_LSTM_5ts	83.33	70.00	77.27	91.67	60.00	77.27	83.33	70.00	77.27
CNN_maxpool_LSTM_5ts	91.67	70.00	81.82	**83.33**	**80.00**	**81.82**	83.33	80.00	81.82
LSTM_10ts	91.67	50.00	72.73	83.33	70.00	77.27	75.00	60.00	68.18
CNN_LSTM_10ts	83.33	70.00	77.27	83.33	70.00	77.27	75.00	80.00	77.27
CNN_maxpool_LSTM_10ts	83.33	70.00	77.27	83.33	70.00	77.27	83.33	80.00	81.82
LSTM_15ts	83.33	50.00	68.18	83.33	60.00	72.73	75.00	50.00	63.64
CNN_LSTM_15ts	91.67	60.00	77.27	91.67	60.00	77.27	75.00	50.00	63.64
CNN_maxpool_LSTM_15ts	91.67	70.00	81.82	91.67	60.00	77.27	83.33	60.00	72.73

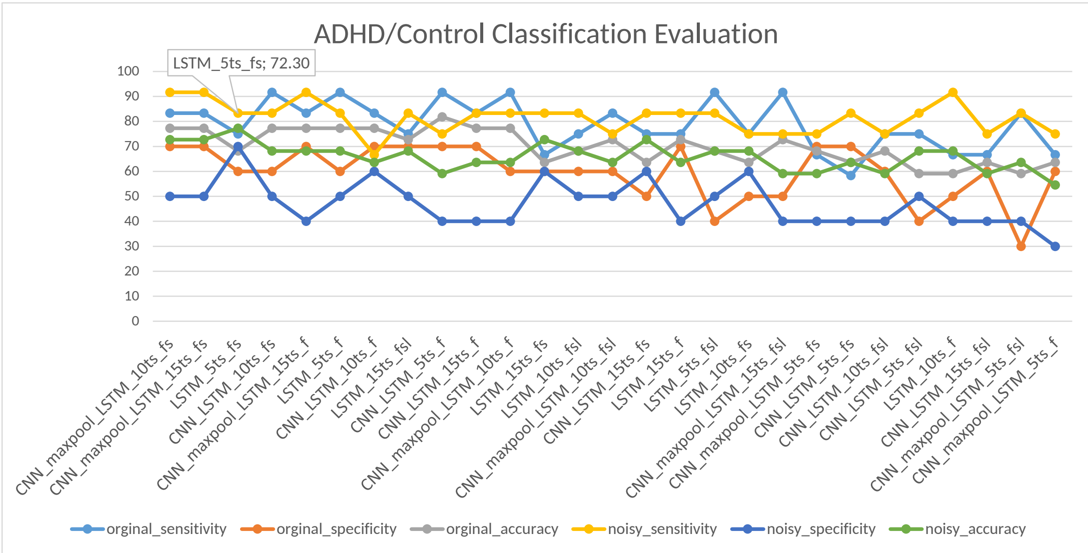


Figure 5.9 : The evaluation of ADHD/Control classification.

According to the results (see Table 5.6, Table 5.7 and Figure 5.9) obtained for the ADHD/Control classification, the most efficient model is the LSTM_5ts_fs. Considering both the noisy dataset and the original dataset, the ADHD/Control classification is found to be a more difficult problem than ASD/Control. The reason is that the control group participants have more common features with the participants with ADHD than the participants with ASD.

The hidden representation of the dataset that CNN algorithm reveals, may cause overfitting and negatively affect the performance. Therefore, the models only using the LSTM becomes more successful for ADHD/Control classification.

Increasing the number of features by combining fixation and saccade affects the results positively. As far as we observe from the results, choosing the time steps small increases the learning performance of the models. Therefore when the time step is selected as 5, the best ADHD/Control results are obtained such as in ASD/Control classification.

Table 5.6 : The classification results for ADHD/Control on the original dataset.

	fixation			fixation+saccade			fixation_layer+saccade_layer		
Classifier/time step	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
LSTM_5ts	91.67	60.00	77.27	**75.00**	**60.00**	**68.18**	91.67	40.00	68.18
CNN_LSTM_5ts	91.67	70.00	81.82	58.33	70.00	63.64	75.00	40.00	59.09
CNN_maxpool_LSTM_5ts	66.67	60.00	63.64	66.67	70.00	68.18	83.33	30.00	59.09
LSTM_10ts	66.67	50.00	59.09	75.00	50.00	63.64	75.00	60.00	68.18
CNN_LSTM_10ts	**83.33**	**70.00**	**77.27**	91.67	60.00	77.27	75.00	60.00	68.18
CNN_maxpool_LSTM_10ts	91.67	60.00	77.27	83.33	70.00	77.27	83.33	60.00	72.73
LSTM_15ts	75.00	70.00	72.73	**66.67**	**60.00**	**63.64**	75.00	70.00	72.73
CNN_LSTM_15ts	83.33	70.00	77.27	75.00	50.00	63.64	66.67	60.00	63.64
CNN_maxpool_LSTM_15ts	83.33	70.00	77.27	83.33	70.00	77.27	91.67	50.00	72.73

Table 5.7 : The classification results for ADHD/Control on the noisy dataset.

	fixation			fixation+saccade			fixation_layer+saccade_layer		
Classifier/time step	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
LSTM_5ts	83.33	50.00	68.18	**83.33**	**70.00**	**77.27**	83.33	50.00	68.18
CNN_LSTM_5ts	75.00	40.00	59.09	83.33	40.00	63.64	83.33	50.00	68.18
CNN_maxpool_LSTM_5ts	75.00	30.00	54.55	75.00	40.00	59.09	83.33	40.00	63.64
LSTM_10ts	91.67	40.00	68.18	75.00	60.00	68.18	83.33	50.00	68.18
CNN_LSTM_10ts	**66.66**	**60.00**	**63.63**	83.33	50.00	68.18	75.00	40.00	59.09
CNN_maxpool_LSTM_10ts	83.33	40.00	63.64	91.67	50.00	72.73	75.00	50.00	63.64
LSTM_15ts	83.33	40.00	63.63	**83.33**	**60.00**	**72.73**	83.33	50.00	68.18
CNN_LSTM_15ts	83.33	40.00	63.63	83.33	60.00	72.73	75.00	40.00	59.09
CNN_maxpool_LSTM_15ts	91.67	40.00	68.18	91.67	50.00	72.73	75.00	40.00	59.09

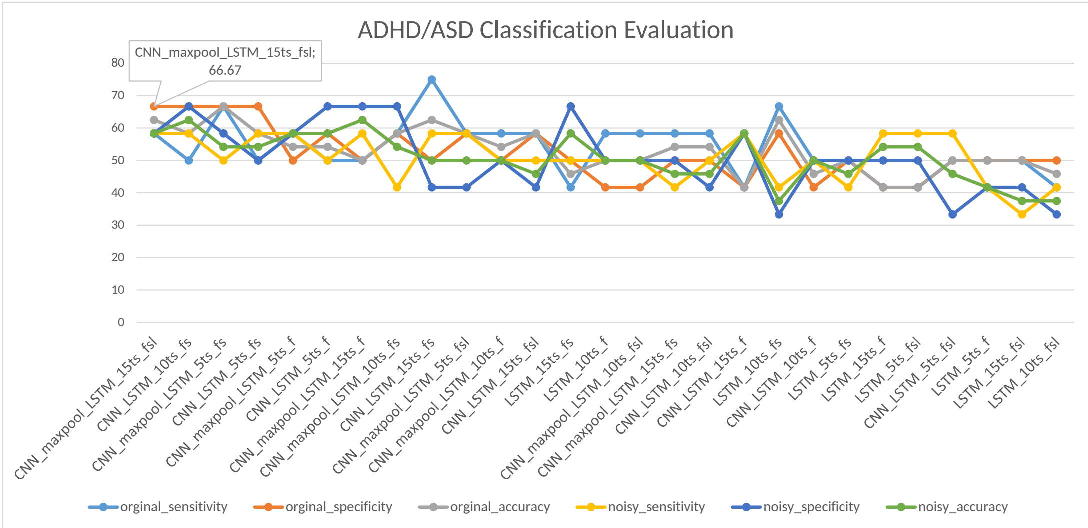


Figure 5.10 : The evaluation of ADHD/ASD classification.

The sensitivity value for ADHD/ASD classification shows the correct estimation rate of participants with ADHD, while specificity indicates the probability of correct classification of participants with ASD. In ASD/Control and ADHD/Control classifications, the correct estimation of the control group is calculated as specificity and the accurate prediction of the disordered groups as sensitivity.

The results achieved in both the noisy dataset and the original dataset are lower than the results obtained from the other two classification problems (see Table 5.8, Table 5.9). Although widely accepted, it suffers from some limitations due to participants with ADHD and participants with ASD have common psychiatric symptoms [2].

Figure 5.10 illustrates that the disordered participants with ADHD and ASD show similar eye movements when they look at emotion-expressing images, and this situation yields a challenge for our proposed models. Therefore, our methods have difficulty in producing good results. Nevertheless, the results demonstrate that the most successful outcomes for all measurements are provided by adding the CNN layer on the front end of the LSTM layer. Unlike other classification problems, the results indicate that ADHD/ASD classification model required the time step value as 15. Further, employing the parallel two pipelines architecture and in this manner interpreting the features across time steps and then fusing them, increase the performance.

Table 5.8 : The classification results for ADHD/ASD on the original dataset.

	fixation			fixation+saccade			fixation_layer+saccade_layer		
Classifier/time step	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
LSTM_5ts	50.00	50.00	50.00	50.00	50.00	50.00	41.67	41.67	41.67
CNN_LSTM_5ts	50.00	58.33	54.17	50.00	66.67	58.33	50.00	50.00	50.00
CNN_maxpool_LSTM_5ts	58.33	50.00	54.17	66.67	66.67	66.67	58.33	58.33	58.33
LSTM_10ts	58.33	41.67	50.00	66.67	58.33	62.50	41.67	50.00	45.83
CNN_LSTM_10ts	50.00	41.67	45.83	50.00	66.67	58.33	58.33	50.00	54.17
CNN_maxpool_LSTM_10ts	58.33	50.00	54.17	58.33	58.33	58.33	58.33	41.67	50.00
LSTM_15ts	41.67	41.67	41.67	41.67	50.00	45.83	50.00	50.00	50.00
CNN_LSTM_15ts	41.67	41.67	41.67	75.00	50.00	62.50	58.33	58.33	58.33
CNN_maxpool_LSTM_15ts	50.00	50.00	50.00	58.33	50.00	54.17	**58.33**	**66.67**	**62.50**

Table 5.9 : The classification results for ADHD/ASD on the noisy dataset.

	fixation			fixation+saccade			fixation_layer+saccade_layer		
Classifier/time step	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.	Acc.
LSTM_5ts	41.67	41.67	41.67	41.67	50.00	45.83	58.33	50.00	54.17
CNN_LSTM_5ts	50.00	66.67	58.33	58.33	50.00	54.17	58.33	33.33	45.83
CNN_maxpool_LSTM_5ts	58.33	58.33	58.33	50.00	58.33	54.17	58.33	41.67	50.00
LSTM_10ts	50.00	50.00	50.00	41.67	33.33	37.50	41.67	33.33	37.50
CNN_LSTM_10ts	50.00	50.00	50.00	58.33	66.67	62.50	50.00	41.67	45.83
CNN_maxpool_LSTM_10ts	50.00	50.00	50.00	41.67	66.67	54.17	50.00	50.00	50.00
LSTM_15ts	58.33	50.00	54.17	50.00	66.67	58.33	33.33	41.67	37.50
CNN_LSTM_15ts	58.33	58.33	58.33	58.33	41.67	50.00	50.00	41.67	45.83
CNN_maxpool_LSTM_15ts	58.33	66.67	62.50	41.67	50.00	45.83	**58.33**	**58.33**	**58.33**

6. CONCLUSIONS AND RECOMMENDATIONS

Children with ADHD and ASD require many hours of individual education while the government reimbursed education hours are only about eight hours per week in Turkey (http://www.memurlar.net/haber/508184/). Use of computer-based education may provide more education hours for these children at a much smaller cost. In order to provide individualized education it is necessary to collect and analyse some information about these people. Therefore, we prepare an experimental environment and analysis tool to collect application log data and sensor data from these type of children. Our computer-based tool can be used as a data collection and analysis tool to help the experts.

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 the first part of the thesis, we have presented the results of the ANOVA statistical analysis method and three classification algorithms.

The experiments have indicated that the usage of APL and fixation data are 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 have presented the statistical significance analysis of different factors, such as pupil diameter, RC and RT. We have used various machine learning techniques, such as outlier removal using the Tomek links, feature relevance and Random forest, Logistic Regression, SVM classifiers. We have achieved the best accuracy results by using the Random forest algorithm. Also, having balanced value for specificity and sensitivity is important. Therefore, distributions of specificity and sensitivity values (see Table 4.2, Table 4.3 , and Table 4.4) have become balanced especially for ASD/Control and ADHD/Control classification problems.

As it is known, Random forest algorithm performs classification process by using the best and most necessary features of the data. Therefore, applying feature selection methods has not significantly affect the classification performance. However, the usage of Tomek link removing method has improved the classification results obtained with ET_log data. In contrast, Tomek link removing method has no effect on the results obtained with raw fixation data. At this point in order to discover exact impact of the raw fixation data, we decide to go a step further and take the advantage of deep learning techniques in the classification of the participant groups.

In the second part of the thesis, we propose a deep learning based framework for the classification of ASD, ADHD and the control groups. To the best of our knowledge, this is the first study that uses eye movement data with LSTM and CNN algorithms to classify individuals with ASD and individuals with ADHD. During the emotion recognition experiment, our model learns the eye gaze behavior of the participants that occurs on the experimental setup and thus it can distinguish the groups. We have used CNN and LSTM together in our proposed models, CNN is used as the first layer since it achieves to extract essential features of the data. Then LSTM takes the selected relevant features to classify the participant groups. The highest performance results for ASD/Control and ADHD/ASD classification problems are obtained when CNN and LSTM are used successively.

In addition to fixation eye movement, saccade eye movement data has been used in the second part of the thesis. In this way the amount of features in the dataset has been increased. We introduce three model, the first one uses only fixation eye gaze data. For the second one, we merge fixation and saccade features of the participants and use them as input to the CNN-LSTM layers. The last one includes parallel two layers, one layer takes the fixation features as inputs and the other layer takes the saccade features as inputs. According to the results, small sized time step usage provides more successful results for ASD/Control and ADHD/Control classification problems. A further important implication is combination of the fixation and the saccade features. We apply the combination process in two different ways; in the first one we basically merge fixation and saccade features and serve it to the proposed single pipeline network model; in the second way we feed two parallel pipelines separately with the fixation and saccade features and after model training, we merge the model outcomes. Both methods achieve successful results for all our diagnosis classification problems.

To illustrate the robustness of the proposed methods, a new dataset is created by adding noise to the original dataset. The developed methods are also tested with this dataset. We consider the methods for each classification problem that provide satisfactory results in both datasets as the final classification results. According to the obtained results, for ASD/Control problem CNN_maxpool_LSTM_5ts_fs and CNN_maxpool_LSTM_5ts_fsl methods; for ADHD/Control problem LSTM_5ts_fs method; for ADHD/ASD problem CNN_maxpool_LSTM_15ts_fsl method achieve the best results.

6.1 Discussion

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

ADHD/ASD classification problem is the hardest problem among the others. As it is widely accepted, discrimination of the ASD group from the ADHD group suffers from some limitations since participants with ADHD and participants with ASD have common psychiatric symptoms [2]. The proposed fusion of the layers model produced considerable results for ADHD/ASD classification problem. Thus, by driving to feed the parallel pipelines structures by multimodal datasets, we demonstrate that pretty hard classification problems can be solved. In the future, we plan to use other data types such as clinical data to be able to increase the classification accuracy for the disordered groups.

According to obtained results for ADHD/ASD classification problem, time step value should be 15 for the ADHD/ASD classification while as the time step value five is sufficient for the other classification problems. This has shown that more data is needed to learn the intrinsic features of the ADHD and ASD groups. Although, the lack of available sample size is a limitation, we try to overcome this problem by means of the proposed padding and truncation processes. On the other hand, available data has been sufficient for ASD/Control and ADHD/Control problems.

6.1.1 Contribution of RC and RT

The feature selection algorithm used by Random forest reveal that the feature rank of the RC feature is lower than the raw fixation and RT features. This result is consistent with the fact that the RC generally does not have a statistical difference between the participants except for the fear emotion (see Table 2.2). On the other hand, the fact that the RT value is statistically significant (see Table 2.3) leads to this feature being ahead of other APL features during the feature selection operation. Supporting our conclusion, [36] discovers that though RC has no statistically significant difference between participants with ASD and typically developing participants. Furthermore, they indicate that the participants with ASD are slower than the control group in emotion recognition. Also, in [100] ADHD, ASD and the control groups are analyzed in terms of their ability of facial affect recognition. They achieve similar conclusion with us. According to their results, the ASD group respond less correctly than the control group. Also the ASD group's response latency is more than the control group.

6.1.2 Contribution of the pupil diameter

Through the eye tracker that participants wear during the experiment, the differences between the pupil diameter of the participant groups are observed. The average normalized pupil diameters of the participants with ADHD are less than the participants with ASD and the control group while they are looking at angry emotions ($p<0.0001$). The pupil size of the participants with ASD is smaller than the control group. This result is consistent with findings of [101] in which the authors reached smaller pupil diameters in participants with ASD while they are shown neutral faces, avatars and objects.

6.1.3 Feature ranking

Random forest algorithm ranks the features according to their importance score. As seen in Figure 4.2, when the Random forest algorithm uses the 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 employs fixation features of the ET_log data as the most distinguishing features (see Figure 4.3). 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 have a fair agreement (according to the Fleiss' kappa value, see Section 2.5.1) in their responses to the images within themselves. However, the participants of the control group have a moderate agreement. Moreover, the RC feature could not discriminate the participant groups statistically. Therefore, the RC is not a distinguishing factor among the other features for the participant groups except the fear emotion. But the RT feature indicates a statistically significant difference between the participant groups, thus the RT feature have a higher feature importance rank than RC (see Section 4.2).

The statistical test results for the pupil diameter in Table 2.4 confirms that the pupil diameter of the participant with ADHD and the control group has a statistically distinguishing effect. In [33] they perform classification of control subjects from mild cognitive impairment subjects by using the eye tracker fixation and saccade data. They claim that, although pupil diameter does not increase the classification accuracy, fixation duration improves their performance. In our study, both pupil diameter and fixation duration are the distinguishing features.

6.1.4 Deep learning models vs. feature-based models

When solving a machine learning problem, a specific workflow is followed. Firstly, input data is determined and then the relevant features are extracted from it. Then a model that describes or predicts the object is created. On the other hand, deep learning models skip the manual extracting the features from data; instead, data is fed directly into the deep learning algorithms which learn the features and predict the objects.

In this work, the results of the RF + ET_log obtained by averaging the classification results based on ET_log and raw fixation are higher than the results obtained with deep learning. According to the result of RF + ET_log; if a person who is not classified correctly by one data type is classified correctly according to the other data type and the accuracy obtained is sufficiently high, it can be concluded that the person is classified correctly. However, in this case, high variance problems may occur. Therefore, in this study, we compare the results obtained with only the raw fixation data with the results obtained by deep learning methods.

When we compare the proposed CNN-LSTM models and Random forest feature based model, we conclude that the presented deep learning techniques achieve better classification results than Random forest with/without Tomeklinks algorithm when the raw fixation data is used. This conclusion can be interpreted as deep learning methods have strong capability of extracting hidden features and non-linear relations from data without any human intervention [102]. The hidden features of the eye movement data is captured by the CNN algorithm better than Random forest in this work. In addition to this, especially CNN algorithm has powerful generalization ability and it is robust to noise [103]. Moreover, LSTM succeeded in the eye movement time series data since long term dependencies are considered by LSTM method.

6.2 Future Work and Limitations

Eye tracker mobile glasses are worn like regular glasses; however, the size of the eye tracker does 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. 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 who are above an IQ of 70, and who can 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.

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.

6.3 Impact

The impact of this thesis study is on scientific knowledge and health.

From the scientific knowledge side, determining social interaction and emotion recognition abilities of the children with ASD or ADHD in terms of detailed sensors data is an emerging research area. Any contribution that presents more detailed information to the specialist is very important and may provide a different perspective to the issue. Our application can be used as a data collection and analysis tool to help the experts.

In terms of health, analysis of the data may support the specialist knowledge by statistically. At the end of the analysis the findings may contribute to the specialist in terms of understanding emotion recognition behaviors of the children with ADHD and ASD. Thanks to the established analysis setup, the proposed system could be used with another datasets, prepared algorithms and programs may be reused for other studies.

REFERENCES

[1] **Kanade, T.**, **Tian, Y. and Cohn, J.F.** (2000). Comprehensive database for facial expression analysis, *fg*, IEEE, p. 46.

[2] **Frances, A.**, **Pincus, H.A. and First, M.** (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*, American Psychiatric Association Washington DC.

[3] **Polanczyk, G.**, **De Lima, M.S.**, **Horta, B.L.**, **Biederman, J. and Rohde, L.A.** (2007). The worldwide prevalence of ADHD: a systematic review and metaregression analysis, *American journal of psychiatry*, *164*(6), 942–948.

[4] **Roter, D.L.**, **Frankel, R.M.**, **Hall, J.A. and Sluyter, D.** (2006). The expression of emotion through nonverbal behavior in medical visits, *Journal of general internal medicine*, *21*(1), 28–34.

[5] **Kleitman, N. and Ramsaroop, A.** (1948). Periodicity in body temperature and heart rate, *Endocrinology*, *43*(1), 1–20.

[6] **Montepare, J.M.**, **Goldstein, S.B. and Clausen, A.** (1987). The identification of emotions from gait information, *Journal of Nonverbal Behavior*, *11*(1), 33–42.

[7] **Radwan, A.M.**, **Birkan, B.**, **Hania, F. and Cataltepe, Z.** (2017). Active machine learning framework for teaching object recognition skills to children with autism, *International Journal of Developmental Disabilities*, *63*(3), 158–169.

[8] **Kliemann, D.**, **Dziobek, I.**, **Hatri, A.**, **Steimke, R. and Heekeren, H.R.** (2010). Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders, *Journal of Neuroscience*, *30*(37), 12281–12287.

[9] **de Wit, T.C.**, **Falck-Ytter, T. and von Hofsten, C.** (2008). Young children with autism spectrum disorder look differently at positive versus negative emotional faces, *Research in Autism Spectrum Disorders*, *2*(4), 651–659.

[10] **Gamer, M. and Büchel, C.** (2009). Amygdala activation predicts gaze toward fearful eyes, *Journal of Neuroscience*, *29*(28), 9123–9126.

[11] **Fink, E.**, **de Rosnay, M.**, **Wierda, M.**, **Koot, H.M. and Begeer, S.** (2014). Brief report: Accuracy and response time for the recognition of facial emotions in a large sample of children with autism spectrum disorders, *Journal of autism and developmental disorders*, *44*(9), 2363–2368.

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

[13] **Bal, E.** (2012). Emotion Recognition and Social Behaviors in Children with Attention-Deficit /Hyperactivity Disorder. University of Illinois at Chicago, *Ph.D. thesis*.

[14] **Bölte, S., Feineis-Matthews, S., Leber, S., Dierks, T., Hubl, D. and Poustka, F.** (2002). The development and evaluation of a computer-based program to test and to teach the recognition of facial affect, *International Journal of Circumpolar Health*, *61*(sup2), 61–68.

[15] **Cabielles-Hernández, D., Pérez-Pérez, J.R., Paule-Ruiz, M., Álvarez-García, V.M. and Fernández-Fernández, S.** (2014). dmTEA: Mobile learning to aid in the diagnosis of autism spectrum disorders, *European Conference on Technology Enhanced Learning*, Springer, pp.29–41.

[16] **Iannaccone, R., Hauser, T.U., Ball, J., Brandeis, D., Walitza, S. and Brem, S.** (2015). Classifying adolescent attention-deficit/hyperactivity disorder (ADHD) based on functional and structural imaging, *European child & adolescent psychiatry*, *24*(10), 1279–1289.

[17] **Zhou, F., Qu, X., Helander, M.G. and Jiao, J.R.** (2011). Affect prediction from physiological measures via visual stimuli, *International Journal of Human-Computer Studies*, *69*(12), 801–819.

[18] **Liu, C., Conn, K., Sarkar, N. and Stone, W.** (2008). Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder, *International journal of human-computer studies*, *66*(9), 662–677.

[19] **Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M. and Castiglioni, I.** (2015). Use of machine learning to identify children with autism and their motor abnormalities, *Journal of autism and developmental disorders*, *45*(7), 2146–2156.

[20] **Wall, D., Kosmicki, J., Deluca, T., Harstad, E. and Fusaro, V.** (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism, *Translational psychiatry*, *2*(4), e100.

[21] **Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., ... and Rutter, M.** (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*(3), 205–223.

[22] **Crawford, T.J., Higham, S., Renvoize, T., Patel, J., Dale, M., Suriya, A. and Tetley, S.** (2005). Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease, *Biological psychiatry*, *57*(9), 1052–1060.

[23] **Noris, B.**, **Benmachiche, K.**, **Meynet, J.**, **Thiran, J.P. and Billard, A.G.**, (2007). Analysis of head-mounted wireless camera videos for early diagnosis of autism, Computer Recognition Systems 2, Springer, pp.663–670.

[24] **Vidal, M.**, **Turner, J.**, **Bulling, A. and Gellersen, H.** (2012). Wearable eye tracking for mental health monitoring, *Computer Communications*, *35*(11), 1306–1311.

[25] **Ting, W.K.C.**, **Perez Velazquez, J.L. and Cusimano, M.D.** (2014). Eye movement measurement in diagnostic assessment of disorders of consciousness, *Frontiers in neurology*, *5*, 137.

[26] **Bayram, S. and Esgin, E.** (2015). Patterns of visual attention and gaze to human and animal faces in children with Autism Spectrum Disorders, *World Journal on Educational Technology*, *4*(3), 165–179.

[27] **Kätsyri, J. and Sams, M.** (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces, *International Journal of Human-Computer Studies*, *66*(4), 233–242.

[28] **Gökler, B.**, **Ünal, F.**, **Pehlivantürk, B.**, **Kültür, E.Ç.**, **Akdemir, D. and Taner, Y.** (2004). Okul Çağı Çocukları için Duygulanım Bozuklukları ve Şizofreni Görüşme Çizelgesi-Şimdi ve Yaşam Boyu Şekli- Türkçe Uyarlamasının Geçerlik ve Güvenirliği, *Çocuk ve Gençlik Ruh Sağliği Dergisi*.

[29] **Kaufman, J.**, **Birmaher, B.**, **Brent, D.**, **Rao, U.**, **Flynn, C.**, **Moreci, P.**, **..... and 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.

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

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

[32] **Scinto, L.F.**, **Daffner, K.R.**, **Castro, L.**, **Weintraub, S.**, **Vavrik, M. and 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.

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

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

[35] **Landis, J.R. and Koch, G.G.** (1977). The measurement of observer agreement for categorical data, *biometrics*, 159–174.

[36] **Bal, E.**, **Harden, E.**, **Lamb, D.**, **Van Hecke, A.V.**, **Denver, J.W. and 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.

[37] **Moore, D.**, **McCabe, G.P. and Craig, B.** (2012). *Introduction to the Practice of Statistics: w/Student CD*, San Francisco, CA: Freeman.

[38] **Cover, T.M. and Thomas, J.A.** (2012). *Elements of information theory*, John Wiley & Sons.

[39] **Shannon, C.E.** (1948). A mathematical theory of communication, *Bell system technical journal*, *27*(3), 379–423.

[40] **Huang, D. and Chow, T.W.** (2005). Effective feature selection scheme using mutual information, *Neurocomputing*, *63*, 325–343.

[41] **Yu, L. and Liu, H.** (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution, *Proceedings of the 20th international conference on machine learning (ICML-03)*, pp.856–863.

[42] **Song, Q.**, **Ni, J. and Wang, G.** (2013). A fast clustering-based feature subset selection algorithm for high-dimensional data, *IEEE transactions on knowledge and data engineering*, *25*(1), 1–14.

[43] **Singh, B.**, **Kushwaha, N. and Vyas, O.P.** (2014). A feature subset selection technique for high dimensional data using symmetric uncertainty, *Journal of Data Analysis and Information Processing*, *2*(04), 95.

[44] **Dash, M. and Liu, H.** (1997). Feature selection for classification, *Intelligent data analysis*, *1*(3), 131–156.

[45] **Kononenko, I.**, **Šimec, E. and Robnik-Šikonja, M.** (1997). Overcoming the myopia of inductive learning algorithms with RELIEFF, *Applied Intelligence*, *7*(1), 39–55.

[46] **Kira, K. and Rendell, L.A.** (1992). The feature selection problem: Traditional methods and a new algorithm, *Aaai*, volume 2, pp.129–134.

[47] **Breiman, L.** (2017). *Classification and regression trees*, Routledge.

[48] **Ho, T.K.** (1995). Random decision forests, *Document analysis and recognition, 1995., proceedings of the third international conference on*, volume 1, IEEE, pp.278–282.

[49] **Cortes, C. and Vapnik, V.** (1995). Support-vector networks, *Machine learning*, *20*(3), 273–297.

[50] **Altman, N.S.** (1992). An introduction to kernel and nearest-neighbor nonparametric regression, *The American Statistician*, *46*(3), 175–185.

[51] **Freund, Y. and Schapire, R.E.** (1997). A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of computer and system sciences*, *55*(1), 119–139.

[52] **Delaherche, E.**, **Chetouani, M.**, **Bigouret, F.**, **Xavier, J.**, **Plaza, M. and 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.

[53] **Ozturk, M.U.**, **Arman, A.R.**, **Bulut, G.C.**, **Findik, O.T.P.**, **Yilmaz, S.S.**, **Genc, H.A.**, **..... and Cataltepe, Z.** (2018). Statistical Analysis and Multimodal Classification on Noisy Eye Tracker and Application Log Data of Children with Autism and ADHD, *Intelligent Automation And Soft Computing*, *24*(4903), 918.

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

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

[56] **Duda, M.**, **Ma, R.**, **Haber, N. and Wall, D.** (2017). Use of machine learning for behavioral distinction of autism and ADHD, *Translational psychiatry*, *6*(2), e732.

[57] **Van Der Geest, J.N.**, **Kemner, C.**, **Verbaten, M.N. and 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), 669–678.

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

[59] **Partala, T. and Surakka, V.** (2003). Pupil size variation as an indication of affective processing, *International journal of human-computer studies*, *59*(1-2), 185–198.

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

[61] **Caffrey, C.** (2008). Using pupillometric, fixation-based and subjective measures to measure the processing effort experienced when viewing subtitled TV anime with pop-up gloss, *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing. Copenhagen: Samfundslitteratur*, 125–144.

[62] **Hvelplund, K.T.** (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.

[63] **Hess, E.H.** (1975). *The tell-tale eye: How your eyes reveal hidden thoughts and emotions.*, Van Nostrand Reinhold.

[64] **Lin, Y.**, **Zhang, W. and 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.

[65] **Lin, Y.**, **Zhang, W.J.**, **Wu, C.**, **Yang, G. and 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.

[66] **Rivera, S.**, **Best, C.**, **Yim, H.**, **Walther, D.**, **Sloutsky, V. and 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.

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

[68] **Colby, J.B.**, **Rudie, J.D.**, **Brown, J.A.**, **Douglas, P.K.**, **Cohen, M.S. and Shehzad, Z.** (2012). Insights into multimodal imaging classification of ADHD, *Frontiers in systems neuroscience*, *6*, 59.

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

[70] **Alpaydin, E.** (2014). *Introduction to machine learning*, MIT press.

[71] **Buhmann, M.D.** (2003). *Radial basis functions: theory and implementations*, volume 12, Cambridge university press.

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

[73] **Elhassan, T.**, **Aljurf, M.**, **Al-Mohanna, F. and 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*, *1*, 1–12.

[74] **Wang, Q.**, **Xin, J.**, **Wu, J. and Zheng, N.** (2017). SVM classification of microaneurysms with imbalanced dataset based on borderline-SMOTE and data cleaning techniques, *Ninth International Conference on Machine Vision (ICMV 2016)*, volume10341, International Society for Optics and Photonics, p.103411S.

[75] **Quinlan, J.R.** (1986). Induction of decision trees, *Machine learning*, *1*(1), 81–106.

[76] **Breiman, L.** (2001). Random forests, *Machine learning*, *45*(1), 5–32.

[77] **Pedregosa, F.**, **Varoquaux, G.**, **Gramfort, A.**, **Michel, V.**, **Thirion, B.**, **Grisel, O.**, **..... and Cournapeau, D.** (2011). Scikit-learn: Machine learning in Python, *Journal of machine learning research*, *12*(Oct), 2825–2830.

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

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

[80] **Pelc, K.**, **Kornreich, C.**, **Foisy, M.L. and Dan, B.** (2006). Recognition of emotional facial expressions in attention-deficit hyperactivity disorder, *Pediatric neurology*, *35*(2), 93–97.

[81] **Yuill, N. and Lyon, J.** (2007). Selective difficulty in recognising facial expressions of emotion in boys with ADHD, *European child & adolescent psychiatry*, *16*(6), 398–404.

[82] **Komogortsev, O.V.**, **Gobert, D.V.**, **Jayarathna, S.**, **Koh, D.H. and Gowda, S.M.** (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors, *IEEE Transactions on Biomedical Engineering*, *57*(11), 2635–2645.

[83] **Brenner, L.A.**, **Turner, K.C. and Müller, R.A.** (2007). Eye movement and visual search: are there elementary abnormalities in autism?, *Journal of autism and developmental disorders*, *37*(7), 1289–1309.

[84] **Dvornek, N.C.**, **Ventola, P.**, **Pelphrey, K.A. and Duncan, J.S.** (2017). Identifying autism from resting-state fMRI using long short-term memory networks, *International Workshop on Machine Learning in Medical Imaging*, Springer, pp.362–370.

[85] **Bone, D.**, **Bishop, S.L.**, **Black, M.P.**, **Goodwin, M.S.**, **Lord, C. and Narayanan, S.S.** (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion, *Journal of Child Psychology and Psychiatry*, *57*(8), 927–937.

[86] **Deshpande, G.**, **Libero, L.**, **Sreenivasan, K.R.**, **Deshpande, H. and Kana, R.K.** (2013). Identification of neural connectivity signatures of autism using machine learning, *Frontiers in human neuroscience*, *7*, 670.

[87] **Kosmicki, J.**, **Sochat, V.**, **Duda, M. and Wall, D.** (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning, *Translational psychiatry*, *5*(2), e514.

[88] **Hochreiter, S. and Schmidhuber, J.** (1997). Long short-term memory, *Neural computation*, *9*(8), 1735–1780.

[89] **Gers, F.A.**, **Schmidhuber, J. and Cummins, F.** (1999). Learning to forget: Continual prediction with LSTM.

[90] **Krizhevsky, A.**, **Sutskever, I. and Hinton, G.E.** (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp.1097–1105.

[91] **Zheng, Y.**, **Liu, Q.**, **Chen, E.**, **Ge, Y. and Zhao, J.L.** (2014). Time series classification using multi-channels deep convolutional neural networks, *International Conference on Web-Age Information Management*, Springer, pp.298–310.

[92] **Qureshi, M.N.I.**, **Oh, J.**, **Min, B.**, **Jo, H.J. and Lee, B.** (2017). Multi-modal, multi-measure, and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain MRI, *Frontiers in human neuroscience*, *11*, 157.

[93] **Choi, E.**, **Bahadori, M.T.**, **Schuetz, A.**, **Stewart, W.F. and Sun, J.** (2016). Doctor ai: Predicting clinical events via recurrent neural networks, *Machine Learning for Healthcare Conference*, pp.301–318.

[94] **Lipton, Z.C.**, **Kale, D.C.**, **Elkan, C. and Wetzel, R.** (2015). Learning to diagnose with LSTM recurrent neural networks, *arXiv preprint arXiv:1511.03677*.

[95] **Lustig, J.**, (2016), Identifying dyslectic gaze pattern: Comparison of methods for identifying dyslectic readers based on eye movement patterns.

[96] **Pei, M.**, **Wu, X.**, **Guo, Y. and Fujita, H.** (2017). Small bowel motility assessment based on fully convolutional networks and long short-term memory, *Knowledge-Based Systems*, *121*, 163–172.

[97] **Lin, X.**, **Zhao, C. and Pan, W.** (2017). Towards accurate binary convolutional neural network, *Advances in Neural Information Processing Systems*, pp.345–353.

[98] **Mishra, A.**, **Dey, K. and Bhattacharyya, P.** (2017). Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pp.377–387.

[99] **Wang, J.**, **Yu, L.C.**, **Lai, K.R. and Zhang, X.** (2016). Dimensional sentiment analysis using a regional CNN-LSTM model, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pp.225–230.

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

[101] **Martineau, J.**, **Hernandez, N.**, **Hiebel, L.**, **Roché, L.**, **Metzger, A. and 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.

[102] **Bengio, Y.**, **Courville, A. and Vincent, P.** (2013). Representation learning: A review and new perspectives, *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798–1828.

[103] **LeCun, Y.**, **Bengio, Y. and Hinton, G.** (2015). Deep learning, *nature*, *521*(7553), 436.

CURRICULUM VITAE

Name Surname: Mahiye Uluyağmur Öztürk

Place and Date of Birth: Kayseri, 1987

E-Mail: muluyagmur@itu.edu.tr

EDUCATION:

- **B.Sc.:** 2009, Erciyes University, Computer Engineering Department
- **M.Sc.:** 2012, Istanbul Technical University, Computer Engineering Department

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2012-2018 Research and Teaching Assistant at Istanbul Technical University, Faculty of Computer and Informatics Engineering.
- 2013-2014 Technology Leaders Graduate Scholarship from Turkcell.
- 2011-2012 Researcher at San-Tez Project 00966.STZ.2011-2.
- 2010-2011 Researcher at Tubitak Project.
- 2009-2010 Student assistant at Istanbul Technical University.

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Ozturk, M. U.**, Arman, A. R., Bulut, G. C., Findik O. T. P., Yilmaz, S. S., Genc H. A., Yazgan, M. Y., Teker, U., Cataltepe Z., 2018. Statistical Analysis and Multimodal Classification on Noisy Eye Tracker and Application Log Data of Children with Autism and ADHD, *Intelligent Automation And Soft Computing,* Vol.24, 4903-918.
- **Ozturk, M. U.**, Arman, A. R., Bulut, G. C., Findik O. T. P., Yilmaz, S. S., Genc H. A., Yazgan, M. Y., Teker, U., Cataltepe Z., 2016. ADHD and ASD Classification Based on Emotion Recognition Data. *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference*, 2016, USA.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Garaylı, L., **Öztürk, M. U.**, Çataltepe, Z. (2017, May). Performing video alignment by using different methods in human computer interaction experiments. In Signal Processing and Communications Applications Conference (SIU), 2017 25th (pp. 1-4). IEEE.

- Garayli, L., **Uluyagmur-Ozturk**, M., Cataltepe, Z. (2016, May). Alignment of eye tracker and camera data for human computer interaction experiments. In Signal Processing and Communication Application Conference (SIU), 2016 24th (pp. 2009-2012). IEEE.

- **M. Uluyagmur**, Z. Cataltepe, E. Tayfur, Method for facilitating personalized movie recommendation for user, Patent No: EP2763421A1, 06 Aug 2014

- Cataltepe, Z., **Uluyagmur, M.**, and Tayfur, E. (2013). Feature Selection for Movie Recommendation. Turk J Elec Eng and Comp Sci, 24, (2016), 833-848.

- Cataltepe, Z., **Uluyagmur, M.**, and Tayfur, E. (2012, April). TV program recommendation using implicit feedback with adaptive regularization. In Signal Processing and Communications Applications Conference (SIU), 2012 20th (pp. 1-4). IEEE.

- **Uluyagmur, M.**, Cataltepe, Z., and Tayfur, E. (2012). Content-based movie recommendation using different feature sets. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 17-24).