ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

MACHINE LEARNING-ENABLED STRESS DETECTION IN CHILDREN USING PHYSIOLOGICAL SIGNALS DURING ROBOT ASSISTED THERAPY

M.Sc. THESIS

Sevgi Nur BİLGİN AKTAŞ

Department of Computer Engineering

Computer Engineering Programme

FEBRUARY 2024

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

MACHINE LEARNING-ENABLED STRESS DETECTION IN CHILDREN USING PHYSIOLOGICAL SIGNALS DURING ROBOT ASSISTED THERAPY

M.Sc. THESIS

Sevgi Nur BİLGİN AKTAŞ
(504201538)

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Hatice KÖSE

FEBRUARY 2024

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ÇOCUKLARDA MAKİNE ÖĞRENMESİ İLE DESTEKLENMİŞ ROBOT İLE YAPILAN TERAPİ SIRASINDA FİZYOLOJİK SİNYALLERLE STRES TESPİTİ

YÜKSEK LİSANS TEZİ

Sevgi Nur BİLGİN AKTAŞ
(504201538)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Hatice KÖSE

ŞUBAT 2024

Sevgi Nur BİLGİN AKTAŞ, a M.Sc. student of ITU Graduate School student ID 504201538 successfully defended the thesis entitled "MACHINE LEARNING-ENABLED STRESS DETECTION IN CHILDREN USING PHYSIOLOGICAL SIGNALS DURING ROBOT ASSISTED THERAPY", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Hatice KÖSE** Istanbul Technical University	....
Jury Members :	**Prof. Dr. Duygun EROL BARKANA** Yeditepe University	....
	Assoc. Prof. Yusuf YASLAN Istanbul Technical University	....

Date of Submission : 3 January 2024
Date of Defense : 22 February 2024

To my family,

FOREWORD

First of all, I would like to start by expressing my sincerest thanks to Prof. Dr. Hatice KÖSE. Her inclusion of me in her research group and her unwavering support and guidance throughout my educational journey were invaluable. Her mentorship and insightful guidance contributed significantly to the success of this project.

The friendship and collaborative spirit in the ITU Cognitive and Social Robotics team has been an enriching experience. I am deeply grateful for the opportunity to work with such an inspiring group.

I extend my heartfelt thanks to Prof. Dr. Duygun Erol BARKANA for invaluable help and guidance during the Emboa project. Additionally, I am grateful to Elif TOPRAK, Buket COŞKUN, and Pınar ULUER for their enjoyable contributions to our collaborations.

Also, I would like to express my deep thanks to Semra MİRMAHMUTOĞULLARI, Head of the Human-Machine Interaction Software Technologies Department, and all my colleagues for their support and encouragement.

Finally, I would like to express my deepest gratitude to my parents, BİLGİN - AKTAŞ family, and my dear husband Cihat AKTAŞ, whose support was the cornerstone of my educational journey.

February 2024 Sevgi Nur BİLGİN AKTAŞ

TABLE OF CONTENTS

ABBREVIATIONS

ASD : Autism Spectrum Disorder
ANS : Autonomic Nervous System
BVP : Blood Volume Pulse
CNN : Convolutional Neural Network
ECG : Electrocardiogram
EDA : Electrodermal Activity
HF : High Frequency
HRV : Heart Rate Variability
LF : Low Frequency
LSTM : Long Short Term Memory
RF : Random Forest
ROC : Receiver Operating Characteristic
SC : Skin Conductance
SHAP : SHapley Additive ExPlanations
ST : Skin Temperature
SVM : Support Vector Machine

LIST OF TABLES

LIST OF FIGURES

Page

MACHINE LEARNING-ENABLED STRESS DETECTION IN CHILDREN USING PHYSIOLOGICAL SIGNALS DURING ROBOT ASSISTED THERAPY

SUMMARY

Human is a system that constantly interacts with the environment involving in different behaviours and emotional states. Some of the emotional states that are an important part of human life can negatively affect a person's health, such as stress. Stress reduces people's quality of daily life and causes health problems if it persists. Understanding stress and emotional state is even more challenging for some people such as people with Autism spectrum disorder (ASD).Individuals with ASD, especially children, have difficulty expressing their emotions with facial gestures therefore facial emotion recognition systems do not work as desired in children with autism. For this reason, physiological components such as Electrodermal Activity (EDA), Blood Volume Pulse (BVP), and Skin Temperature (ST), can be used for emotion and stress detection., and can be an alternative for the facial systems.

This thesis is a part of the EMBOA Erasmus+ project where the aim is to detect the emotion and stress of children from different modalities during robot-assisted autism therapy sessions. Various studies were carried out with children within the scope of this thesis. Our first study was conducted with data recorded during robot therapy with children with ASD within the scope of the EMBOA project. The major challenges in this workare the absence of a similar dataset with children with ASD and the challenge in labelling all of the data collected during the clinical studies. For this reason, other known and labelled databases from similar domains are used in the analysis as well. Signal analysis and filtering were provided and feature extractions were made. The changes of children in therapy sessions with different robots were examined. In addition, the relationship between physiological data and age groups for children in different age groups in different countries was examined. Analyzes were performed using ANOVA and T-test statistical analysis methods. Finally, in the test conducted with typically developing primary school students, the setup was created using the Stroop test. Thanks to this setup, the data is labeled. The obtained data were bandpass filtered and EDA peaks were calculated. Afterwards, approximately 61 features were extracted from the physiological data obtained. The number of features was reduced to 10 using the MrMR feature selection algorithm. Afterwards, the XGBoost classification algorithm was used and classifications were made according to the situations separated in the test. The trained model was analyzed using the SHAP algorithm. It was determined which features were more effective on the model. Finally, the model was trained using the LSTM algorithm, and the outputs were shared.

ÇOCUKLARDA MAKİNE ÖĞRENMESİ İLE DESTEKLENMİŞ ROBOT İLE YAPILAN TERAPİ SIRASINDA FİZYOLOJİK SİNYALLERLE STRES TESPİTİ

ÖZET

İnsan sürekli çevreyle etkileşim halindedir, içinde bulunduğu ortam ve olaylar kişiyi farklı duygu durumlarına dahil edebilmektedir. Günlük yaşamın önemli bir bileşeni olan duygu durumlarından bazıları ise insanın sağlığını kötü yönde etkileyebilmektedir. Bu duyguların başında stres gelmektedir. Stres kişilerin günlük yaşam kalitesini düşürmektedir, sürekli hale geldiği durumlarda ise mental ve fiziksel sorunlara sebep olmaktadır. Bu sebeple stres ve duygu tespiti amacıyla literatürde laboratuvar ve doğal yaşam ortamında farklı yöntemlerle çalışmalar yapılmıştır. Çalışmalarda genel olarak giyilebilir teknolojilerden yardım alınmıştır, testlerde ise farklı duygu hissiyatı tetiklenebilecek ortamlar oluşturulmuştur. Test ortamları oluşturulmasında görsel ve işitsel bileşenler kullanılmıştır. Literatürdeki stres ve duygu tespiti çalışmaları çoğunlukla yetişkin bireylerde yapılmıştır, çocuklarla ilgili yayınlanan bir veriseti bulunmamaktadır. WESAD (Wearable Stress and Affect Detection), DEAP gibi yetişkinlerin verilerinin bulunduğu datasetler bulunmaktadır. Bu datasetlerde yetişkinlere bazı stres duygusunu aktifleştirecek etkinlikler yaptırılıp o anki verileri stresli olarak işaretlenmektedir. Bu stress etkinlikleri, toplu alanda sunum yapma mental aritmetik vb. etkinliklerdir. Genel olarak duyguları tetiklemek için video klip ve görseller kullanılmaktadır.

Stres - duygu tespiti çalışmaları sağlık, pazarlama, reklamcılık, insan makine-robot etkileşimi gibi konularda sıklıkla kullanılmaktadır. Robotların insanlarla etkileşime girmesi sebebiyle, sosyal davranış sergileyen robotlar için duygu tespitini giderek daha önemli hale getirmektedir. Otizm Spektrum Disorder (OSD) sendromuna sahip çocukların duygu aktarım metabolizması tipik gelişim gösteren çocuklara göre farklıdır. OSD'li bireyler özellikle çocuklar, duygularını ifade etmekte zorlanırlar. Bazı görsel duygu algılama sistemleri yetişkinlerde yüksek doğruluk oranlarında çalışırken bu sistemler otizmli çocuklarda beklenilen doğrulukta çalışamazlar. Bunun sebebi otizmli çocukların duygularını tipik gelişimli çocuklar gibi ifade edememesidir. Bu sebeple görsel sistemlerdense stres ve duygu durumlarının otomatik olarak etkilediği fizyolojik bileşenlerin stres analizinde etkin rol oynadığı belirtilmektedir. Çevre koşullara uyum sağlanması ve günlük yaşamına devam edebilmesi amacıyla OSD'li bireylerle Robotlu terapi seansları gerçekleştirilmektedir. Bu terapi sistemlerinin amaçları OSD'li çocukları doğal yaşam aşamalarına alıştırmak ve günlük hayatta kendilerini daha iyi ifade edebilmelerini sağlamaktadır. Tasarlanan sosyal robotlarla birlikte insanlarla iletişim gerçekleştiren robotların, sosyal farkındalık göstermesi ve karşısındaki kişisin psikolojik durumuna göre hareketlerini şekillendirebilmesi gerekmektedir. Bu amaç doğrultusunda fizyolojik verilerden duygu tahimini önemli bir konuma sahiptir.

“Otizmli çocuklar için bir müdahale aracı olarak sosyal yardımcı robotlarda duygusal döngü (EMBOA)” isimli Erasmus projesi araştırma projesi tez konularından biridir. Çalışmanın amacı OSD’li çocuklarla rehabilitasyon sürecinde robotları kullanarak çalışmalar gerçekleştirmektir. EMBOA projesi kapsamında OSD’li çocuklarda yapılan robotla terapi seansları sırasında, giyilebilir akıllı saatler yardımıyla fizyolojik veriler kayıt altına alınmıştır. Empatica E4 bilekliği kullanılmıştır, bileklik yardımıyla birçok fizyolojik veri kaydı alınmıştır. Sürekli değişmekte olan çevre koşullarına uyum sağlamak için hücreler Otonom Sinir Sistemi tarafından kontrol altında olan biyolojik bazı geri bildirimler kullanır. Otonom Sinir Sistemi vücut içindeki homeostazı (dengeyi) korumak için hücrelerin tepkilerini düzenler. Bu fizyolojik veriler Elektrodermal Aktivite (EDA), Kan Hacmi Nabzı (BVP), Deri Sıcaklığı (ST) gibi bileşenlerdir. EDA deri yüzeyine temas eden özel sensörlerle ölçülmektedir stres sonucu ortaya çıkan ter bezleri aktivitesi elektriksel iletkenliği arttırır ve ilgili bileşende bir değişim meydana getirir. BVP, sensörü kızılötesi ışıkla geçen kan hacmini ölçerek kalp atışlarını tespit edilerek elde edilmektedir. BVP’den kalp atış hızı ve kalp atış hızı değişkenliğini (HRV) hesaplanabilmektedir. Empatica E4 bilekliği içerisinde tüm bu fizyolojik verileri kaydetmek için çeşitli sensörler içermektedir. Bu sensörlerin farklı örnekleme frekansı bulunmaktadır. EDA ve ST 4 Hz, BVP 64 Hz örnekleme frekansına sahiptir. Çalışma uluslararası bir kapsamdadır. 4 ülke katılım sağlamıştır. Bu ülkeler Polonya, Makedonya, Ingiltere ve Türkiye’dir. Toplamda 65’ten fazla OSD’li çocuktan kayıtlar alınmıştır. Çalışmalar belirli zaman aralığından birden fazla oturumla gerçekl eştirilmiştir. Sosyal robot olarak Kaspar kullanılmıştır.

Kaspar 55 cm yüksekliğinde ve 45 cm genişliğinde olan insansı bir robottur. Çoğunlukla çocuklar olmak üzere insanlarla oturur pozisyonda etkileşime girmektedir. Çevresini algılamak ve etkileşime geçmek için gömülü 2D kameralar, dokunmatik sensörler ve mikrofonlar bulunmaktadır. Silikon yüze sahip olan robot, basit ifade özelliklerine sahip ve farklı duygu durumlarını gösterebilmektedir. Otizmli çocuklar için robot destekli terapi ve sosyal robotik çalışmaları için kullanıldı. Kaspar bir sosyal aracı olarak kullanılabilmekte ve çocukların çevreleriyle daha iyi etkileşim ve iletişim kurmasına, temel duyguları keşfetmesine ve tanımasına, sosyal olarak kabul edilebilir dokunsal etkileşimi öğrenmesine ve taklit, işbirliği becerileri gibi temel sosyal becerileri öğrenmesine yardımcı olmaktadır. Boyut olarak da çocuk boyutlarına yakın olduğundan dolayı çocuklar tarafından kolaylıkla oyun arkadaşı olarak kabul edilebilmektedir.

EMBOA projesi kapsamında çocuklarla çeşitli çalışmalar yürütülmüştür. İlk çalışmamız EMBOA projesi ile OSD’li çocuklarla robotla terapi süresince kayıt altına alınan verilerle yapılmıştır. Yetişkin verilerinde duygu değişimleri kişilere farklı görevler verilerek etiketleme ve anketler ışığında yapılmaktadır, bu durum OSD’li çocuklarda yapılamadığı için Erasmus projesiyle katılım sağlayan okulların topladığı verilerde etiket bulunmamaktadır. Bu sebeple ilk aşama olarak bu verilerle literatürdeki araştırmalar sonucu elde edilen veriler ile değerlendirmeler yapılmıştır. Sinyal analizleri ve filtrelemeleri sağlanmış ve özellik çıkarımları yapılmıştır. Tüm verilerden toplamda 63 adet özellik çıkarımı yapıştır. Yapılan çalışmalardaki stres belirten özellikler araştırılmış ve EDA’ya bağlı olarak çıkartılan EDA peak sayısının duygu durum değişiminde önemli belirteçlerden olduğu tespit edilmiştir. Bu bilgiler ışığında çocukların farklı robotlarla terapi seanslarındaki değişimleri incelenmiştir. Ayrıca farklı ülkelerdeki farklı yaş gruplarındaki çocuklar için fizyolojik verileriyle yaş gruplarının ilişkisi incelenmiştir. ANOVA ve T-test istatistiksel analiz yöntemi kullanarak analizler yapılmıştır. Grupların birbirinden ayırt edici olan özellikleri belirlenmiştir. Özellik seçim çalışmasında fizyolojik sinyallerden çıkarılan özellikler (2 dakikalık bölümler) kullanıldı. Toplanan ham sinyaller, özellik çıkarma adımından önce filtrelendi ve normalleştirildi. Her segment için BVP için 7680 x 1 ve EDA ve ST için 480 x 1 vektörleri elde edildi. BVP verileri, 18 dB durdurma bandı zayıflamasına (Rs) ve 0,1 Hz normalleştirilmiş durdurma bandı kenar frekansına (Wn) sahip altıncı dereceden bir Chebyshev II filtresi kullanılarak önceden işlendi. Ayrıca filtrelenen BVP sinyali 0-100 arasında normalleştirildi. EDA sinyalini filtrelemek için çerçeve uzunluğu değeri 11 olan beşinci dereceden Savitzky-Golay filtresi kullanıldı. BVP'ye benzer şekilde EDA sinyali de 0-100 arasında normalleştirildi. Filtrelenen ve normalize edilen sinyal, fazik (cilt iletkenlik tepkisi (SCR)) ve tonik bileşenlere (cilt iletkenlik düzeyi (SCL)) ayrıştırıldı. ST sinyalinde herhangi bir gürültü veya aykırı değer gözlemlenmediğinden, özellik çıkarma işlemi için doğrudan ham ST sinyali kullanıldı.

İkinci çalışmada ise, tipik gelişimli ilkokul öğrencileriyle Stroop test kullanılarak setup oluşturulmuştur. Stroop test, zihinsel aktiflik ve dikkat yeteneğini ölçen bir psikolojik testtir. Stroop test çoğu makalede kullanılan bir fizyolojik duygu uyarımı olan bir oyundur. Katılımcılara kısa sürede renk ve yazılar arasında çelişki olan oyunu bitirmeleri görev olarak verilmiştir. Katılımcıya ekranda farklı renklerde renk isimleri belirtilerek doğru renklerin söylenmesi istenir. Renk ve yazılar farklı olduğu için zihinde çatışmaya sebep olmaktadır. Katılımcılar ilkokul öğrencileri olduğu için Stroop test uygulanabilmiştir. Her katılımcı için aynı süre sürecek olan Stroop test oyunu oluşturulmuştur, bu setup sayesinde veriler etiketlenmiştir. Elde edilen veriler bandpass filtreleme yapılıp EDA peakleri hesaplanmıştır. Sonrasında ise elde edilen fizyolojik verilerden yaklaşık 61 adet özellik çıkartılmıştır. Bu özelliklerden bazılarının ilişkileri konvolüsyon matrisleri ile incelenmiştir. MrMR özellik seçim algoritması kullanılarak özellik sayısı 10'a düşürülmüştür. Böylelikle birbiriyle ilişkili olan özellik boyutları elimine edilmiştir. Sınıflandırma yöntemleri araştırılmış ve literatürde en çok kullanılan sınıflandırma yöntemleriyle çalışmalar yapılmıştır çalışmada XGBoost ve LSTM kullanılmıştır. XGBoost sınıflandırma algoritması kullanılmış ve analiz edilmiştir Sınıflandırma metrikleri MrMR feature selection sonrası seçilen 10 özellik ile yapılmıştır. Test doğruluğunun %88 olduğu gözlemlenmiştir. Sınıflandırma yapılan özellik sayısı 61 olduğunda ise stest doğruluğu %96 olarak görülmüştür. Eğitim sonucunda SHAP (SHapley Additive ExPlanations) algoritması kullanarak eğitim yapılan modelin analizi yapılmıştır. SHAP sayesinde model üzerinde hangi özelliklerin model üzerinde daha etkili olduğu görselleştirilmiştir. Bu görseller üzerinden analiz ve sonuç çıkarımı yapılmıştır. Model üzerinde en etkin olan değerin EDA_max_min olduğu gözlemlenmiştir. Son olarak LSTM algoritması kullanılarak model eğitimi yapılmış ve çıktıları paylaşılmıştır. LSTM ile Stroop test kayıtlarından elde edilen özellik çıkartma yapılmamış zaman serisi verileri üzerinden işlem yapılmıştır. Oluşturulan modelin binary-class sınıflandırma sonuçlarının multi-class sınıflandırmaya göre daha iyi olduğu gözlemlenmiştir. Multi-class sınıflandırmaya göre baseline durumunun ayırt ediciliği diğer sınıflara göre daha düşüktür. WESAD datası kullanılarak LSTM eğitimi yapılıp çocuk katılımcılardan aldığımız verilerle transfer learning yapılmıştır, fakat kayda değer bir sonuç alınamamıştır.

Özet olarak, ilgili tezde çocuklarla stres tespiti amacıyla farklı yaklaşımlar gerçekleştirilmiştir. Amaç OSD'li çocuklarda yapılan terapilerde fizyolojik verilerden hareketle duygu tahmini yapıp robotun çevreye olan duyarlılığını daha arttırmaktadır. Çalışmadaki kısıtlamalar, otizmli çocuklarla yetişkinlerle yapılan stres görevlerinin gerçekleştirilememesi ve test sırasında takılan giyilebilir teknoloji aletlerini takmak istememeleri, test sırasında çıkarmaya çalışmaları ve benzeri durumlardır. Gelecek çalışmalarda elde edilen veriseti boyutu büyütülüp ikili sinyaller ile eğitimler yapılıp karşılaştırmalar yapılabilir. Deep learning algoritmalarıyla genişletilmiş veri setleriyle ek çalışmalar yapılması doğruluk oranı arttırılabilir.

1. INTRODUCTION

In order to survive, people must adapt to changes in the environment. The system that ensures this balance is called homeostasis. Homeostasis regulates parameters such as oxygen supply, body temperature, and heart rate to keep people alive [1]. Hormones regulate the Autonomic Nervous Systems (ANS) that maintain balance in our body and enable us to adapt to the environment. Some physiological hormones and data respond to stress encountered in daily life. The human body is protected by physiological mediators including cortisol, adrenaline, and noradrenaline, as well as modifications in immunological and metabolic characteristics, provided that these processes occur in balance [2].

Several factors affect physiological and mental health, the most indirect of which is stress. Long-term stress in people causes multiple physiological and psychological problems [3]. Stress is the major reason for many diseases that affect both physiological and mental health issues. People's long-term stress is associated with various physiologic and psychological complications. Stress is an inevitable component of contemporary living, and long-term stress can negatively affect physical and psychological health. Emotional recognition skills help understand individuals' emotional states. It helps to manage difficult circumstances better and achieve emotional regulation.

The neurodevelopmental disorder known as Autism Spectrum Disorder (ASD) is characterized by restricted and recurring activities as well as difficulties with social communication and engagement [4]. It is a comprehensive disorder with a broad range of symptoms, including limited interests repetitive behaviors, and delays or deficits in language development, social interaction, and communication. People who have ASD frequently struggle to recognize and communicate their feelings as well as manage and control their emotional reactions [5,6]. Since individuals with ASD have difficulty representing their emotions, emotion and stress recognition applications are useful for therapy sessions.

Emotion recognition-interpretation through physiological signals has emerged as an important research field that has attracted significant attention in various interdisciplinary fields such as psychology, healthcare, and human-robot interaction.

Various studies are being carried out in this field from different perspectives. discussed the use of wireless sensors for emotion recognition from physiological signals and emphasized the importance of existing technologies in this context [7]. Presented a comprehensive review of emotion recognition using physiological signals and highlighted the widespread use of Support Vector Machines (SVM) in this field [8]. Focused on emotion recognition based on physiological changes during music listening and shed light on the limitations associated with using physiological signals for emotion recognition [9]. Collectively, these references underscore the importance of physiological signals and the challenges associated with using them in emotion recognition.

Song et al., 2019 also highlighted the performance of Long Short-Term Memory (LSTM) in emotion recognition using physiological signals, while acknowledging the limitations of certain sequences in contributing to effective classification models [10]. Emphasized the reliability of physiological signals for emotion recognition because they cannot be deliberately controlled by the subject, thus underlining the robustness of physiological data in inferring emotions [11]. Additionally, Wei et al., 2018 proposed a decision-level weight fusion strategy for emotion recognition using multi-channel physiological signals, demonstrating the potential of advanced signal processing techniques in this field [12].

The application of physiological signals for emotion recognition extends to various fields, as evidenced by the discussion of the use of AI cloud and edge sensors for the recognition of affective, emotional, and physiological states in various applications [13]. Emphasized the use of machine learning algorithms such as Fisher linear discriminant and support vector machines for emotion recognition using physiological signals [14]. These references demonstrate various applications and methodologies using physiological signals for emotion recognition.

As a result, the use of physiological signals for emotion recognition has gained much attention, focusing on overcoming challenges, leveraging advanced signal processing techniques, and exploring various applications across fields. The robustness and reliability of physiological signals make them valuable for inferring emotions, paving the way for advances in emotional computing, healthcare, and human-computer interaction.

1.1 Purpose of Thesis

This study focuses on the detailed categorization of different physiological conditions and a thorough analysis of the relationships between children with ASD and social robots within the context of the EMBOA project. Additionally, the study examines the emotional awareness of typically developing children. The overall objective of this multifaceted study is to explore various dimensions of engagement between these children and social robots, particularly during therapeutic sessions aimed at promoting the children's development and well-being.

The study aims to identify the complex range of emotional reactions demonstrated by the children during their multiple interactions with adults. These emotional signals and trends have the potential to significantly enhance the effectiveness and scope of specialized therapeutic interventions developed for children with autism.

The following research questions inspired us to conduct this thesis.

1. Is there a connection between stress and physiological data?
2. Which physiological characteristics are more determining in emotion analysis?
3. Are there any significant differences between ages in emotion analysis?
4. Do children experience emotional and physiological differences in child-robot interaction?
5. Which classification methods could be used in this type of classification problem?
6. What kind of setup is required for the stress identification system?

1.2 Thesis Overview

We organize the thesis as follows:

- Chapter 1 provides an overview of the research objectives, methodology, and key hypotheses.
- Chapter 2 analyzes existing scholarly work, identifying gaps and laying the foundation for the current research.
- Chapter 3 details the approach, tools, and techniques employed in data collection and analysis and presents the findings derived from the conducted research.
- Chapter 4 summarizes the key findings, significance, limitations, and potential future research avenues.

2. LITERATURE REVIEW

A literature review was conducted in the fields of emotion recognition and ASD to give an idea of the general framework in stress detection. The research conducted is grouped under two headings as follows.

2.1 Emotion Recognition

Memar et al., 2021 investigate the relationship between human stress levels and Electro Dermal Activity (EDA) [15]. The *EDA* is also called *Galvanic Skin Response*. The main aim of the paper is stress level classification using physiological sensors. They utilized a driver dataset collected during real-world driving, available on PHYSIONET. The age interval of drivers is not mentioned. Their data set signals are taken from hand and foot EDA sensors. To classify stress, they used certain metrics, including rise time and amplitude. They verified that foot signals are more reliable. Their stress recognition accuracy is 95.83% by using ANOVA.

Shukla et al., 2019 they try to classify people's emotions [16]. They make a wide comparison of literature works and implement various methods. They compared the EDA signal with three different feature selection methodologies. These methods are Joint Mutual Information (JMI), Conditional Mutual Information Maximization (CMIM), Double Input Symmetrical Relevance (DISR), and machine learning techniques. They searched 25 studies. They implemented these methods in the AMIGOS dataset. This dataset has 40 participants, and the data is collected while they are watching different types of movies (happy, sad, etc.). They used Mel-Frequency Cepstral Coefficients (MFCC) for analysis. They try to identify the most significant feature to find valence and arousal data. Three feature selection methods verified that an avg. of 95 features are used for arousal recognition and an average of 96 features for valence recognition. Results have an average accuracy of 85.75% for arousal and an average of 83.9% percent for valence identification. The study also examined the findings from 25 research papers on the same topic, providing a comprehensive overview of the existing literature.

The aim of the work [17] is to identify children's engagement with an adult during social interaction. The dataset consisted of 51 child-adult data. The adult participant rated the children with the aspect of engagement. Support Vector Machines (SVM) were used to assess the level of engagement of the children. Children between the ages of 15 and 30 months were tested in the paper. The average age of the children was 21 months. To eliminate motion artifacts, they used the Hanning filter. The window of the filter is 1 second was chosen. They also make normalization and decomposition. They extract these features: mean, standard deviation, area under the curve, relative positions of maximum and minimum values, slope, and peaks. They used a 10-fold cross-validation protocol with SEM. They found that a combination of features from electrodermal activity (EDA) and features capturing the physiological synchrony between the child and the adult resulted in the highest classification accuracy. According to the paper, the Hanning filter can be used to have better signal quality for EDA signals.

Zangroniz et al., 2017 created a new monitoring device to record physiological signals. The paper also has a schematic of their electrical system to create a smart band. Also, another aim is to classify calm and stress conditions using EDA (Electro Dermal Activity). There are 50 participants involved in the test. They used the International Affective Picture System to control the experiment. They preprocessed the signal using a 1.5 Hz cut-off low-pass FIR filter with order N = 32. They extracted temporal, morphological, and frequency features. For statistical analysis they used ANOVA. They report around 89% accuracy to identify distress and calm conditions. They verified that there is a correlation between EDA and calm condition [18].

Lutin et al., 2021 collected different feature extraction methods these are: trough-to-peak features, decomposition-based features, frequency features, and time-frequency features to verify stress conditions. They used Ledalab, cvxEDA, and sparsEDA applications to extract these features. Their dataset has 20 people who performed 3 different stress tasks. Support Vector Machine (SVM) classifier with

Leave-One-Subject-Out Cross Validation is used. They used the number of responses (from Ledalab). The classifier accuracy is 88.52%, sensitivity 72.50%, and specificity 93.65%. The metrics showed that EDA can be used as stress detection, also they used Ledalab, cvxEDA, and sparsEDA in their applications [19].

Zantone et. al, 2019 analyzes the stress in a driver during unexpected events or taxing situations. They used both ECG and EDA signals. They used the Supervised Machine Learning technique. They make their test in professional simulations that have wide screens and cockpits. The test scene also has different environments like highways and unexpected objects and situations. Their dataset has 16 participants with the age of 22-47. The study has a balanced accuracy in stress detection of 77.59%. In the work, their test simulation seems successful in feeling different emotions [20].

The aim of the study [21] is to demonstrate a method for detecting stress levels. Liu et al. 2017 aim to use a single physiological signal to detect to make it more practical. They used linear discriminant analysis (LDA) to separate stress levels as low, medium, and high. MIT Media lab 'stress database' is used. 11-foot data are chosen for the experiment. Five-minute data segments were used to extract EDA characteristics while driving in three different conditions: in the city, on the open road, and at rest. To categorize stress levels, they used linear discriminant analysis and Fisher projection. To categorize stress levels, they use linear discriminant analysis and Fisher projection. The accuracy is 81.82% measured.

The main purpose of the project [22] is to detect stress in real life. The main difference is that all tests in this work are conducted in real-world scenarios rather than in a laboratory environment. Can et al., 2019 organized a 9-day algorithmic programming contest event. The event has different sections. Data was collected from 21 participants. They used Empatica E4 and Samsung S devices to collect physical data. They used Hart Rate (HR), skin conductance(EDA, SC), and accelerometer signals. They used the EDA Explorer tool to eliminate noise from the signal. They eliminate the data if the EDA explorer labels the batch as noisy. They eliminate the artificial peaks brought on by elevated body temperature or physical exertion. (using accelerometer data). They used the cvxEda tool to extract features. They extracted the mean, standard deviation, peak, strong peak, 20th percentile, 80th percentile, and quartile deviation of the EDA. They obtained 90.40% accuracy by using Empatica E4 devices, whereas the accuracy with Samsung S devices was 84.67%. The paper has well-prepared literature research on it. Also EDA Explorer tool is easy to use.

Can et al., 2020 mentioned that high stress affects life quality and health. They aim to create a stress level detection system. They try to improve the performance of daily stress detection. They compare the daily life, hybrid, and laboratory-daily life models. In the laboratory data collection side, they have a controlled environment, they apply Trier Social Test (TSST) to participants during test tasks. 14 participants were involved in the test. The steps in their testing are as follows: baseline pre-stress measurements; stress-inducing TSST (Trier Social Stress Test); and post-stress recovery measurements (recovery). Participants complete the 14-question Perceived Stress Scale (PSS) in the pre-stress measurements section while at rest. During the test, they are told to prepare a speech explaining why they study and why they would be a good fit for their dream job. This speech will be recorded and reviewed by the research psychologists. You have five minutes to prepare, and your time starts now.

After time passes interviewers ask the participant to continue in English. Then give them to number subtraction task. They also make tests in out-of-lab environments. While making daily life tests, participants were responsible for filling out the PSS-5 questionnaire. In the feature selection part, they extracted these features Mean value, Standard deviation, Number of peaks, Number of strong peaks, twenty-eighth percentile, eightieth percentile, and quarterile deviation using cvxEda tools. They used 5 ml algorithms to classify these as multi-layer perceptron (MLP), random forest (RF), k-nearest neighbor (kNN), support vector machine (SVM), and logistic regression (LR). They obtained a maximum of 94.4% accuracy with HR, 86.70% with EDA, and 92.30% with the combination of HRV and EDA [23].

Kyamakya et al., 2021 aim is to list all the methods and sensors used in stress detection. They made a literature review in this field and categorized it. They categorized these papers as Stress level recognition, Wearable body sensors, Dermatological sensors, and Facial expression recognition. They mentioned the AffectiveROAD dataset and

Empatica E4 device. Also, stress detection can be done with EEG, ECG, and EDA sensors and different ML methods are used in literature such as RF, CNN, kNN, and SVM. The paper has a wide literature review in this area [24].

The project aims to explore the effect of emotions evoked by music on HRV (Heart Rate Variability). Thirteen pianists participated in the test. Nakahara et al., 2009 perform expressive and nonexpressive performances during tests. The pianists are 10 females and 3 males 22–32 years of age. They firstly play expressive music, then play music without emotions, thirdly listen to their expressive performance with emotions, and finally listen to their nonexpressive-performance music. The dating process was statistically tested using two-way ANOVA. They verified that expressive performance has a higher HR than the other tasks, and nonexpressive has the lowest HR. In our work with children, it is quite hard to express some emotion. In the work they demonstrate that there is a relation between emotion status and music [25].

Tonacci et al., 2020 performed a relaxation test environment based on a short audio and video, and analyzed signals related to the autonomic nervous system (ANS) activity. They used electrocardiogram (ECG) and galvanic skin response (GSR). 24 volunteers participated in the test. Test setup is made in two ways, the first group testing schedule performs an audio-video + video-only relaxation, and the second group performs an audio-video + audio-only protocol. They state that the first group significantly differs in the HRV-related SDNN parameter. They used a neural network and have an accuracy of 79.2% [26].

In the research, Castro et al. 2022 list a low-cost wearable. These are OpenBCI and three other open-hardware custom-made designs that communicate via Bluetooth applications for data synchronization and storage. They did all of the tests in the laboratory but did not mention the count of participants. They access electroencephalography (EEG), electrocardiography (ECG), breathing rate (BR), electrodermal activity (EDA), and skin temperature (ST). They used a low-pass filter for EDA, ST, and BR, and they used a cutoff frequency of 5 Hz, based on a Gaussian window length of 50 implemented in Ledalab. They created a test that had a stress task session after that test period. Stress task has arithmetic operations. They have a label for their data. At the and they have a low-cost wearable watch [27].

Chen et al., 2021 assume that there is an acceptable correlation between stress and pain. They are trying to create a system to detect pain and stress to wearable. They listed different signals heart activity, brain activity, muscle activity, electrodermal activity, respiratory, blood volume pulse, and skin temperature. They list all of the stress tests. These are the Stroop Color-Word Inference Test, Trier Social Stress Test, Cold Pressor Test/Hot Water Immersion Test, and International Affective Picture System Test. Also, the signals that have the emotion information are listed. Sensors options are listed. The paper has wide research that has different alternatives [28].

Van Lier et al., 2019 created a protocol to choose valid data for electrodermal activity and cardiovascular activity signals. They used E4 Empatica wearable. They defined the steps to determine the cross-correlation for EDA data is: 1. Down and up sample the data to the same frequency., 2. Normalize and detrend the data, 3. Determine cross-correlation at multiple time lags, Find the highest cross-correlation with the corresponding time lag and plot these in a histogram. Also in the paper they have some metrics in a normal state like the Number of SCRs are 1–3 per minute, and high arousal state of 20-25 per minute. Normal SCR amplitude etc. They proposed their validity protocol in the paper. The paper only has the protocol to choose valid data [29].

Kyriakou et al., 2019 thought that there is limited work in the real world. Environmental impact is considered in the work. They aim to detect stressful moments. They propose a rule-based method to detect stress. They validate their algorithm in the real world. 11 women and 8 men participate in the test as a volunteer. Their rule-based algorithm has these rules, GSR (Galvanic Skin Response) increase, ST Decrease, Rising Time, Response Slope, Duration has high importance. They evaluate their data with these metrics and have 84% accuracy. Their rule-based system can be improved with temperature sensor output. When the temperature data is not acceptable, it is better to eliminate the data [30].

Kleckner et al., 2017 created an assessment procedure to classify EDA signals. Their methods 20 participants with autism there are 5 females in 5-13 years. 181 hours of

EDA data were collected in their home using Affectiva Q Sensor during 8 weeks. Their aim to classify invalid data. These four rules are as follows: Is the EDA out of range? Does it alter swiftly? Does temperature contain reliable information? Their process confirmed that 21% of the data is erroneous. Additionally, they expertly validate the outcomes. Compare the output produced by the algorithm with the EDA signal expert. The sensitivity of outputs is 91%. Their rules to validate signals are so easy to implement and have high accuracy. These rules can be used in the project by adding some new rules [31].

In the work [32], Chen et al., 2015 propose a method to remove artifacts from EDA signals. They use Stationary Wavelet transform (SWT). They modeled the wavelet coefficients as a Gaussian mixture distribution. They compare their methods with three other methods and find out their methods better. The difference between the paper and others they use SWT in the frequency domain.

This is a review paper and has 164 references. Schmidt et al., 2018 summarise psychological models, emotion models, emotion test types, wearable assessments, and wearables. They also demonstrate data processing chain, and review, feature extraction, and classification. They also listed the most used datasets these are Eight-Emotion, Deap, MAHNOB-HCI, StudentLife, DECAF, etc. This paper has a huge reference and literature view [33].

Redd et al., 2020 aim is to continuously monitor emotion prediction for school-age children. Their models correctly classify the behavioral state of a child with 68% mean accuracy and up to 85%. They used the Empatica E4 device. 5 children aged between 8 and 12 were involved in the study. During the tests, professionals label the status as happy or not. After the test Logistic Regression (LR), Support Vector Machine (SVM), and Decision Trees (DT) are used. They stated that a person-based model is more successful [34].

The purpose of the study is to compare the analysis of four physiological signals. These signals are EDA, Heart Rate (HR), Skin Temperature (SKT), and Blood Volume Pulse (BVP). Oleander et al., 2016 used Empatica E4. 21 people participated in the test aged 17 to 25. They have mental arithmetic sessions to stress makers. They used RF, kNN, and Ada boost Gradient boost. To filter noise they used a Savitzky-Golay filter for SKT and EDA signal. The filter smooths the signal. For HR and BVP, they used a median filter. They extracted median, mode, standard deviation, minimum, maximum, and first derivative of the signals. They calculated the Spearman Rank Correlation factor. They compare these signals with classification methods. The highest accuracy is 99.92% while using 4 signals. EDA and HR also have high accuracy like 99.09% [35].

Conducted experiments to verify a proposed emotion recognition method using a constructed video library, demonstrating the practical application of the method for emotion recognition validation analysis [36]. This study provides empirical evidence of the effectiveness of the proposed emotion recognition approach. Chun-Yan and Wang utilized the Bagging-CHAID model for emotion recognition validation analysis based on the correlation study between chaos features of physiological signals and emotions using the information gain method [37]. This approach highlights the utilization of advanced modeling techniques for verifying emotions from physiological signals.

Furthermore, emphasizes the reliability of physiological signals for emotion recognition, highlighting their robustness as a form of signals for inferring emotions [11]. This underscores the fundamental role of physiological signals in verifying emotions due to their inherent reliability.

In conclusion, the verification of emotions using physiological signals involves a diverse array of methods, including machine learning algorithms, data fusion, and advanced signal processing techniques. These approaches contribute to the development of robust and reliable emotion recognition systems based on physiological data.

2.2 Autism Spectrum Disorder

ASD refers to a complex neuro-developmental condition that consists of deficits and inabilities in social engagement with other individuals and repetitive behaviors. ASD is a complex disease involving multifactorial causes, subtypes, and varying developmental courses. People with autism spectrum disorder have problems interacting and communicating, have narrow ranges with little diversity in interests, and show stereotypical and repetitive behaviors [38,39]. It is associated with dysfunctions in core behavioral dimensions such as compulsions, social awkwardness, and verbal abnormalities. There have been studies showing that neurobiology could be used as an appropriate susceptibility or risk biomarker for ASD. Studies have shown that the incidence or occurrence of ASD is increasing, and their range from 0.07 % - to 2.64 %). Moreover, people diagnosed with ASD can suffer from several other psychopathologies, like depression, ADHD, or OCD. Dependable and valid measurement instruments are necessary here too.

ASD can cause a variety of challenges, including limited interests and repetitive behavior, as well as difficulties with social interaction, language development, and communication [39]. Moreover, children diagnosed with autism spectrum disorder (ASD) may experience challenges with self-harm, physical exercise, and eating habits selectivity. In addition, they might experience hindrances while mastering a second language and show problems in explicit and implicit social cognitive tests [40,41]. In addition, eating disorders and difficulty swallowing may be experienced by some children with AS D, and there are reports that these children have problems using electronics as well [42,43]. This does not change in hospital settings as members of society become more conscious of children with autism spectrum disorder.

Finally, autism spectrum disorder is complex with many facets as well as numerous problems. Understanding the complex nature of autism spectrum disorder will make it possible to offer proper assistance and interventions that can be useful to the members of society who are living with the condition.

3. STRESS RECOGNITION SYSTEM

In order to verify emotions like stress and happiness using physiological signals, various methods and techniques have been explored in the literature. These methods encompass a wide range of approaches, including the use of machine learning algorithms, physiological signal processing, and data fusion. The referenced studiesprovide insights into the diverse methodologies for verifying emotions through physiological signals: presented a comprehensive review of physiological signal-based emotion recognition, covering emotion models, elicitation methods, datasets, features, classifiers, and the overall framework for emotion recognition based on physiological signals [8].

In Chapter 2, we provided an overview of the existing literature, methodologies employed, and datasets used. A notable absence in the literature exists regarding the recognition of emotions and stress levels in both individuals with Autism Spectrum Disorder (ASD) and children generally. Our study seeks to address this gap by providing additional insights. Our work focused on three key areas:

1. **Studies under the Emboa Project:** As part of the Emboa project, physiological data from children with ASD during robot interactions were collected and analyzed across various countries. Signal processing methods were applied to analyze the data. However, due to limitations in conducting stressful tasks comparable to those used in adult studies, statistical analysis was conducted based on findings from comprehensive literature reviews.

2. **Stress Analysis using the Stroop Test:** We employed the Stroop test to recognize emotions in typical primary school students. This test involved three stages, and the resulting data were categorized according to their specific objectives.

3.1 Data Collection Using Empatica E4

Empatica E4 [44] physiological wristband was used in all studies within the scope of the thesis. The E4 is equipped with sensors that record specific physiological data. It is a wearable device and allows users to continuously monitor their physiological data. It can measure many biometric data such as heart rate (BVP - 64 Hz), skin conductance (EDA - 4 Hz), body temperature (ST - 4 Hz), and movement (ACC - 32 Hz) as seen in Figure 3.1 and Table 3.1. This data helps researchers monitor stress levels, emotional states, or physical activities. Used in healthcare and academic research, such devices are becoming important tools for understanding and monitoring biological responses.

Figure 3.1 : Empatica E4 features.

The wristband can collect multiple sensor data and display data in real-time with the help of Bluetooth as seen in Figure 3.2. As a result of the test, the data can be backed up and processed with Empatica Cloud. Although more than one sensor data was obtained from the wristband, EDA and BVP data were used because it is more decisive in defining stress within the scope of the literature.

Table 3.1 : Empatica E4 sensors: sampling frequencies.

Physiological Signal	Sampling Frequency
EDA	4 Hz
BVP	64 Hz
ACC	32 Hz
ST	4 Hz

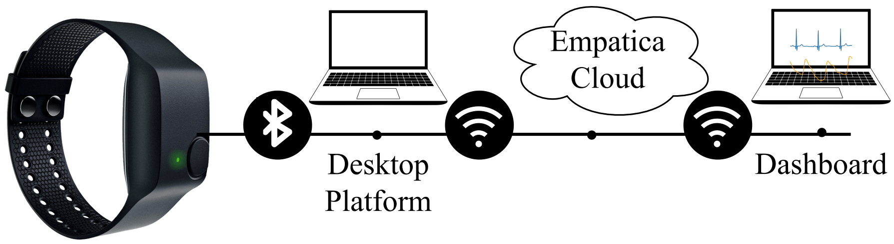

Figure 3.2 : Empatica E4 data recording.

3.2 Feature Extraction

Specifically, we analyzed the Blood Volume Pulse (BVP), Skin Temperature (ST), and Skin Conductance (SC) signals obtained from the Empatica wristband. To address differences in sampling frequencies among the signals, we utilized techniques such as filtering, normalization, and decomposition to extract meaningful features. Firstly, the raw BVP signals were subjected to a sixth-order Chebyshev II filter with an attenuation coefficient of 18 dB and a normalized stopband edge frequency of 0.1 Hz to remove noise and artifacts. Subsequently, the signals were normalized between 0-100 to facilitate further analysis. Similarly, the BVP and EDA signals pass a fifth-order Savitzky-Golay filter with a frame length of 11 to remove noise and enhance their spectral characteristics. Following normalization between 0-100, these signals were decomposed into phasic and tonic components using a linear regression model.

Regarding the ST signal, which is less susceptible to noise compared to other physiological signals, no preprocessing steps were deemed necessary due to its excellent quality. Consequently, the raw ST signal was used directly for feature extraction without any modifications. Notably, a total of 80 features were derived from the raw data, including both statistical and mathematical features obtained through various methods applied to each physiological signal separately or jointly. These findings contribute significantly to the development of more accurate machine-learning models for emotion recognition systems based on wearable sensor technology.

According to the information provided, a total of 80 features were derived from the original data. These features were obtained by analyzing the mentioned signals. The extracted features are presented in Table 3.2.

Table 3.2 : Selected elements and their abbreviations.

Abbreviations	Description
Signal_mean	Mean of the signal
Signal_med	Median of the signal
Signal_std	Standard deviation of the signal
Signal_min	Minimum of the signal
Signal_max	Maximum of the signal
Signal_skew	Skewness of the signal
Signal_kurt	Kurtosis of the signal
Signal_q05	0.05 Quantile of the signal
Signal_q25	0.25 Quantile of the signal
Signal_q75	0.75 Quantile of the signal
Signal_q95	0.95 Quantile of the signal
Signal_var	Variance of the signal
Signal_ran	Range of the signal
Signal_max_min	$Signal_{max}$ / $Signal_{min}$
Signal_iqr	Inter-quartile-range of the signal
Signal_coeffvar	Coefficient of variation of signal

“Signal" represents BVP, SCR or SCL for EDA, and ST

3.3 Emboa Project

3.3.1 Data collection & test setup

The goal of this research is to develop a novel strategy for establishing an affective loop in robot-based therapy for children with ASD to enhance the intervention’s impact on the development of emotional intelligence. A series of interaction scenarios were implemented using the Kaspar robot to investigate the technologies and techniques that are currently available and fulfilling this purpose, the setup as seen in Figure 3.3 and Figure 3.4.

Every interaction scenario was built around the concepts of imitating others and switching roles. These required linguistic abilities on emotions, animals, and bodily

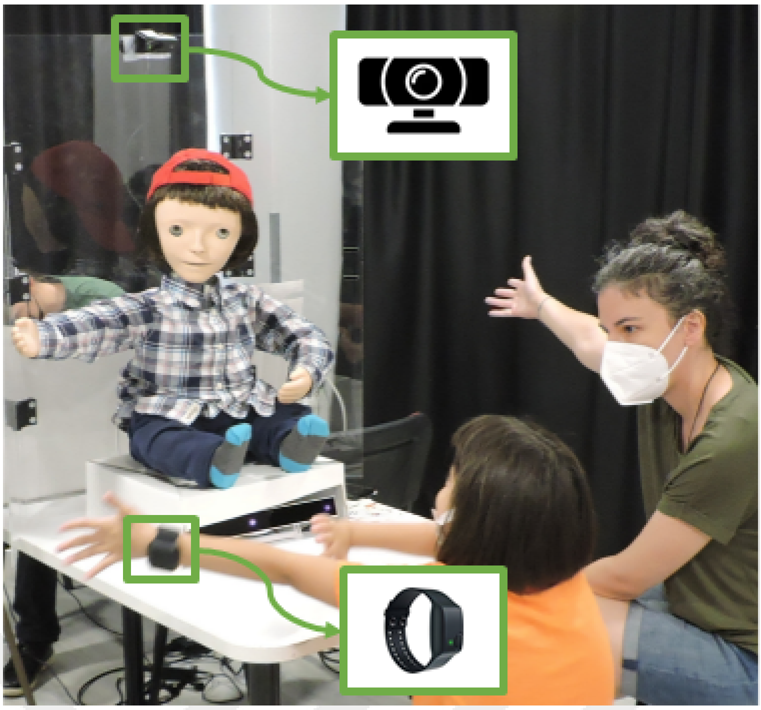

Figure 3.3 : Experimental setup for the child-robot interaction sessions.

parts in addition to a basic understanding of English. Throughout the session, Kaspar guided the kids with positive, neutral, or negative feedback using simple phrases and a few behavioral signals. The children's profiles informed the therapist's determination of the interaction flow and its adaptation.

Multi-modal interaction data was collected during the interaction experiments to assess and examine the suggested intervention plan. The E4 wristband recorded physiological signals; the Gazepoint Eye Tracker recorded eye gaze motions, duration, and fixation data; the H4n Pro sound system recorded audio; the two video cameras recorded video, one above the head of the robot to record children's reactions and the other on the robot's right side to record the entire interaction session; the children's demographic profile and diagnostic history made up the remaining data.

The interaction studies were carried out by four collaborating countries: North Macedonia (MAAP), Poland (GUT), the United Kingdom (UH), and Turkey. Table 3.3 shows the 29 children (25M, 4F) with an ASD diagnosis who were part of the study. The children varied in age from 2 to 12 years old, with moderate to high levels of verbal abilities and linguistic awareness. Each child took part in the interaction research at least once, and it was repeated in two to eleven sessions based on the therapist's assess-

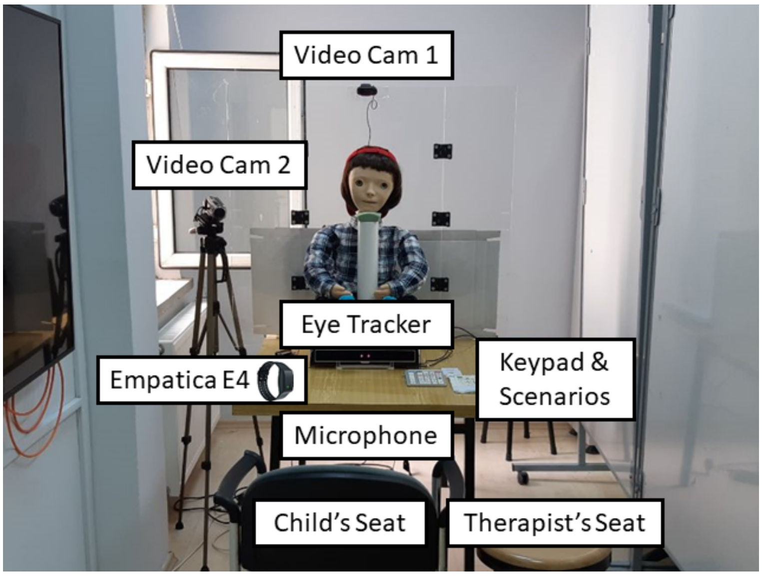

Figure 3.4 : Setup overview.

ment. The ethics commission authorized the interaction research projects at Gdansk University of Technology (Poland). In addition, the children's family members gave written consent and were informed about the study methodology.

Table 3.3 : Group features.

Group Number	Count (Gender)	Age Range
Group 1 (G1)	10 (9M,1F)	2-5
Group 2 (G2)	14 (12M,2F)	6-8
Group 3 (G3)	8 (7M,1F)	9-12

3.3.2 Signal analysis

This phase of the study involves a stress detection method based on signal analysis in children with autism. The aim is to develop an approach using social and assistive robots during therapy sessions for children with autism. Data were collected from children of various ages and nationalities as part of the EMBOA project. During interactions with a robot in a game scenario, the E4 smart bracelet recorded and analyzed Electrodermal Activity (EDA) and Blood Volume Pulse (BVP) values.

As the collected data lacked labels, analysis was conducted based on existing literature. The EDA signal's number of peaks and amplitude values were processed using average lower limit values from previous studies to determine stress levels in children.

Additionally, Low Frequency (LF) and High Frequency (HF) values from the BVP signal were meticulously examined. A comparison was drawn between these values and stress indicators.

Physiological data of children with autism, such as BVP, ST and EDA, were collected with the help of the E4 wristband during therapy sessions with Kaspar. The BVP signal was sampled at 64 Hz, while the EDA and ST signals were sampled at 4 Hz. This was due to the sampling frequency of the bracelet. The E4 bracelet had to be worn tightly on the child's wrist without obstructing blood circulation, so some children felt uncomfortable.

After data collection, data processing started after all data was backed up to Empatica Cloud. BVP features were extracted using the HeartPy Python Heart Rate Analysis Toolset. The most important BVP components extracted are SDNN, PNN50, RMSSD, HR, and frequency domain features, which are LF and HF. The usability of the features was verified using the value ranges in the literature [45]–[50]. The literature's ranges were used to rate the signal qualities.

The peak count and average amplitude of the EDA signal, along with the LF and HF values of the usable BVP signal were used to identify stress in children. A prior publication in [50] included the reference LF and HF values for kids with ASD. Research has shown that under stressed conditions, LF power rose while HF power fell. Consequently, the baseline for stress detection has been the previously published mean values for HF (M = 3127ms^2) and LF (M = 2243 ms^2). Stressed-out children were those whose computed LF value was higher and whose HF value was lower than the reference. Stressed-out children were those whose computed characteristic was higher than the reference.

Average peak counts, or "SCRs," and amplitude values were employed for the EDA signal's recognition of stress. The baseline metrics that are provided in from [51] and [52] have been used while detecting stress with EDA. Prior research revealed that the range of EDA signal is from 0 to 20 µS [53]. The average number of SCRs per minute is 1-3. In case of high arousal, it is 20-25 [52]. Additionally, the resting mean amplitude is 0.66±0.13 µS for ASD [51]. MIT EDA explorer tool [54] and cvxEDA library were used to extract peak counts, amplitudes, Area Under the Curve (AUC), skin conductance response (SCR) width, decay time, rise time, phasic component, and tonic component features from EDA. Threshold 0.02, offset 1, Rise time 4, Decay Time 4, these constants are used to extract the peak values from the signal. While detecting peaks, default constants in the library were used [55]. SCR amplitude should be higher than the threshold (0.02) to be counted as a peak with this threshold, we aim to eliminate small changes. Rise time also called onset-to-peak time varies between 0.5 and 5s [56].

3.3.3 Statistical analysis

There are cases of using statistical methods in studies to investigate differences between groups for stress detection analysis. One-way ANOVA was used in the study conducted by Shakiba et al. The purpose of ANOVA is to analyze whether there is a relationship between the levels of stress experienced by individuals within families or communities and physical responses to stress. This evaluation shows whether the data collected for both groups are related to each other [57]. Acute stress was found to impair the health of memory through the ANOVA study conducted by Merz and colleagues (2020), demonstrating that stress is associated with cognitive effects and the utility of ANOVA in exploring these phenomena [58].

Statistical tests such as ANOVA and t-tests have been used in many studies to examine the impact of stress on different aspects, including physiological responses and behavior. These methods helped researchers compare groups exposed to different levels of stress and identify differences in stress responses [59]–[61].

Within the scope of the Emboa project, physiological data of children with autism from different age groups were collected in more than one country. Since labeling could not be done during or after the session, physiological data differences between these age groups were made on a feature basis. During this analysis, comparisons among all groups were conducted using ANOVA, while pairwise comparisons were performed using T-tests.

3.3.3.1 ANOVA

Since there were different age groups between the groups, children with similar age groups were divided into G1, G2, and G3. One-way ANOVA test was applied with existing groups, shown in Table 3.3. As a result of the analysis, when the differences between age groups were examined, it was noticed that the amplitude-dependent changes of BVP were more distinctive.

Among all EDA features, those with the highest maxima, 95th quantile, and amplitude ranges were found to be most distinguishing. Additionally, the standard deviation, minimum, and maximum-to-minimum ratio of the Skin Conductance Response (SCR) signal were also significant. These results suggest that both the raw and phasic components of the EDA signal are affected by age, with the phasic component potentially being more pronounced, shown in Table 3.4.

Of the 16 ST features examined, 12 were found to be statistically distinguishable across age groups. However, the standard deviation, variance, interquartile range, and coefficient of variation of these signals were not particularly informative. Interestingly, the minimum, maximum, and range of signal amplitude varied significantly between the three modalities of BVP, EDA, and ST.

In essence, an examination of the 39 BVP, 48 EDA, and 16 ST features revealed that 3 BVP features, 13 EDA features, and 12 ST features were significantly different among the various age groups. Overall, our findings highlight the importance of considering age when analyzing physiological signals, as it can impact the interpretation of the results. Specifically, certain features of the EDA and ST signals were found to be more distinguishing across age groups, while others were not. These insights can help guide future research and clinical applications in the field.

3.3.3.2 T-test

A pairwise comparative t-test statistical analysis method which is a widely used method for comparing the changes between different groups' characteristics [62] was performed between age groups in this study. All groups were compared with each

Table 3.4 : ANOVA results between ages (just significant values (p-value <0.05)).

Signal	Features	F-Value	p-Value
Blood Volume Pulse	BVP_min	4.3	0.03
	BVP_max	3.8	0.04
	BVP_ran	4.3	0.03
Skin Conductance	SCR_std	5.5	0.01
	SCR_min	8.8	<0.01
	EDA_max	4.0	0.03
	SCR_max	10.7	<0.01
	SCL_max	4.0	0.03
	EDA_q05	5.4	0.01
	EDA_q95	3.8	0.03
	SCR_q95	4.5	0.02
	SCL_q95	3.8	0.03
	EDA_ran	4.0	0.03
	SCR_ran	10.4	<0.01
	SCL_ran	4.1	0.03
	SCR_max_min	5.1	0.01
Skin Temparature	ST_mean	206.6	<0.01
	ST_med	3.8	0.04
	ST_min	18.8	<0.01
	ST_max	4.5	0.02
	ST_skew	6.9	<0.01
	ST_kurt	4.4	0.02
	ST_q05	5.6	0.01
	ST_q25	4.0	0.03
	ST_q75	3.7	0.04
	ST_q95	3.6	0.04
	ST_ran	5.8	0.01
	ST_max_min	6.9	<0.01

other using ANOVA. T-test was used to compare paired groups with each other. It was observed that 9 acquired features were more distinctive between Group 1 and Group 2 (Table 3.5). It also revealed that the distinctive features that distinguished Group 1 from Group 3 were based on ST.

Distinguishing features were examined in pairwise comparison for Group 2 and Group 3. It showed that their averages differ significantly depending on the feature extracted from the EDA signal, as shown in Table 3.5. When the important features of the groups were examined, more features that showed statistically significant differences between Group 2 and Group 3 were found (Table 3.5). Between Group 1 and Group 3, there were 4 features with a p-value less than 0.05 (Table 3.5). When the results were examined between groups, it was seen that EDA-derived features were more prominent.

3.3.4 Experimental results & discussions

3.3.4.1 Signal analysis

It was checked whether the data collected from all countries were within the required ranges. Feature extraction was made from physiological data such as BVP and EDA. Each of the sessions performed by the children with the robots was examined. Within the scope of the study, examinations were made on the physiological analysis of the participant anonymously labelled as C07. To extract signal features, 2-minute windows were selected and features were extracted. The child received 7 sessions of therapy with Kaspar. BVP HF - LF values and EDA peak number and amplitudes are important to us in the therapies. Data regarding this are presented in Figure 3.5 and Figure 3.6. The participants' average data between sessions and baseline data found in the literature were also examined. When the pictures were examined, it was observed that the 4th and last session was not stressful according to BVP features. When we examine the inferences in the EDA data, these results are not supported. The peaks are lower than the baseline but higher than the individual mean. The amplitude value is also higher than the average. According to these data, we can say that the boy was more

Table 3.5 : T-Test results between ages (just significant values (p-value <0.05)).

Groups	Features	T-Stats	p-Value
G1 vs G2	BVP_med	-3.06	<0.01
	BVP_min	-2.6	0.02
	BVP_max	2.5	0.02
	BVP_ran	2.6	0.02
	SCR_min	2.9	<0.01
	SCR_max	-3.2	<0.01
	SCL_max	4.0	0.03
	SCR_ran	-3.1	<0.01
	ST_skew	2.58	0.02
G1 vs G3	ST_min	3.8	<0.01
	ST_skew	3.04	<0.01
	ST_kurt	-2.2	0.04
	ST_max_min	-2.15	0.05
G2 vs G3	EDA_mean	3.8	<0.01
	SCL_mean	2.54	0.02
	EDA_med	2.25	0.03
	SCL_med	2.24	0.03
	SCR_std	3.25	<0.01
	SCR_min	-3.33	<0.01
	EDA_max	2.72	0.01
	SCR_max	4.05	<0.01
	SCL_max	2.7	0.01
	SCR_skew	-2.5	0.02
	EDA_kurt	-2.6	0.02
	SCR_kurt	-2.08	0.05
	SCR_q05	3.08	<0.01
	EDA_q25	2.13	0.04
	SCR_q25	-2.14	0.04
	SCL_q25	2.16	0.04
	EDA_q75	2.36	0.03
	SCR_q75	2.47	0.02
	SCL_q75	2.34	0.03
	EDA_q95	2.66	0.01
	SCR_q95	3.08	<0.01
	SCL_q95	2.63	0.01
	SCR_var	2.24	0.03
	EDA_ran	2.72	0.01
	SCR_ran	0.001	3.78
	SCL_ran	0.012	2.68
	SCR_max_min	0.003	3.15
	SCR_iqr	2.35	0.07

stimulated compared to the beginning of the therapy, which may have occurred due to his interaction with Kaspar.

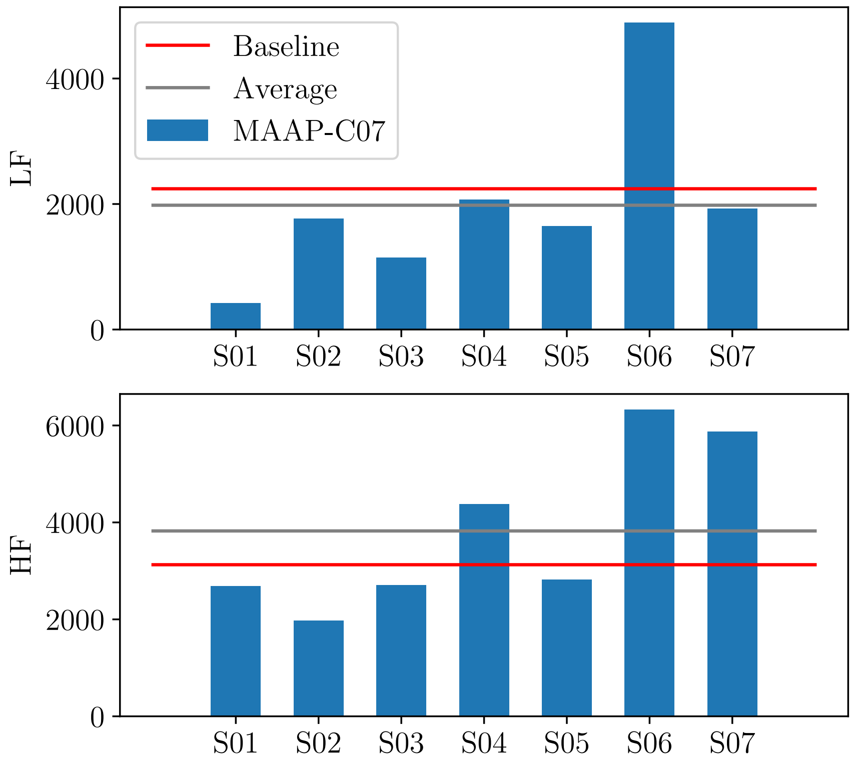

Figure 3.5 : BVP features for all sessions.

BVP and EDA signals, which we examined in the data between sessions, gave us different information compared to the information we received from the literature. For this reason, the 4th session was examined in more detail. With windowing, the data obtained during the 4th session was divided into parts, and feature extraction was performed again. As seen in the figures, if we look at the literature data, it was determined that the low and high-frequency values were compatible with the amplitude values and the 4-6 minute intervals corresponded to stress. When we examine all sessions and features, we can say that all signals between minutes 2-4 indicate that the child is stressed, shown in Figure 3.7 and Figure 3.8.

The data for the 4th session are presented in Table 3.6 at 2-minute intervals. The table includes the baseline value taken from the literature and the person's threshold value. Emotion states were determined considering the data. This chapter is based on the paper "Stress Detection of Children With ASD Using Physiological Signals" [63].

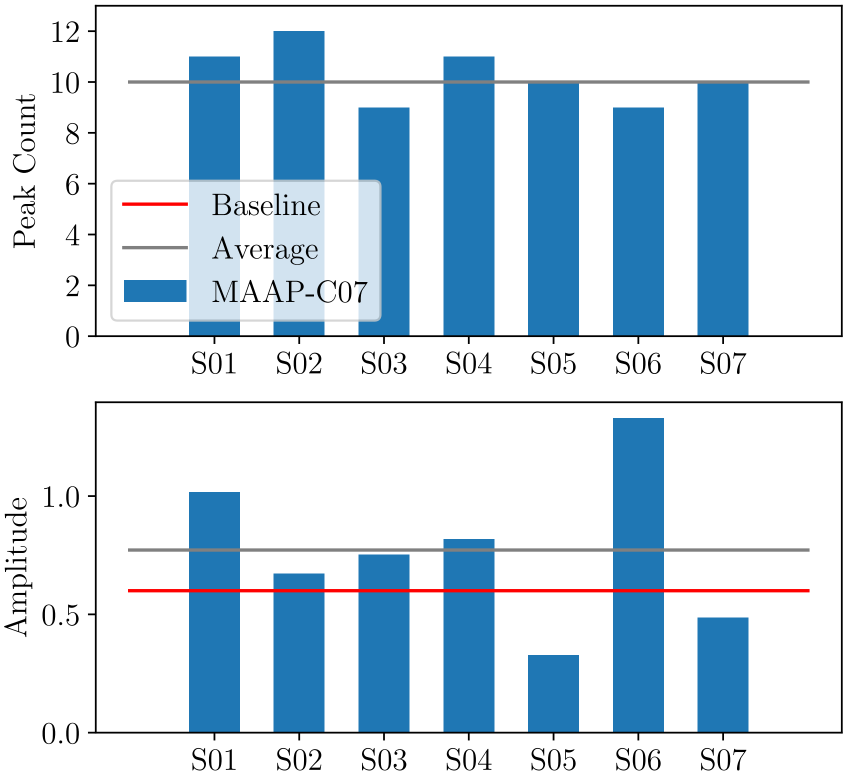

Figure 3.6 : EDA features for all sessions.

3.3.4.2 Statistical analysis

In this part of the study, it was examined whether there was a significant difference in physiological signals between the groups of children participating in the Emboa project ($p < 0.05$). The children were divided into 3 groups according to their age range. The results showed that the amplitude-related features of BVP were the most distinctive between age groups for all 3 groups.

Table 3.6 : 4th session statistics for MAAP-C07.

Interval	LF	HF	Peaks	Amp	Emotion
0-2*	2565.34	3691.96	16	0.7127	sad
2-4**	2912.78	3565.82	13	0.9349	sad
4-6*	2439.87	2852.65	8	0.8198	happy
6-9	980.68	2036.96	8	0.8049	happy
Baseline	2243.00	3127.00	25	0.6600	-
Threshold	1783.05	4062.36	10	0.7355	sad

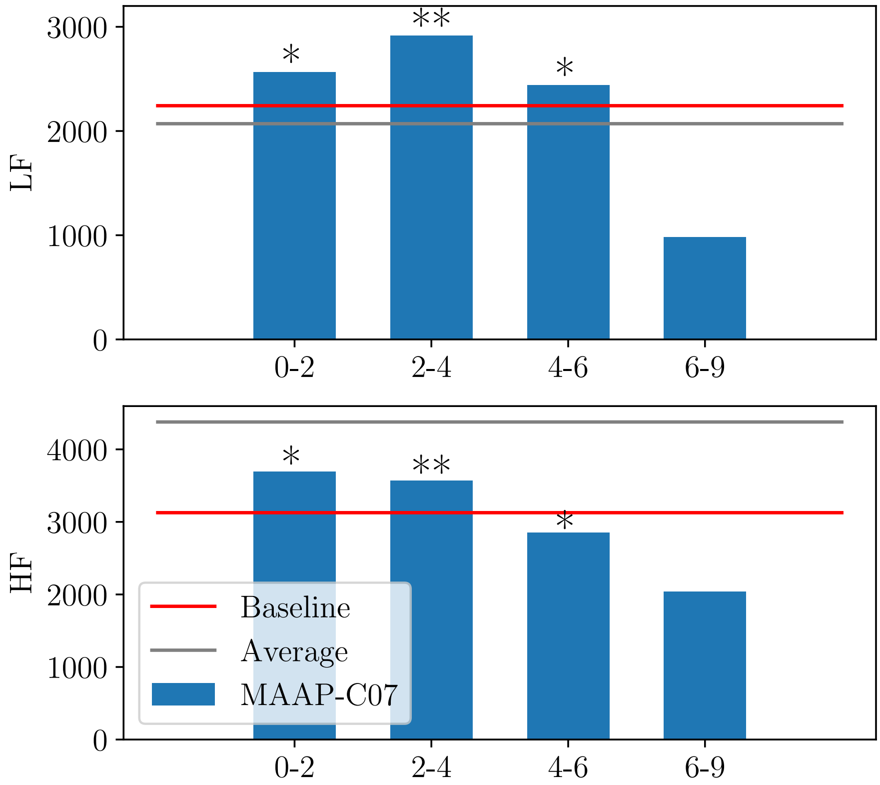

Figure 3.7 : BVP features for the 4th session.

It has been observed that features with the maximum, 95th quantile, and amplitude range for EDA, SCR, and SCL are distinctive. According to the results, phasic and tonic components of EDA play an active role as distinguishing features. 12 of 16 ST features were found to be statistically significant 3.5) and 3.4). Overall, 3 of 39 BVP features, 13 of 48 EDA features, and 12 of 16 ST features were statistically significant. Findings show that the ST signal is more pronounced among different age groups. This may be due to environmental conditions if the watch is not seated properly. This chapter is based on the paper "Investigation of Physiological Features by Age Groups in Children with Autism" [64].

3.4 Stress Analysis Using the Stroop Test

Within the scope of the Emboa project, the physiological data of the participants were analyzed according to the metrics in the literature review, comparing age groups and stress differences between sessions. Since there were no labels in the sessions held within the scope of the Emboa project, a different test was conducted using the Stroop

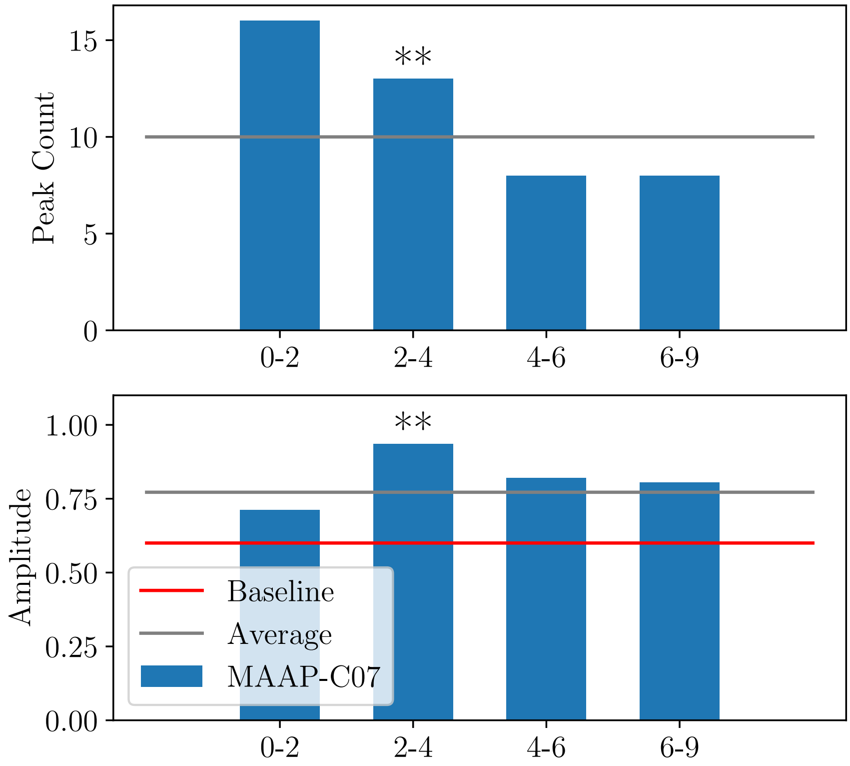

Figure 3.8 : EDA features for the 4th session.

test with typically developed primary school children to classify stress detection in children.

Stroop test is a psychological test used to measure cognitive abilities such as attention and focus. This test creates a challenge for the person regarding the names of the colors. For example, if the word "Green" is written in red, the person is asked to name the written color of the word (i.e., expected to say "red"), while ignoring the meaning of the word. This conflict inhibits the brain's tendency to automatically respond based on the meaning of the word and may cause the individual to respond more slowly or incorrectly. Incorrect answers make the person feel under intense pressure in short-term tests. The Stroop test is frequently used in psychology and neuropsychological evaluations.

In this test, sessions in which the Stroop test was performed are labeled as stressful. As seen in the Figure 3.9 and Figure 3.2, it took 5 minutes in total. First, a baseline recording was taken. In the stress area, a Stroop test was performed in a limited time.

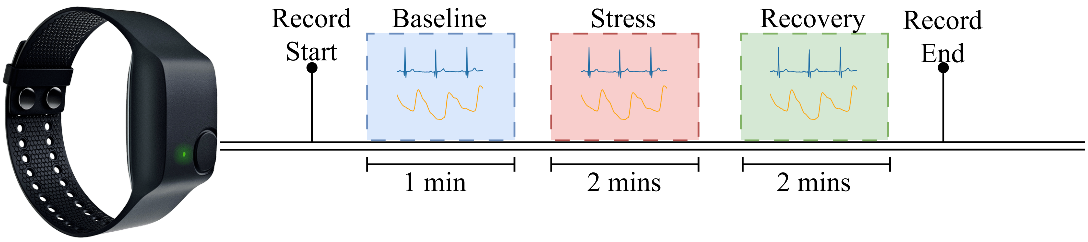

Figure 3.9 : Stroop test procedure.

3.4.1 Signal analysis

The data obtained in the Stroop test were collected using Empatica E4, as in other tests. Physiological data were transferred to Excel with the help of Cloud. First, raw data was created on a student basis to visualize the sessions.

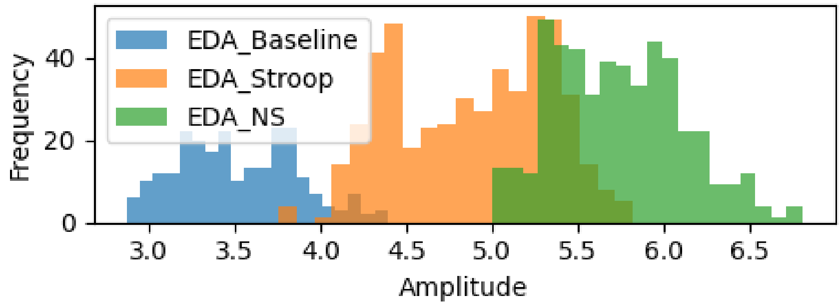

Figure 3.10 : Signal histograms.

Distributions between sessions were visualized with histogram graphics as shown in Figure 3.10. In addition, Boxplot graphs were drawn to visualize the central tendency, distribution, outliers, and comparison of different groups of the data set, seen in Figure 3.11. It is aimed to remove the noise components seen in the boxplots by performing filtering operations. In the relevant images, It is aimed to remove the noise components seen in the boxplots by performing bandpass filtering operations. In the relevant stages, some signals were filtered.

3.4.2 Feature selection

Due to their propensity to maximize classification system performance, two well-performed feature selection algorithms k−best and mRMR (minimum redundancy maximum relevance) have attracted a lot of interest in the machine learning community in recent years. While the mRMR algorithm concentrates on choosing features with

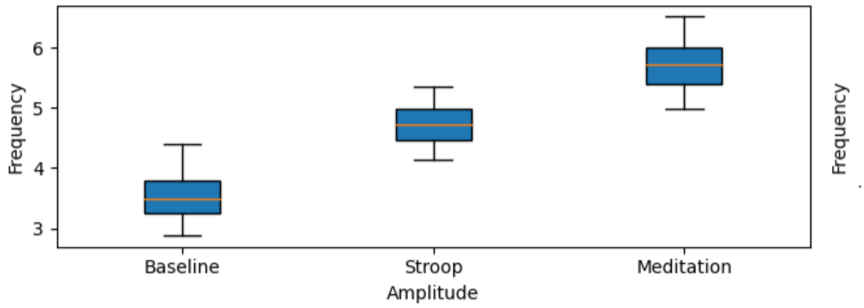

Figure 3.11 : Signal box plots.

the least amount of redundancy and the greatest relevance to the target variable, the k-best method chooses the top k most informative features from a given dataset.

Numerous studies have demonstrated the efficacy of these algorithms across various application areas, including human activity recognition, cancer classification using gene expression profiling, natural language parsing, and medical data analysis. For instance, Fang et al., 2020 [65] leveraged the mRMR algorithm to recognize human activities with high accuracy, while Alshamlan et al., 2015 [66] employed the same method for cancer classification. Similarly, Hameed et al., 2021 [67] utilized the k-best algorithm for efficient medical data analysis.

To further improve the performance of classification systems, researchers have developed hybrid feature selection algorithms that integrate the strengths of multiple methods. For example, Alshamlan et al., 2015 [66] proposed the mRMR-ABC algorithm, which combines the mRMR algorithm with an additional filter to reduce the impact of irrelevant genetic variants on cancer classification. Ünler et al., 2011 [68] developed the mr2PSO algorithm, which employs particle swarm optimization to identify the optimal combination of feature selection techniques for support vector machine classification.

The versatility and significance of these algorithms are evident in their widespread adoption across numerous domains. They have been successfully applied in healthcare and biology, natural language processing, and engineering, highlighting their potential to enhance the efficiency and accuracy of machine learning models. As the demand for advanced AI technologies continues to grow, the importance of effective feature selection algorithms will only increase, underscoring the need for continued research and innovation in this area.

Feature selection was performed using mRMR feature selection and kbest. 10 features were selected for the two methods, and EDA_max_min, EDA_coeffvar, SCR_Peaks and SCL_coeffvar features were selected as common elements. The relevant Figure 3.12 shows the correlation matrix created for 10 features selected with the mRMR algorithm. After the feature selection step, the dataset was created with labels 0, 1, and 2 representing distinct states: 0 for the baseline, 1 indicating stress conditions, and 2 signifying recovery steps. These labels categorically define the conditions under which dataset samples were collected. In machine learning, such designated labels play a crucial role in classifying or identifying these diverse states. This systematic labeling aids in both analyzing the dataset's different conditions and integrating these distinctions into models.

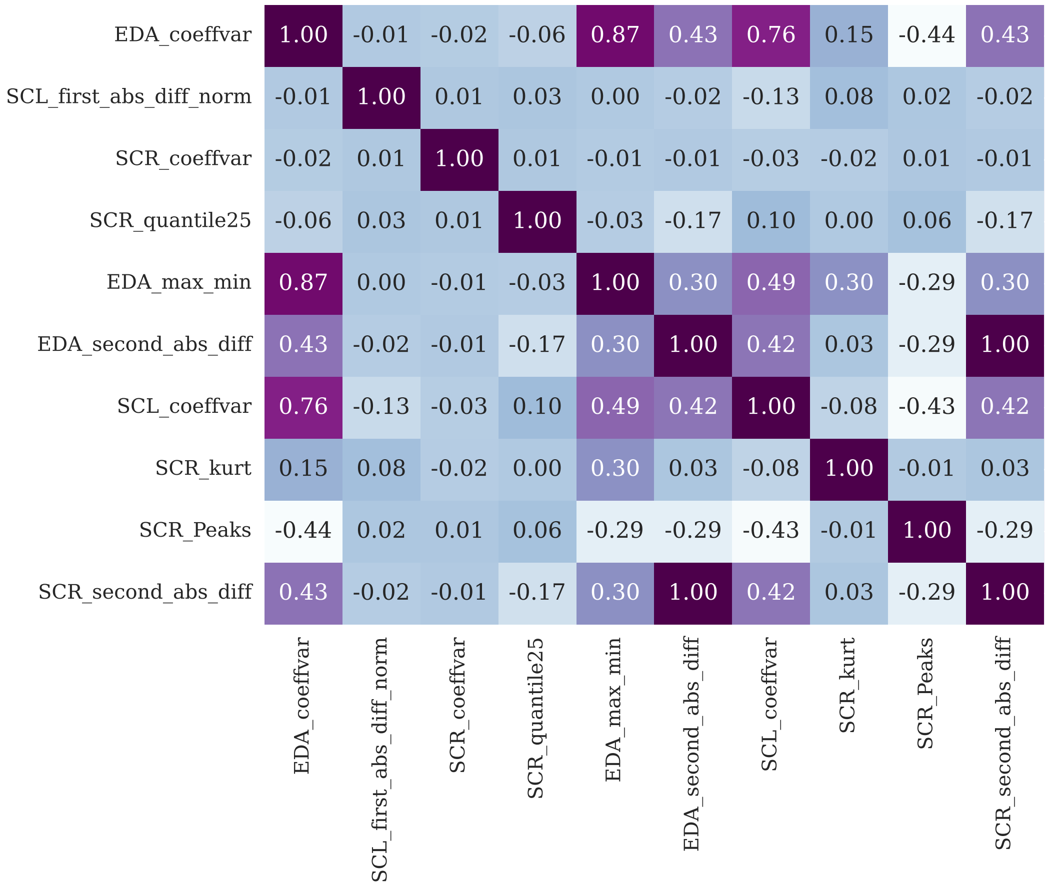

Figure 3.12 : mRMR features correlation matrix.

3.4.3 Classification

3.4.3.1 XGBoost classification

XGBoost is a versatile algorithm that is used as a classification method in many fields. In the field of artificial intelligence, XGBoost (Extreme Gradient Boosting) has emerged as the learning algorithm that has gained widespread recognition. What sets XGBoost apart from others lies in its capacity to handle large datasets, large feature spaces, and diversified applications with ease. The essence of XGBoost lies in its boosting mechanism, which involves combining a large number of weak learners to create an outstanding classifier. With the strong structure created, model errors can be eliminated and the models can iteratively reveal the relationships within the data professionally. XGBoost increases performance with gradient boosting and also creates more reliable predictions. Chen et al. described XGBoost as a boosting system and emphasized its ability to achieve good results on classification problems [69]. Gu et al. proposed the powerful machine learning algorithm XGBoost to identify the root causes of failures with exceptional accuracy [70]. They took advantage of the strengths of XGBoost by implementing quadratic Taylor expansion in the model. They stated that the proposed method achieved superior performance compared to others.

XGBoost has proven successful in various application areas. XGBoost is a powerful ensemble approach that has shown to be extremely effective in a wide range of applications, such as social analysis, defect detection, healthcare, and environmental monitoring. Its remarkable success may be ascribed to its capacity to provide dependable predictions, highlight significant features using feature significance measures, and manage class imbalance concerns with ease. Its adaptability has made it easier to integrate smoothly with other optimization methods like particle swarm optimization and genetic algorithms, which has increased its potential. XGBoost has become a standard tool for academics and practitioners alike because of its broad usage in a variety of sectors.

3.4.3.2 LSTM classification

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) designed to efficiently find and process long-term relationships in sequential data. Because of their unique feedback connections, which set them apart from traditional feedforward neural network architecture, long-term storage and retrieval of historical data is made possible for LSTM networks. This characteristic makes them especially well-suited for processing sequential data types, such as those found in time series analysis, natural language processing, physiological signal processing, and other related fields. LSTMs expand upon the basic structure of RNNs, allowing them to better handle dependencies over long time intervals. Normal RNNs tend to lose historical knowledge over time, and they might have trouble comprehending lengthy dependency processes. By including a feature known as "cell state," LSTMs enhance information storage and control mechanisms to address this issue. They are more equipped to deal with chronic addictions in this way.

LSTM has been used to the analysis of EDA and other physiological signals for stress categorization and detection in several studies. For example, Zulqarnain et al., 2023 [71] showed that an E-LSTM technique performed better in stress detection classification than other approaches, proving the usefulness of LSTM in stress classification tasks. Furthermore, Phutela et al., 2022 [72] demonstrated the application of LSTM in stress classification based on physiological signals by contrasting it with the MLP and LSTM for categorizing stress and non-stress groups.

In conclusion, LSTM has proven its strength in tasks like stress classification and fault detection, whereas XGBoost has been useful in applications like network intrusion detection and power price predictions. The particular needs and features of the issue at hand may influence which of XGBoost and LSTM is best.

3.4.4 Experimental results & discussions

Within the scope of this study, various setups were established to leverage labeled data following assessments conducted with children. Models were constructed using XGBoost and LSTM algorithms utilizing test data obtained from the Stroop test, and their accuracies were subsequently analyzed.

3.4.4.1 XGBoost classification

The 10 significant features selected by mRMR were classified using XGBoost classification. The test classification result is as shown in Table 3.8, for train shown in Table 3.7. Balanced accuracy is a measure used to evaluate the performance of the classification model on data sets with class imbalance. It was observed that the sensitivity of detecting the baseline state was higher than detecting the stress and recovery states.

Confusion matrices for train and test classification for XGBoost are shown in Figure 3.13. In test classifications, the best stress situation was determined correctly by the confusion matrix. The distinctiveness of this class is greater than others.

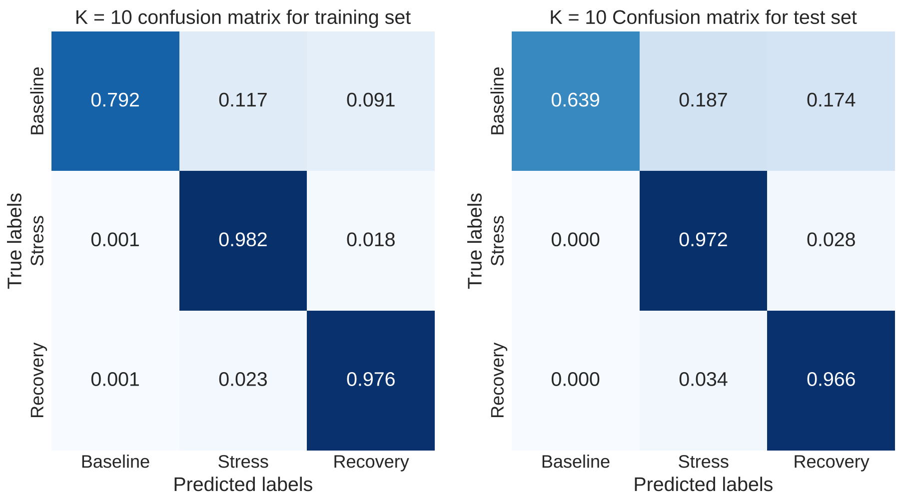

Figure 3.13 : XGBoost training and test confusion matrix.

All of the metrics are calculated after training. Figure 3.14 and 3.15, showed that, recall, precision, and roc curves. Precision measures how many positive predictions are correct. Indicates the rate of false positives. Recall measures how many true positives were predicted correctly. Indicates the rate of false negatives. Precision is about false

Table 3.7 : XGBoost train accuracy.

Classes	Precision	Recall	F1-Score
Baseline	0.9936	0.8619	0.9231
Stress	0.9645	0.9897	0.9769
Recovery	0.9600	0.9753	0.9676
Balanced Accuracy			0.9423
Accuracy			0.9662
Macro Avg.	0.9727	0.9423	0.9559
Weighted Avg.	0.9668	0.9662	0.9658

Table 3.8 : XGBoost test accuracy.

Classes	Precision	Recall	F1-Score
Baseline	0.9796	0.7111	0.8240
Stress	0.9294	0.9809	0.9545
Recovery	0.9282	0.9562	0.9420
Balanced Accuracy			0.8827
Accuracy			0.9340
Macro Avg.	0.9457	0.8827	0.9068
Weighted Avg.	0.9359	0.9340	0.9316

positives, recall is about false negatives. Ideally, both high precision and high recall are desired, but the two are often at odds with each other. That's why balance is important.

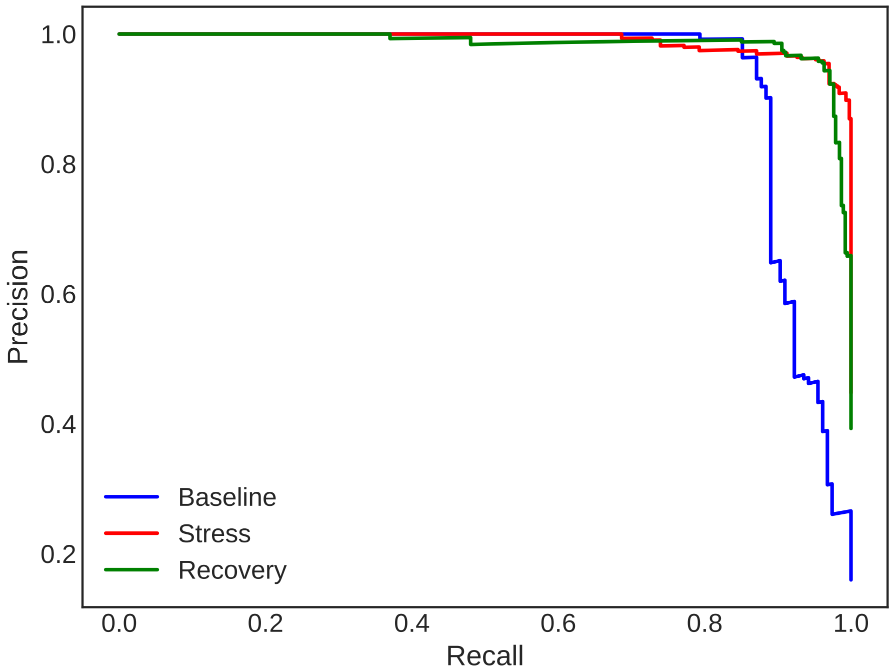

Figure 3.14 : XGBoost precision and recall curve.

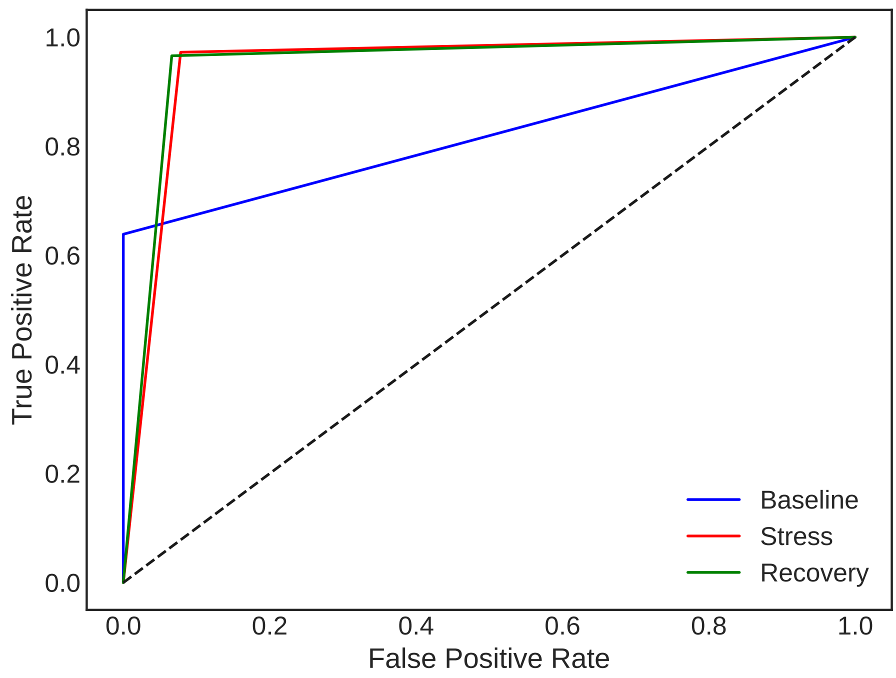

Figure 3.15 : XGBoost ROC curve.

The contribution of the features used in the model to the classification of the model was examined. For this, the SHAP (SHapley Additive Explanations) method was used. In this way, it was possible to comment on the model based on the outputs. According to the model result we created with the XGBoost algorithm, EDA_max_min is the most decisive feature among all features, shown in Figure 3.16.

In Figure 3.17, the SHAP summary chart visualizes the impact of different features on the model's predictions. Each point represents the Shapley value calculated for a specific attribute in the dataset. The Shapley values quantify the contribution of each feature to the model output. These values are computed individually for each data point, resulting in a calculation for every cell used in training, which accounts for the longer computation time associated with SHAP.

The y-axis lists the features used in training, ordered by their importance in influencing the model's output, with the most impactful features at the top. Notably, in this graph, the attribute that appears to have the most significant impact on the model's predictions is EDA_max_min.

On the x-axis, the Shapley values are plotted. Positive values indicate contributions that positively influence the prediction of the data's class, while negative values signify contributions that have a negative effect on the prediction.

Observing the graph, it's apparent that a low SCR_quantile25 value negatively impacts the model's prediction. Conversely, areas, where the relevant values are depicted as red (indicating high values), show features that have a positive effect on the prediction.

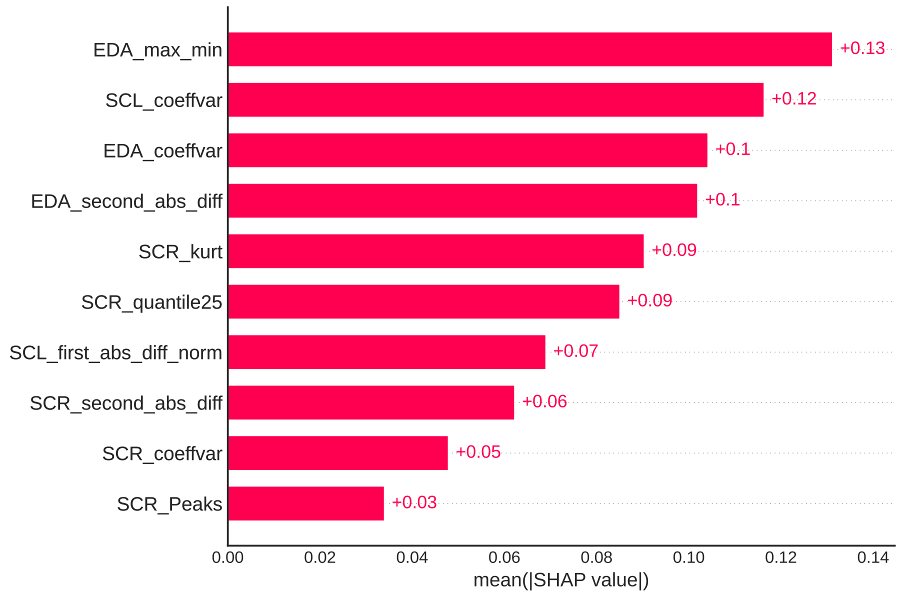

Figure 3.16 : Feature importance.

This visual representation aids in understanding how individual features contribute to the model's decisions and which attributes hold more sway over the predicted outcomes.

3.4.4.2 LSTM classification

In this part of the study, the LSTM method, which is mostly used in sequential data, was used. Features were extracted from the data and stress classification operations were performed in XGBoost. In LSTM, raw data was processed and classification operations were performed with certain window intervals. Batch normalization was used. With this method, the outputs at each layer of the network are normalized and the training process is improved. Afterwards, the Relu activation function was used. The system operates more quickly and effectively because relu activation functions, which are only active at positive values, prevent the neural network's intermediate

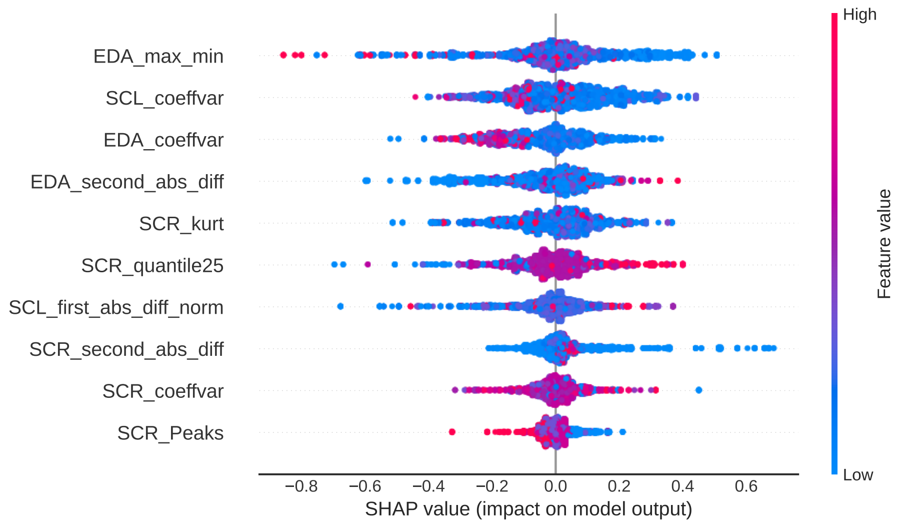

Figure 3.17 : Shap value impact graph.

layers from activating a potential negative output and, as a result, prevent the neural network from being able to activate all of its neurons at once. Dropout layer was used to prevent overfitting in the intermediate layers. Through Dropout layout, some neurons are forgotten and the model is prevented from memorizing the data. At the model output, the sigmoid activation function was used for binary classification, and the Softmax activation function was also used for multiple classification.

In Table 3.9, it is shown that the accuracy of the model is 0.73. Baseline recall and F1-scores are 0.24 and 0.34 respectively. The F1 score is a measure that evaluates the accuracy of a classification model. F1 score is the harmonic mean of precision and recall values. A low F1 score may indicate that the model is underperforming in terms of both precision and recall. This may indicate that the model produces both false positive results (low precision) and false negative results (low recall). That is, a low F1 score indicates that the model is poor or unstable at identifying given classes. In this case, steps such as better training the model, improving features, or choosing a more suitable algorithm can be considered. This situation may also be due to the lack of data for the baseline class. In our data, baseline data is less than in other classes.

The test was performed as a binary classification with the same model. The recovery data in the model was combined with baseline data and named No-Stress. Table 3.10, it is shown the result of binary classification with the LSTM model. According to the result, the accuracy is 0.77 and the metrics are more acceptable than the first test. Our model can better classify non-stress and stress situations from raw data.

Table 3.9 : LSTM multi-class test accuracy.

Classes	Precision	Recall	F1-Score
Baseline	0.62	0.24	0.34
Stress	0.79	0.76	0.78
Recovery	0.69	0.87	0.77
Accuracy			0.73
Macro Avg.	0.70	0.62	0.63
Weighted Avg.	0.73	0.73	0.71

Table 3.10 : LSTM binary-class test accuracy.

Classes	Precision	Recall	F1-Score
Non-Stress	0.74	0.87	0.80
Stress	0.82	0.66	0.73
Accuracy			0.77
Macro Avg.	0.78	0.76	0.76
Weighted Avg.	0.77	0.66	0.76

4. CONCLUSION

In the current study, unlike most of the studies in the literature involving adult test subjects, research has been conducted on stress detection in children. Due to difficulties in obtaining a large and labelled dataset,future studies will focus on expanding the child dataset and employing deep learning methods.

In the research on robot therapies in children, emotional states were analyzed and the following findings were obtained: There is a significant correlation between physiological data and stress states. It has been observed that especially in children, the EDA_max_min feature is more important than other features during stress periods. In addition, in the analyses made between ages, it was seen that stress status markers had different characteristics. It has been observed that the XGBoost classification algorithm works with a higher accuracy rate compared to other methods in stress classification in the child data we have.

The relationships between child datasets and adults will be investigated. Furthermore, binary groupings of obtained signals or determining the most influential features of each signal can facilitate classification methods. Using child datasets, the aim is to compare and contrast the relationship of signals between children and adults.

The emotional recognition feature with children can be utilized in our studies in the field of human-robot interaction. This feature can be employed to provide feedback to therapies and rsobots, enhancing their effectiveness.

REFERENCES

[1] **O'Connor, D.B.**, **Thayer, J.F. and Vedhara, K.** (2021). Stress and health: A review of psychobiological processes, *Annual review of psychology*, *72*, 663–688.

[2] **Fisher, S.E. and Reason, J.E.** (1988). *Handbook of life stress, cognition and health.*, John Wiley & Sons.

[3] **Schneiderman, N.**, **Ironson, G. and Siegel, S.D.** (2005). Stress and health: psychological, behavioral, and biological determinants, *Annu. Rev. Clin. Psychol.*, *1*, 607–628.

[4] **Howard, P.L.**, **Liversedge, S.P. and Benson, V.** (2017). Processing of co-reference in autism spectrum disorder, *Autism Research*, *10*, 1968–1980.

[5] **Mazefsky, C.**, **Herrington, J.D.**, **Siegel, M.**, **Scarpa, A.**, **Maddox, B.B.**, **Scahill, L. and White, S.** (2013). The role of emotion regulation in autism spectrum disorder, *Journal of the American Academy of Child Amp; Adolescent Psychiatry*, *52*, 679–688.

[6] **Harms, M.B.**, **Martin, A. and Wallace, G.L.** (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies, *Neuropsychology Review*, *20*, 290–322.

[7] **Nasoz, F.**, **Alvarez, K.**, **Lisetti, L. and Finkelstein, N.** (2004). Emotion Recognition From Physiological Signals Using Wireless Sensors for Presence Technologies, *Cognition Technology & Work*.

[8] **Shu, L.**, **Xie, J.**, **Yang, M.**, **Li, Z.**, **Li, Z.**, **Liao, D.**, **Xu, X. and Xinyi, Y.** (2018). A Review of Emotion Recognition Using Physiological Signals, *Sensors*.

[9] **Kim, J.H. and André, E.** (2008). Emotion Recognition Based on Physiological Changes in Music Listening, *Ieee Transactions on Pattern Analysis and Machine Intelligence*.

[10] **Song, T.**, **Zheng, W.**, **Cheng, L.**, **Zong, Y.**, **Zhang, X. and Cui, Z.** (2019). MPED: A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition, *Ieee Access*.

[11] **Ahmad, Z. and Khan, N.** (2022). A Survey on Physiological Signal-Based Emotion Recognition, *Bioengineering*.

[12] **Wei, W.**, **Jia, Q.**, **Feng, Y. and Chen, G.** (2018). Emotion Recognition Based on Weighted Fusion Strategy of Multichannel Physiological Signals, *Computational Intelligence and Neuroscience*.

[13] **Kaklauskas, A.**, **Abraham, A.**, **Ubarte, I.**, **Kliukas, R.**, **Luksaite, V.**, **Binkyte-Veliene, A.**, **Vetloviene, I. and Kaklauskiene, L.** (2022). A Review of AI Cloud and Edge Sensors, Methods, and Applications for the Recognition of Emotional, Affective and Physiological States, *Sensors*.

[14] **Park, B.J.** (2013). Design of Prototype-Based Emotion Recognizer Using Physiological Signals, *Etri Journal*.

[15] **Memar, M. and Mokaribolhassan, A.** (2021). Stress level classification using statistical analysis of skin conductance signal while driving, *SN Applied Sciences*, *3*(1), 1–9.

[16] **Shukla, J.**, **Barreda-Ángeles, M.**, **Oliver, J.**, **Nandi, G. and Puig, D.** (2019). Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity, *IEEE Transactions on Affective Computing*, *PP*, 1–1.

[17] **Hernandez, J.**, **Riobo, I.**, **Rozga, A.**, **Abowd, G. and Picard, R.** (2014). Using Electrodermal Activity to Recognize Ease of Engagement in Children during Social Interactions, *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*.

[18] **Zangroniz, R.**, **Martinez Rodrigo, A.**, **Pastor García, J.M.**, **López Bonal, M. and Fernández-Caballero, A.** (2017). Electrodermal Activity Sensor for Classification of Calm/Distress Condition, *Sensors*, *17*, 2324.

[19] **Lutin, E.**, **Hashimoto, R.**, **De Raedt, W. and Van Hoof, C.** (2021). Feature Extraction for Stress Detection in Electrodermal Activity., *BIOSIGNALS*, Vienna, Austria, pp.177–185.

[20] **Zontone, P.**, **Affanni, A.**, **Bernardini, R.**, **Piras, A. and Rinaldo, R.** (2019). Stress Detection Through Electrodermal Activity (EDA) and Electrocardiogram (ECG) Analysis in Car Drivers, *2019 27th European Signal Processing Conference (EUSIPCO)*, pp.1–5.

[21] **Liu, Y. and Siqing, D.** (2017). Psychological Stress Level Detection Based on Electrodermal Activity, *Behavioural Brain Research*, *341*.

[22] **Can, Y.**, **Chalabianloo, N.**, **Ekiz, D. and Ersoy, C.** (2019). Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study, *Sensors*, *19*.

[23] **Can, Y.**, **Gokay, D.**, **Kılıç, D.**, **Ekiz, D.**, **Chalabianloo, N. and Ersoy, C.** (2020). How Laboratory Experiments Can Be Exploited for Monitoring Stress in the Wild: A Bridge Between Laboratory and Daily Life, *Sensors*, *20*.

[24] **Kyamakya, K.**, **Al Machot, F.**, **Mosa, A.**, **Bouchachia, H.**, **Chedjou, J. and Bagula, A.** (2021). Emotion and Stress Recognition Related Sensors and Machine Learning Technologies, *Sensors*, *21*, 2273.

[25] **Nakahara, H.**, **Furuya, S.**, **Obata, S.**, **Masuko, T. and Kinoshita, H.** (2009). Emotion-related Changes in Heart Rate and Its Variability during Performance and Perception of Music, *Annals of the New York Academy of Sciences*, *1169*, 359–62.

[26] **Tonacci, A.**, **Dellabate, A.**, **Dieni, A.**, **Bachi, L.**, **Sansone, F.**, **Conte, R. and Billeci, L.** (2020). Can Machine Learning Predict Stress Reduction Based on Wearable Sensors' Data Following Relaxation at Workplace? A Pilot Study, *Processes*, *8*, 448.

[27] **Castro, J.**, **Cantero, A.**, **Gomez Gonzalez, I.M.**, **Arroyo, S. and Merino Monge, M.** (2022). Towards Human Stress and Activity Recognition: A Review and a First Approach Based on Low-Cost Wearables, *Electronics*, *11*, 155.

[28] **Chen, J.**, **Abbod, M. and Shieh, J.S.** (2021). Pain and Stress Detection Using Wearable Sensors and Devices—A Review, *Sensors*, *21*, 1030.

[29] **van Lier, H.**, **Pieterse, M.**, **Garde, A.**, **Postel, M.**, **de Haan, H.**, **Vollenbroek Hutten, M.**, **Schraagen, J.M. and Noordzij, M.** (2019). A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor, *Behavior Research Methods*, *52*.

[30] **Kyriakou, K.**, **Resch, B.**, **Sagl, G.**, **Petutschnig, A.**, **Werner, C.**, **Niederseer, D.**, **Liedlgruber, M.**, **Wilhelm, F.H.**, **Osborne, T. and Pykett, J.** (2019). Detecting moments of stress from measurements of wearable physiological sensors, *Sensors*, *19*(17), 3805.

[31] **Kleckner, I.**, **Jones, R.**, **Wilder-Smith, O.**, **Wormwood, J.**, **Akcakaya, M.**, **Quigley, K.**, **Lord, C. and Goodwin, M.** (2017). Simple, Transparent, and Flexible Automated Quality Assessment Procedures for Ambulatory Electrodermal Activity Data, *IEEE Transactions on Biomedical Engineering*, *PP*, 1–1.

[32] **Chen, W.**, **Jaques, N.**, **Taylor, S.**, **Sano, A.**, **Fedor, S. and Picard, R.** (2015). Wavelet-based motion artifact removal for electrodermal activity.

[33] **Schmidt, P.**, **Reiss, A.**, **Duerichen, R. and Van Laerhoven, K.** (2018). Wearable affect and stress recognition: A review.

[34] **Redd, C.B.**, **Silvera-Tawil, D.**, **Hopp, D.**, **Zandberg, D.**, **Martiniuk, A.**, **Dietrich, C. and Karunanithi, M.K.** (2020). Physiological Signal Monitoring for Identification of Emotional Dysregulation in Children, *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, pp.4273–4277.

[35] **Ollander, S.**, **Godin, C.**, **Campagne, A. and Charbonnier, S.** (2016). A comparison of wearable and stationary sensors for stress detection, pp.004362–004366.

[36] **Zhang, H.** (2020). Expression-Eeg Based Collaborative Multimodal Emotion Recognition Using Deep AutoEncoder, *Ieee Access*.

[37] **Chun-yan, N. and Wang, H.** (2022). Correlation Study Between Chaos Features of Physiological Signals and Emotions Based on the Information Gain Method, *Journal of Physics Conference Series*.

[38] **Rahman, M.M.**, **Usman, O.L.**, **Muniyandi, R.C.**, **Sahran, S.**, **Mohamed, S. and Razak, R.A.** (2020). A Review of Machine Learning Methods of Feature Selection and Classification for Autism Spectrum Disorder, *Brain Sciences*.

[39] **Wang, H. and Doering, L.C.** (2015). Autism Spectrum Disorders: Emerging Mechanisms and Mechanism-Based Treatment, *Frontiers in Cellular Neuroscience*.

[40] **Hashim, H.U.**, **Yunus, M.M. and Norman, H.** (2021). English as Secondary Language Learning and Autism Spectrum Disorder: The Obstacles in Teaching and Learning the Language, *Arab World English Journal*.

[41] **Callenmark, B.**, **Kjellin, L.**, **Rönnqvist, L. and Bölte, S.** (2013). Explicit Versus Implicit Social Cognition Testing in Autism Spectrum Disorder, *Autism*.

[42] **Viviers, M.M.**, **Jongh, M.d.**, **Dickonson, L.**, **Malan, R. and Pike, T.** (2020). Parent-Reported Feeding and Swallowing Difficulties of Children With Autism Spectrum Disorders (Aged 3 to 5 Years) Compared to Typically Developing Peers: A South African Study, *African Health Sciences*.

[43] **MacMullin, J.A.**, **Lunsky, Y. and Weiss, J.A.** (2015). Plugged In: Electronics Use in Youth and Young Adults With Autism Spectrum Disorder, *Autism*.

[44] (2023). `https://www.empatica.com/`.

[45] **Berkoff, D.**, **Cairns, C.**, **Sanchez, L. and Moorman, C.** (2007). Heart rate variability in elite American track and Field athletes, *J. Strength Cond. Res.*, *21*, 227–31.

[46] **Seppälä, S.**, **Laitinen, T.**, **Tarvainen, M.**, **Tompuri, T.**, **Veijalainen, A.**, **Savonen, K. and Lakka, T.** (2013). Normal values for heart rate variability parameters in children 6-8 years of age: The PANIC Study, *Clinical physiology and functional imaging*, *34*.

[47] **Silvetti, M.**, **Drago, F. and Ragonese, P.** (2002). Heart rate variability in healthy children and adolescents is partially related to age and gender, *International journal of cardiology*, *81*, 169–74.

[48] **Karabulut, M.** (2015). Sağlıklı Çocuklarda kalp hızı değişkenliği, *Fırat Medical Journal*, *20*, 152–55.

[49] **Fazli, C.** (2019). Pediatric heart rate variability normative values related to average heart rate and age in a developing Country, *Online Journal of Cardiovascular Research*, *2*.

[50] **Tiinanen, S.**, **Mättä, A.**, **Silfverhuth, M.**, **Suominen, K.**, **Jansson-Verkasalo, E. and Seppänen, T.** (2011). HRV and EEG based indicators of stress in children with Asperger syndrome in audio-visual stimulus test, *2011 Annual Conf. Proc. IEEE Eng. Med. Biol. Soc.*, IEEE, pp.2021–2024.

[51] **Bujnakova, I.**, **Ondrejka, I.**, **Mestanik, M.**, **Visnovcova, Z.**, **Mestanikova, A.**, **Hrtanek, I.**, **Fleskova, D.**, **Calkovska, A. and Tonhajzerova, I.** (2016). Autism spectrum disorder is associated with autonomic underarousal., *Physiological research*, *65*.

[52] **van Lier, H.**, **Pieterse, M.**, **Garde, A.**, **Postel, M.**, **de Haan, H.**, **Vollenbroek Hutten, M.**, **Schraagen, J.M. and Noordzij, M.** (2019). A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor, *Behavior Research Methods*, *52*.

[53] **Braithwaite, J.J.**, **Watson, D.P.Z.**, **Jones, R.O. and Rowe, M.A.** (2013). Guide for Analysing Electrodermal Activity & Skin Conductance Responses for Psychological Experiments, *CTIT technical reports series*.

[54] **Taylor, S.**, **Jaques, N.**, **Chen, W.**, **Fedor, S.**, **Sano, A. and Picard, R.** (2015). Automatic identification of artifacts in electrodermal activity data, *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp.1934–1937.

[55] **Lab., M.** (2015). *EDA Explorer*, `https://eda-explorer.media.mit.edu/`, accessed: 2022-02-23.

[56] **Posada-Quintero, H. and Chon, K.** (2020). Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review, *Sensors*, *20*, 479.

[57] **Shakiba, N.**, **Ellis, B.J.**, **Bush, N.R. and Boyce, W.T.** (2019). Biological sensitivity to context: a test of the hypothesized u-shaped relation between early adversity and stress responsivity, *Development and Psychopathology*, *32*, 641–660.

[58] **Merz, C.J.**, **Eichholtz, A. and Wolf, O.T.** (2020). Acute stress reduces out-group related safety signaling during fear reinstatement in women, *Scientific Reports*, *10*.

[59] **Chen, J.**, **He, X.**, **Shen, Y.**, **Zhou, Q.**, **Cui, C. and Liu, Y.** (2022). Demographic differences in public acceptance of waste-to-energy incinerators in china:

high perceived stress group vs. low perceived stress group, *Frontiers in Psychology*, *13*.

[60] **Liu, J.J.W.**, **Vickers, K.S.**, **Reed, M.J. and Hadad, M.** (2017). Re-conceptualizing stress: shifting views on the consequences of stress and its effects on stress reactivity, *Plos One*, *12*, e0173188.

[61] **Mika, A.**, **Mazur, G.J.**, **Hoffman, A.N.**, **Talboom, J.S.**, **Bimonte-Nelson, H.A.**, **Sanabria, F. and Conrad, C.D.** (2012). Chronic stress impairs prefrontal cortex-dependent response inhibition and spatial working memory., *Behavioral Neuroscience*, *126*, 605–619.

[62] **Kyun, K.T.** (2015). T test as a parametric statistic, *kja*, *68*(6), 540–546, `http://www.e-sciencecentral.org/articles/?scid=1156170`, `http://www.e-sciencecentral.org/articles/?scid=1156170`.

[63] **Aktaş, S.N.B.**, **Uluer, P.**, **Coşkun, B.**, **Toprak, E.**, **Barkana, D.E.**, **Köse, H.**, **Zorcec, T.**, **Robins, B. and Landowska, A.** (2022). Stress Detection of Children With ASD Using Physiological Signals, *2022 30th Signal Processing and Communications Applications Conference (SIU)*, pp.1–4.

[64] **Toprak, E.**, **Bilgin Aktas, S.N.**, **Coşkun, B.**, **Uluer, P.**, **Kose, H. and Barkana, D.E.** (2023). Investigation of Physiological Features by Age Groups in Children with Autism, *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pp.287–292.

[65] **Fang, H.**, **Tang, P. and Si, H.** (2020). Feature Selections Using Minimal Redundancy Maximal Relevance Algorithm for Human Activity Recognition in Smart Home Environments, *Journal of Healthcare Engineering*.

[66] **Alshamlan, H.**, **Badr, G. and Al-Ohali, Y.** (2015). mRMR-ABC: A Hybrid Gene Selection Algorithm for Cancer Classification Using Microarray Gene Expression Profiling, *Biomed Research International*.

[67] **Zeeshan Hameed, B.M.**, **Shah, M.**, **Naik, N.**, **Khanuja, H.S.**, **Paul, R. and Somani, B.** (2021). Application of Artificial Intelligence-Based Classifiers to Predict the Outcome Measures and Stone-Free Status Following Percutaneous Nephrolithotomy for Staghorn Calculi: Cross-Validation of Data and Estimation of Accuracy, *Journal of Endourology*.

[68] **Ünler, A.**, **Murat, A. and Chinnam, R.B.** (2011). Mr2PSO: A Maximum Relevance Minimum Redundancy Feature Selection Method Based on Swarm Intelligence for Support Vector Machine Classification, *Information Sciences*.

[69] **Chen, T. and Guestrin, C.** (2016). Xgboost: a scalable tree boosting system.

[70] **Gu, K.**, **Wang, J.**, **Qian, H. and Su, X.** (2021). Study on intelligent diagnosis of rotor fault causes with the pso-xgboost algorithm, *Mathematical Problems in Engineering*, *2021*, 1–17.

[71] **Zulqarnain, M.**, **Shah, H.**, **Ghazali, R.**, **Alqahtani, O.**, **Sheikh, R. and Asadullah, M.** (2023). Attention aware deep learning approaches for an efficient stress classification model, *Brain Sciences*, *13*, 994.

[72] **Phutela, N.**, **Relan, D.**, **Gabrani, G.**, **Kumaraguru, P. and Samuel, M.** (2022). Stress classification using brain signals based on lstm network, *Computational Intelligence and Neuroscience*, *2022*, 1–13.

CURRICULUM VITAE

Name Surname : Sevgi Nur BİLGİN AKTAŞ

EDUCATION :

- **M.Sc.** : 2024, Istanbul Technical University, Institute of Graduate Studies, Department of Computer Engineering
- **B.Sc.** : 2018, Fatih Sultan Mehmet Foundation University, Faculty of Engineering and Technology, Department of Biomedical Engineering
- **B.Sc.** : 2018, Fatih Sultan Mehmet Foundation University, Faculty of Engineering and Technology, Department of Computer Engineering

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Aktaş, S. N. B.**, Uluer, P., Coşkun, B., Toprak, E., Barkana, D. E., Köse, H., Zorcec, T., Robins, B., & Landowska, A. (2022, May). Stress Detection of Children With ASD Using Physiological Signals. *Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. (Presentation Instance)

- Toprak, E. and **Aktaş, S. N. B.** and Coşkun, B. and Uluer, P. and Kose, H. and Barkana, D. E. (2023). Investigation of Physiological Features by Age Groups in Children with Autism. *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 287–292). IEEE. (Presentation Instance)