A WEB-BASED TOOLKIT FOR RECORDING AND ANALYSIS OF PHYSIOLOGICAL SIGNALS IN RESPONSE TO AFFECTIVE AND COGNITIVE STRESSORS

(DUYGUSAL VE BILIŞSEL STRES ETKENLERINE KARŞI OLUŞAN FIZYOLOJIK İŞARETLERIN KAYIT VE ANALIZI IÇIN WEB TABANLI BIR ARAÇ TAKIMI)

by

KEMAL BEŞKARDEŞLER, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Nov 2019

This is to certify that the thesis entitled

A WEB-BASED TOOLKIT FOR RECORDING AND ANALYSIS OF PHYSIOLOGICAL SIGNALS IN RESPONSE TO AFFECTIVE AND COGNITIVE STRESSORS

prepared by **KEMAL BEŞKARDEŞLER** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department** at the **Galatasaray University** is approved by the

Examining Committee:

Assist. Prof. Dr. R. Burak ARSLAN
Supervisor, **Computer Engineering Department, Galatasaray University**

Assoc. Prof. Dr. Özlem DURMAZ İNCEL
Computer Engineering Department, Galatasaray University

Assist. Prof. Dr. Berk GÖKBERK
Computer Engineering Department, MEF University

Date:

ACKNOWLEDGEMENTS

I would like to thank my friends for their support with the experiments and especially to my wife and family for their support throughout the work.

TABLE OF CONTENTS

LIST OF SYMBOLS

ACC	Accelerometer
ANN	Artificial neural network
API	Application program interface
BLE	Bluetooth low energy
BP	Blood pressure
BVP	Blood volume pulse
CPT	Cold pressor test
CWT	(Stroop) Color word test
ECG	Electrocardiogram
EEG	Electroencephalography
EDA	Electrodermal activity
EMG	Electromyogram
F	Female
FN	False negative
FP	False positive
GSR	Galvanic skin response
HF	High frequency
HMM	Hidden Markov model
HR	Heart rate
HRV	Heart rate variability
Hz	Hertz
IAPS	International affective picture system
IBI	Inter-beat interval
KNN	k-Nearest neighbour
LDA	Linear discriminant analysis
LF	Low frequency
LR	Logistic regression
M	Male
MAST	Mental arithmetic stress test
MIST	Montreal imaging stress task
MS	Milisecond
NB	Naive Bayes classifier
PASA	Primary appraisal secondary appraisal
PD	Pupil diameter
PNN	Probabilistic neural network
PPG	Photoplethysmogram
RANN	Recurrent artificial neural network
REST	Representational State Transfer
RMS	Root mean square
SC	Skin conductance

SCL	Skin conductance level
SCR	Skin conductance response
SD	Standard deviation
SQD	Squared deviation
SECPT	Socially evaluated cold pressor test
ST	Skin temperature
SVM	Support vector machine
TN	True negative
TP	True positive
TSST	Trier social stress test
uS	Microsiemens
VAS	Visual analogue scale
VR	Virtual reality

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Stress is one of most important problems of humanity related to several disorders like depression, obesity, cardiovascular and diabetic diseases. We may understand what stress changes in human physiology by monitoring physiological signals in human body and may prevent these disorders. Physiological signals can be monitored by electrodermal activity (EDA), electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG), photoplethysmography (PPG) and analysis of changes in the human body like temperature, respiration, iris and pupil parameters.

In this study, analysis of existing studies regarding stress detection using physiological signals is performed. For the purpose of producing experimental dataset and comparison with existing results, a website to perform stroop color word (CWT), image and calculation tests are developed as controlled environments. Dataset for analyses are collected with Empatica E3 wristband having EDA, temperature, photoplethysmograph and 3-axis accelerometer sensors, and Bitalino kit with EDA and ECG sensors.

Analyses are carried out on test type basis with various classification algorithms and binary/trio classifications of rest, stressed, non-stressed, low-stress and high-stress phases are studied. Best results for stroop test are achieved with 83.09% accuracy for "rest vs stressed" classification. 89.61%, 88.35%, 85.78% and 77.46% accuracies are achieved for symbol digit modalities & math test for "rest vs stressed", "rest vs low stress", "rest vs high stress" and "low vs high stress" classifications, respectively. 85.06% accuracy is achieved in image tests' "rest vs stressed" analyses.

Keywords : Stress Test, Physiological Signals, Electrodermal Activity, EDA, ECG

RÉSUMÉ

Le stress est l'un des problèmes les plus importants de l'humanité lié à plusieurs troubles tels que la dépression, l'obésité, les maladies cardiovasculaires et diabétiques. Nous pouvons comprendre les effets du stress sur la physiologie humaine en surveillant les signaux physiologiques dans le corps humain; et peut prévenir ces troubles. Les signaux physiologiques peuvent être surveillés par activité électrodermale (EDA), électrocardiographie (ECG), électroencéphalographie (EEG), électromyographie (EMG), photopléthysmographie (PPG) et analyse des modifications du corps humain telles que la température, la respiration, les paramètres de la pupille.

Dans cette étude, l'analyse des études existantes concernant la détection de stress à l'aide de signaux physiologiques est réalisée. Afin de produire un jeu de données expérimental et de comparer les résultats existants, un site Web permettant de réaliser des tests de mots de couleur en coup (CWT), des tests d'image et des tests de calcul est développé en tant qu'environnements contrôlés. Les ensembles de données pour les analyses sont collectés avec le bracelet Empatica E3 avec capteurs EDA, de température, de photopléthysmographe et d'accéléromètre à 3 axes, et le kit Bitalino avec capteurs EDA et ECG.

Les analyses sont effectuées sur une base de test avec différents algorithmes de classification; et les classifications binaires / trio des phases de repos, stressée, non stressée, stress faible et stress élevé sont étudiées. Les meilleurs résultats pour le test stroop sont obtenus avec une précision de 83.09% pour la classification "repos vs stressé". Des précisions de 89.61%, 88.35%, 85.78% et 77.46% sont obtenues pour les modalités du nombre de symboles et test mathématique pour les classifications "repos vs stressé", "repos vs stress faible", "repos vs stress élevé" et "stress faible ou élevé", respectivement. Une précision de 85,06% est obtenue dans les analyses "repos vs stressé" des tests d'image.

Mots Clés : Test de Stress, Signaux physiologiques, Activité Électrodermale, EDA,ECG

ÖZET

Stres; depresyon, obezite, kardiyovasküler ve diyabetik hastalıklar gibi çeşitli hastalıklarla ilişkilendirilen insanlığın en önemli sorunlarından biridir. İnsan vücudundaki fizyolojik sinyalleri izleyerek, stresin insan fizyolojisinde neleri değiştirdiğini anlayabilir ve bu hastalıkları önleyebiliriz. Fizyolojik sinyaller elektrodermal aktivite (EDA), elektrokardiyografi (EKG), elektroensefalografi (EEG), elektromiyografi (EMG), fotopletismografi (PPG) ve insan vücudundaki sıcaklık, solunum, iris ve gözbebeği parametreleri gibi değişikliklerin analizi ile izlenebilir.

Bu çalışmada, fizyolojik sinyaller kullanılarak stres tespiti ile ilgili mevcut çalışmaların analizi yapılmıştır. Deneysel veri kümesi üretmek ve var olan sonuçlarla karşılaştırmak amacıyla, stroop renk kelime testi (CWT), resim testi ve hesaplama testlerinin gerçekleştirildiği bir web sitesi kontrollü test ortamı olarak geliştirilmiştir. Analizler için gerekli veriseti EDA, sıcaklık, fotopletismograf ve 3 eksenli ivmeölçer sensörlerine sahip Empatica E3 bilekliği ve EDA ve EKG sensörlerine sahip Bitalino kiti ile toplanmıştır.

Analizler test tipi bazında, çeşitli sınıflandırma algoritmaları kullanılarak yapılmış; dinlenme, stresli, stressiz, düşük stresli ve yüksek stresli fazlarının ikili ve üçlü sınıflandırılmaları üzerinde çalışılmıştır. Stroop testi için en iyi sonuç %83.09 doğrulukla "dinlenme - stresli" sınıflandırması için elde edilmiştir. Sembol rakam yöntemi ve matematik testi için "dinlenme - stresli", "dinlenme - düşük stresli", "dinlenme - yüksek stresli" ve "düşük stresli - yüksek stresli" sınıflandırmalarında sırasıyla %89.61, %88.35, %85.78 ve %77.46 doğruluk elde edilmiştir. Resim testlerinin "dinlenme - stresli" analizlerinde %85.06 doğruluk elde edilmiştir.

Anahtar Kelimeler : Stres Testi, Fizyolojik Sinyaller, Elektrodermal Aktivite, EDA, EKG

1 INTRODUCTION

1.1 Motivation

Many people suffer from several short-term or long-term disorders caused by stress, like depression, irritability, anxiety, obesity, memory and concentration problems, mood swings, cardiovascular and diabetic diseases. These disorders result in low life quality, social problems, drop in success and economical expenses to. For example, if you have health problems, you will have life quality. You may have social problems with your friends, coworkers or students if you have irritability for example. Or, concentration problems may reduce your success at school or work. And you will try to recover these disorders with medications or therapies which will come with economic expenses to you. It has also economic expenses for country if medicines are imported. Or, for companies, efficiency of the workers can decline with stress which means economic loss for the company. However, prevention from stress would increase life quality and improve efficiency and success in life. In order to prevent stress, one must be aware of it and here, stress detection techniques come into play. With the advent of non-invasive devices that can monitor physiological signals of human body, it became easier to collect data to be analyzed which can be used to detect stress. While collecting data became an easy task, it is not so for detecting stress yet. Because of differences between hardware and quality of devices, and the environments the users in, it is still hard to define a formula to detect stress with these data. Analysis of data collected with different devices on similar conditions would provide the aspect of hardware effects and minimize environmental effects on stress detection.

1.2 Purpose

The objective of this study is to analyze existing studies regarding stress detection in terms of the efficiency and the accuracy of the techniques, sensors and parameters used and compare the results with those tests in our controlled environment including stroop, image and calculation tests. The developed platform can be easily reached online and used by non-developers such as psychologists, which would bring different approaches in application of tests and analyses of dataset. Also, the created dataset would enable comparisons with existing studies to understand stress better. The platform can be a promising platform for future studies which can be extended with new test types by modifying or contributing to the open source project.

1.3 Thesis Outline

Chapter 2 contains analysis of existing studies regarding stressors used in test environments, used sensors, extracted parameters from the sensors, methods used to differentiate stress from no stress condition and discussion of literature review.

Chapter 3 presents the test environment to perform tests online, Empatica E3 and Bitalino devices used to collect physiological data, the Android application used to control devices and the data server used to aggregate and store dataset.

Chapter 4 contains summaries of experiments executed, describes the analyses performed and evaluates the results derived from these analyses.

Chapter 5 presents conclusions, a summary of stress detection results from literature and our experiments, contributions and future work.

2 LITERATURE REVIEW

Previous works regarding stress monitoring and detection includes both tests in controlled environments and real-life scenarios. While some test environments are fully controlled environments like stroop test, math test, memorization test, simulated driving, sleeping ; some tests are conducted in life-like or real-life environments such as Trier Social Stress Test (TSST), during student exam, working in call center or driving bus. Also, varying sensors and analysis methods are used for stress detection in these environments. Electrodermal activity (EDA), electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG), photoplethysmography (PPG), accelerometer (ACC), temperature, respiration sensors are used in reviewed studies. Parameters derived from raw sensor data such as heart rate (HR), heart rate variability (HRV), blood volume pulse (BVP) are other sources used as stress indicators in analyses. Also, audio and video recordings are used as source of physical parameters like voice characteristics, eye gaze, blinking, pupil diameter (PD), face movement, head movements, levels of mouth openness, which makes measurement of stress unobtrusively. Other than physical parameters, response times and error rates calculated for tests give a sign for stress. More details about stressors, sensors, analyses and results are described in the following sections.

2.1 Stressors

Stressors used in reviewed studies are explained in this section and summarized in Table **2.1**. There are several ways to trigger stress in laboratory environments as can be seen from the table.

Stroop Color Word Task (CWT) is a well known test technique used in psychological experiments. In this test, the subjects are asked to tell the color or the pronunciation of a color word, in which mismatch between the ink color and word triggers stress.

Trier Social Stress Test (TSST) is protocol that the subjects are asks to perform a free speech about given subject (i.e. reading text loudly, simulating a job interview) followed by a mental arithmetic task in front of audience (Kirschbaum et al., 1993b).

There is not a single way of conducting Mental Arithmetic Stress Test (MAST) but it is a common way to perform test by counting down by 7 or 13 from a given value. In order to raise stress of the subjects, they may be told that it was easy for other participants or asked to start over from scratch for every wrong answer.

D2 is a test of attention in which the subjects are asks to find all 'd(2 dashes)p' patterns in a row of 47 characters consisting of d or p letters and 1 to 4 dashes in between. The whole test contains 14 lines of such rows. While it is mostly related with attention, it can be stress inducer with time pressure.

In Cold Pressor Test (CPT), the subject is asked to immerse his/her hand in to an ice water for a minute, which affects cardiovascular system that can be monitored by the changes in heart rate and blood pressure.

The difference between CPT and SECPT is that in SECPT, along with physiological stress, psychological stress of the subjects is also monitored (i.e. By informing participants that they will be filmed in order to evaluate their facial expressions, and are asked to look into the camera in (Skoluda et al., 2015).)

In verbal test in (Wilson et al., 2014), a set of verbal analogies are used in tests with acceptance, mindfulness, suppression, endurance categories selected for each participant, along with a control group, where the user is asked to select right choice.

In (Riera et al., 2012), a faked an unexpected blood sample extraction performed by a (false) male nurse at the end of the test to increase the stress level of the subjects.

In (Skoluda et al., 2015), the individual is said to prepare a talk about a certain topic in order to give a speech in front of a recording camera like in TSST test.

In (Jentsch et al., 2014), as a math task, the user has to indicate his choice on the correctness of the arithmetic operations and as an audio task, the user has to indicate the alphabetic precedence between the current letter and the previous one.

Table **2.1**: Stressors used in previous studies

Stressor	Source
CWT	(Riera et al., 2012), (Skoluda et al., 2015), (Tulen et al., 1989), (Zhai and Barreto, 2006b), (Frank et al., 2013), (Zhang et al., 2014), (Zhai and Barreto, 2006a), (Ren et al., 2013), (Jing Zhai et al., 2005)
TSST	(Chen et al., 2014), (Kirschbaum et al., 1993a), (Riera et al., 2012), (R Saslow et al., 2013), (Shi et al., 2010), (Skoluda et al., 2015), (Ollander, 2015)
MAST	(Hassellund et al., 2010), (Setz et al., 2010), (McDuff et al., 2014), (Flaa et al., 2006), (Shi et al., 2010), (Wijsman et al., 2011), (Yamaoka, 2014), (Zhang et al., 2014), (Wenhui Liao et al., 2005)
CPT	(Hassellund et al., 2010), (Flaa et al., 2006), (Hassellund et al., 2010), (Skoluda et al., 2015)
SECPT	(Minkley et al., 2014), (Jentsch et al., 2014), (Schwabe et al., 2008), (Ollander, 2015)
Physical	(Frank et al., 2013)
D2	(Wassenberg et al., 2008), (Ollander, 2015)
Verbal	(Wilson et al., 2014), (Wenhui Liao et al., 2005)
Video	(Holzinger et al., 2013)
Blood sample	(Riera et al., 2012)
Talk preparation	(de Santos Sierra et al., 2011)

2.2 Sensors and Extracted Parameters

Physiological signals can be monitored by electrodermal activity (EDA), electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG), photoplethysmography (PPG) and analysis of changes in the human body like temperature, respiration, iris and pupil parameters. EDA is changes in electrical properties of skin which is generated by sweat glads as response to autonomic activities. ECG is measurement of electrical activity of heart which is generally used for diagnosing heart related diseases. EEG is measurement of electrical activity of brain generated by neurons. Similarly, EMG is measurement of electrical activity produced by muscles. Different from these electrical measurement, PPG measures blood volume changes using optical techniques. Using these measurements, several parameters can be extracted and be used in identifying stress. Also, feature extraction from raw sensor signals makes it easier to analyze data and makes it possible for generalized predictions. Table **2.2** summarizes parameters that can be extracted from sensors.

Table **2.2**: Sensors and Parameters

Source	Extractable Parameters
EDA	Peak height, slope, rise time, energy, mean level, deviation, SQD
ECG	HR, HRV, BVP, LF power, HF power, LF/HF power ratio, deviation, SD, SQD, "P, Q, R, S" waves
EEG	alpha asymmetry, alpha/beta ratio
EMG	RMS, static/median/peak load
PPG	BVP, BP
Thermometer	ST (mean, deviation, SQD)
Respiration	Rate, volume, mean period
Physical measures	Gestures, eye gaze, pupil diameter, blink rate, voice

2.3 Stress vs No Stress

In order to make stress measurements relative to a relaxed state, different techniques are applied in reviewed studies, including deep breathing, listening to music, slowing down test, resting (Sharma and Gedeon, 2012) (Salai et al., 2016). For example, in (Zhang et al., 2014) the relaxed state is obtained by resting 5 minutes before test and resting 10 minutes after test. A different technique used in (Frank et al., 2013) was recording physical load (i.e. running, walking, sitting) before performing CWT test, to compare and differentiate mental stress from physical stress arose by physical activity. In some studies, stress levels of subjects are also collected from the subjects themselves by questionnaires in order to compare and verify the results with the feedbacks. For example, in (Skoluda et al., 2015) Primary Appraisal Secondary Appraisal (PASA) and Visual Analogue Scale (VAS) measures were used to get stress level feedbacks before, during and after tests. Here, primary stress appraisal represents threat and challenge feelings; and secondary appraisal represents self concept of own abilities and control expectancy of the subject. While PASA provides 4 scales to choose from, VAS provides an analog range (i.e. 10 cm line) to indicate applicability of the questionnaire.

2.4 Classifiers

Parameters extracted from raw sensor data (detailed in section 2.2) may have information about stress. But not all parameters provide correct sign for stress. So, they should be studied to find the best matching parameter or set of parameters that define stress. The problem of detecting stress and no stress state, or stress level estimation from sensor data is tried to be solved by classifiers.

Common methods used for classification, ranked by use to model stress, from top to bottom, according to (Sharma and Gedeon, 2012) are support vector machine (SVM), recurrent artificial neural network (RANN) (which is an ANN that contains feedback connections), adaptive neuro-fuzzy system, ANN, hidden Markov model (HMM), decision tree, Naive Bayesian network and fuzzy clustering (a hybrid of fuzzy and clustering techniques). Along with these methods, k-nearest neighbour (KNN) and probabilistic neural network (PNN) classifiers are other methods used in reviewed studies (Palanisamy et al., 2013) and (de Santos Sierra et al., 2011).

2.5 Conclusions from Literature

With different techniques used to detect stress, like analyzing different signal parameters or using different algorithms, different accuracies are achieved. For example, by using HRV parameter, 93.75% accuracy is achieved in (Palanisamy et al., 2013) with KNN algorithm, while 90% accuracy is achieved in (Costin et al., 2012) with Minimum Distance Classifier algorithm. Or, by using EDA, 82.8% accuracy is achieved in (Setz et al., 2010) with LDA algorithm, while it was 91% in (Zubair et al., 2015) with LR algorithm. Likewise different accuracies are achieved using combinations of parameters. Besides the differences in used techniques, different datasets used in analyses, that are collected with different devices with different sensors and in various test environments, also make difference in derived results. Also person based analyses give better results compared to subject independent analyses. Because of these differences, it is difficult to say that one result would apply to other exactly.

3 MATERIALS AND METHODS

It is an important problem that conclusions made according to person-dependent tests needs to be verified. In order to improve suggested works, they should be further analyzed, and further tests should be performed with similar ways so as to compare the results for better conclusions. The developed set of controlled environments in this paper would give this chance to perform stress triggering tests in a common environment, so that varying analyses with varying devices may be compared and result in better conclusions.

In this section, the developed environment consisting of a web application, devices, android application and data server is explained. The web application provides stress tests; the android application controls Bitalino device and gathers signals through Bluetooth low energy (BLE) and sends them to the data server. The data server is where sensor data and test metadata are collected and saved for future analyses. The general architecture of the system is shown in Figure **3.1**. With Bitalino setup, tests

are started simultaneously using websocket connection between the Android application and the web application, but since the built in recording ability of E3 device is used, tests are started manually from web user interface when using E3 and the synchronization of the record with test metadata is achieved using accelerometer data which is made obvious by shaking hand when starting test.

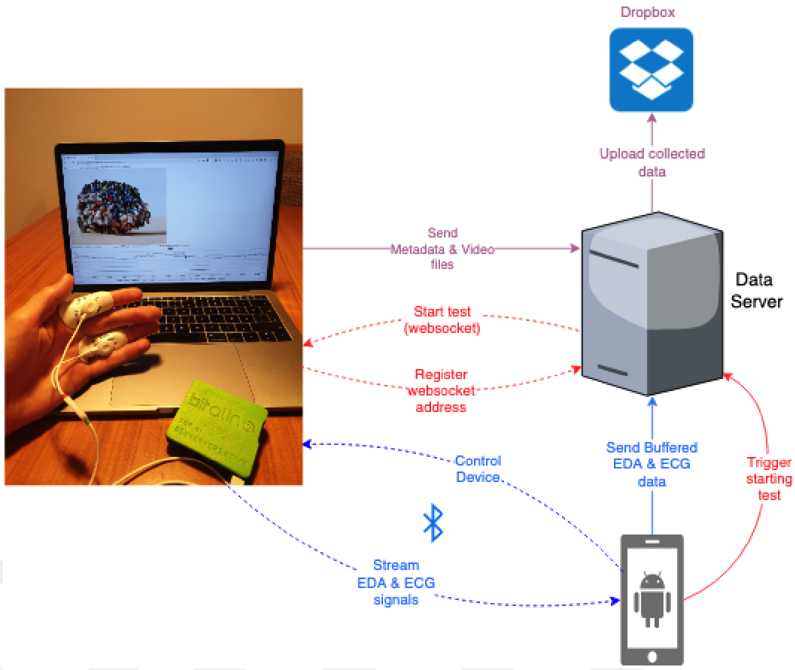

Figure **3.1**: General architecture of the system

3.1 Web Application

The developed web application performs tests and visualizes recorded physiological signals. Stroop, image, word & math, SDMT & math tests and an experimental image test in virtual reality (VR) environment are developed as stressors.

In the entry screen, the test choices are listed (Figure **3.2**) and age and sex information are obtained from the subject using the settings menu (Figure **3.3**). A predefined amount of waiting time is reserved at the beginning and the user is guided to close his/her eyes until a warning voice is heard in order to collect a reference relaxed state data and start the test in a relaxed state.

Combination of tests are also used in reviewed studies such as stroop color word test done for two minutes and subsequently arithmetic test done with limited time to elicit the stress state in (Salafi and Kah, 2015). In this paper, while image and CWT tests are applied individually, successive tests of word/symbol and math tests are used as a combination.

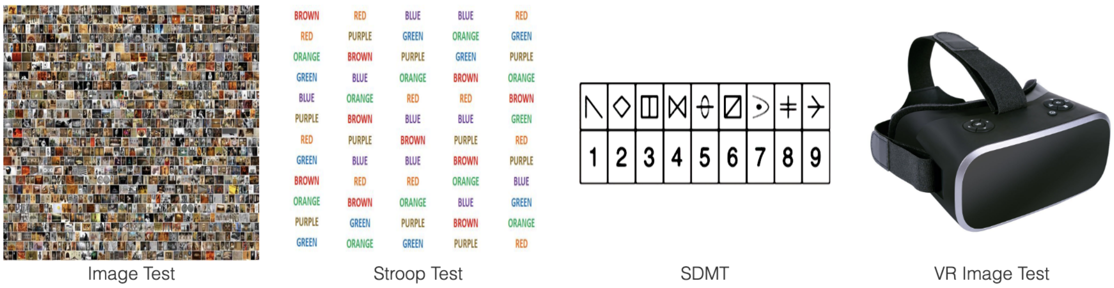

Figure **3.2**: Homepage

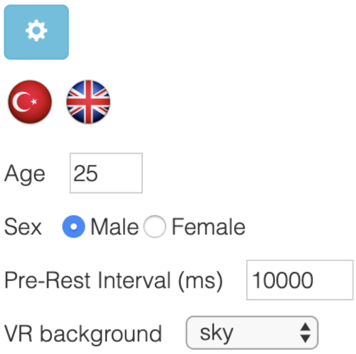

Figure **3.3**: Settings Menu

3.1.1 Image Test

In image test (Figure **3.4**), categorized images are displayed in sequence and a chirp sound is given in every three images. The categories we used in our tests include beauty, disgust, emotion, fear, food, funny, happiness, neutral, pity, sexual, violence, weird and abstract. The sequence of images are generated randomly when the image test page is refreshed. The static resources folder (/resources/images/categories) of the web application is used to scan images to display. Time (in milliseconds) for images to be shifted can be configured before starting test. Test is started by clicking "Start Test" button or by triggering from the Android application. The procedure applied for this test when using Bitalino setup is described in APPENDIX A.

The metadata file recorded in the background is in json format and the properties are age, gender, interval between images shown consecutively (interval), rest interval in miliseconds before test starts (preRestInterval) and list of shown images' informations (imagesShownTimeList) holding shown image's file path (imageFile) and the epoch

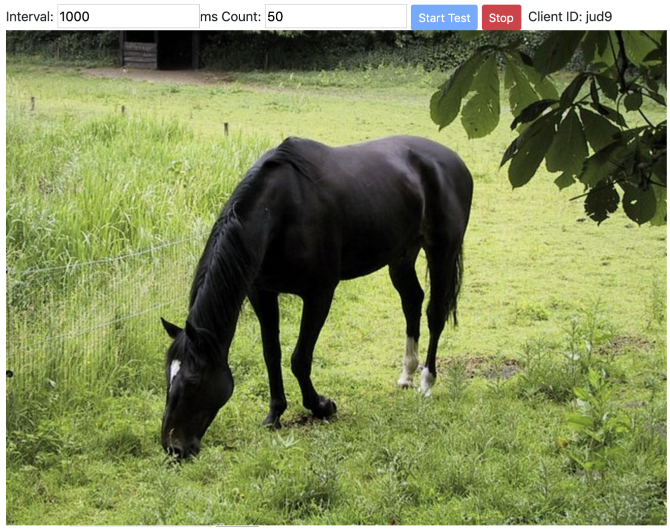

Figure **3.4**: Image Test

time when the image is displayed (date). It is uploaded to corresponding test's folder in Dropbox when the test is finished. Sample metadata file (metadata.json) content for image test can be found in APPENDIX B.

3.1.2 Stroop Color Word Test

Stroop color word test is a common test technique used for monitoring stress of users. In this test, user interacts with web application to give answers. The correct answer type (color or word) can be chosen before starting test. For the sample shown in Figure **3.5**, when color is selected as correct answer type, blue will be the correct answer, otherwise green would be accepted as correct. Users can give answers by clicking Red, Yellow, Green, Blue buttons under the images or by using 1, 2, 8, 9 keys using keyboard. The procedure applied for this test when using Bitalino setup is described in APPENDIX C.

Different from image test, correctness and epoch time of answers are also recorded in this test. The properties are age, gender, rest interval in miliseconds before test starts (preRestInterval), type of record (which can be either color or word) and list of information for each answer (imagesShownTimeList) holding shown image's file path (imageFile), the epoch time when the image is shown (date), the epoch time when answer is given (dateAnswered) and correctness of the answer (correct) fields. Sample metadata file (metadata.json) content for stroop test can be found in APPENDIX

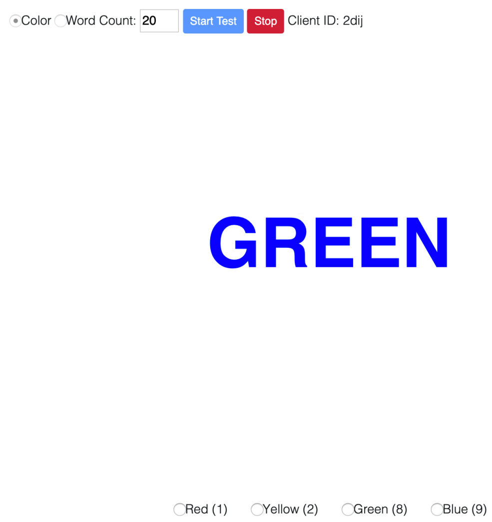

Figure **3.5**: Stroop Color Word Test

D.

3.1.3 Math and Word Test

In this combined test, the following procedure is applied :
1) Start with a math test, which is subtracting given amount successively from a given initial value (Figure **3.6**) ;

2) And then continue with word test which starts with a given word and asks to find another words successively starting with the last character of the found word (Figure **3.7**) for a configured amount of time ;
3) And subsequently continue with another math test of multiplication of two numbers which is getting harder after each answer.

Duration for each task can be configured before starting test, as can be seen in figures. Additionally, videos of the users are recorded using webcam to make video analyses possible.

3.1.4 SDMT and Math Test

SDMT is a common test technique used to measure cognitive functionality of brain. With the combination of mathematics test, we aim to increase difficulty of the test to

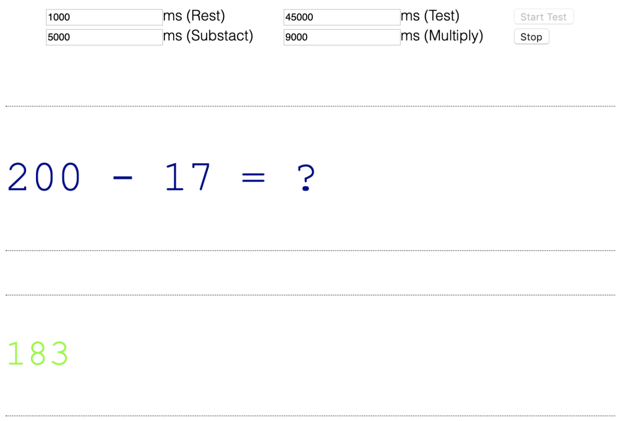

Figure **3.6**: Subtract Test

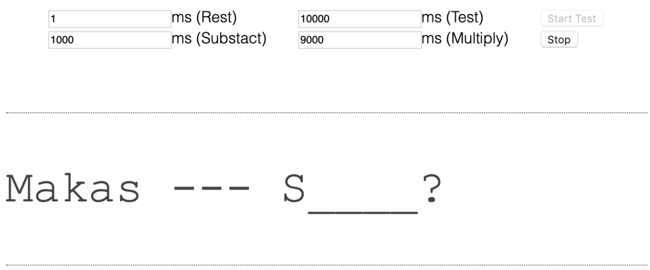

Figure **3.7**: Word Finding Test

trigger more stress. In SDMT & Math test (Figure **3.8**), a symbol is displayed and

the equivalent digit is asked for several times which is followed by accumulative adding of symbols. If the user enters a wrong sum value, the accumulation test is started over again. Correctness and epoch time of answers are also recorded during the test. The sequence of symbols are generated randomly when the test page is refreshed where the images of symbols are scanned from the static resources folder (/resources/images/sdmt) of the web application. The same sequence is used for the mathematical part of the test. Additionally, videos of the users are recorded using webcam to make video analyses possible. The procedure applied for this test when using Bitalino setup is described in APPENDIX E.

The metadata file of this test is alike that is for CWT test, except the "typeOfRecord" field which is irrelevant here ; and image file paths are of sdmt folder. Sample metadata file (metadata.json) content for SDMT & Math test can be found in APPENDIX F.

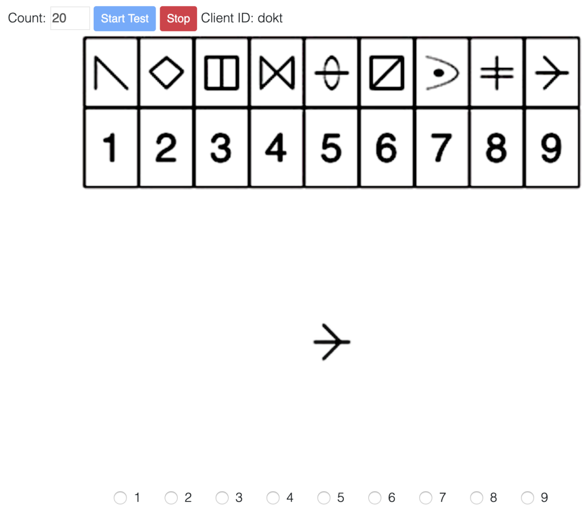

Figure **3.8**: SDMT & Math Test

3.2 Hardware

Empatica E3 wristband (Garbarino et al., 2014) having built-in memory, EDA, temperature, photoplethysmograph and 3-axis accelerometer sensors and Bitalino kits (*Bitalino Plugged Kit BLE*, n.d.) with EDA (*Plux EDA Sensor*, n.d.) and ECG (*Plux ECG Sensor*, n.d.) sensors are used to gather physiological signals during tests. Both devices have Bluetooth low energy (BLE) hardware to send data wirelessly. The details and usage are described in the following sections.

3.2.1 Empatica E3

Providing said sensors' raw data along with calculated HR and IBI data, Empatica E3 (Figure **3.9**) is easy to use unobtrusive wristband to be used for stress monitoring.

Other than sending data through BLE, Empatica E3 wristband has built-in memory to make recordings. We found it more consistent to use in-memory recordings instead of streaming data through BLE in our experiments. But it required synchronization of the records with the timings during tests. We figured out this problem by shaking hand when starting test, which made it obvious in accelerometer data (Figure **3.10**).

The EDA sensor of E3 can measure conductance in [0.01, 100]uS. The sampling rates of EDA, PPG, ACC and temperature sensors are respectively 4Hz, 64Hz, 32Hz and

Figure **3.9**: Empatica E3

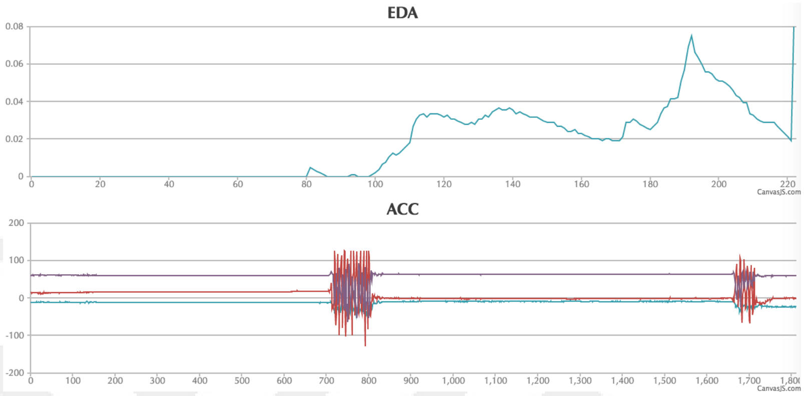

Figure **3.10**: EDA and ACC signals during test

4Hz. Also, calculated heart rate is provided once per second.

3.2.2 Bitalino Setup

Bitalino setup (Figure **3.11**) is not an ideally unobtrusive device and there is no built-in memory utility in Bitalino. The data is collected through BLE which we get connected through the Android application, which is described in Section 3.3. Pluggable ECG and EDA sensors are used with this setup where HR and heart rate variability (HRV) are extracted manually.

The conductance range of the used EDA sensor is [0-25]uS. While Bitalino kit can sample with 1, 10, 100 or 1000Hz, the EDA sensor's bandwidth is in [0-3]Hz and the ECG sensor's bandwidth is in [0.5-100]Hz.

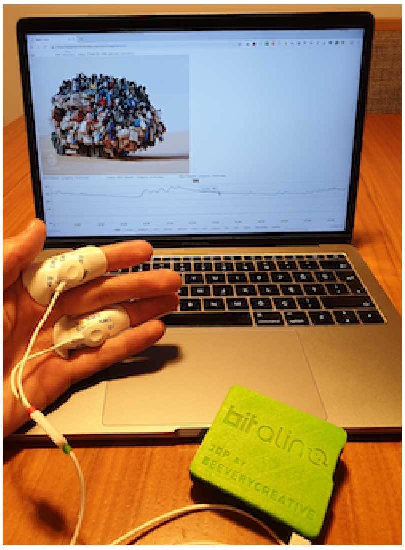

Figure **3.11**: Bitalino setup with plugged EDA sensor

3.3 Android Application

The developed Android application (Figure **3.12**) acts as a bridge between devices and the web application. It connects to devices through BLE to collect physiological data and sends the buffered data to the Data Server described in Section 3.4. When using Empatica E3 device for test, we synchronize test metadata with sensor data with the help of accelerometer data. But when using Bitalino setup, since the Android application can connect both the device and the web application, by starting the test from the Android application, we can start recording sensor data and start the test in the web application simultaneously. This ensures test metadata and sensor data become synchronized. In order to pair the Android application with the web application, the client id generated by the web application must be entered in the Android application. Only then the web application can start the test automatically. Before starting test, test type must be chosen in order to save record data to corresponding folder in Dropbox. When using Empatica E3 device, since we used built-in recording capability, only real-time data visualization is done for this wristband in the Android application.

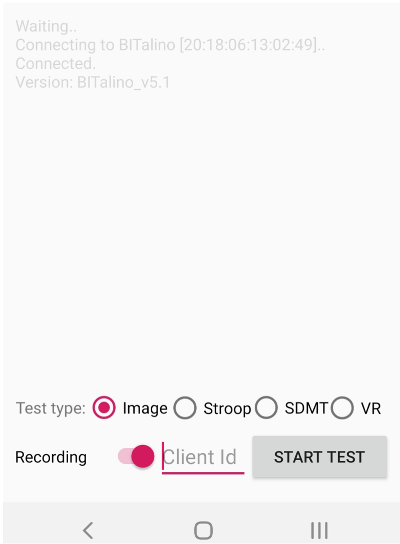

Figure **3.12**: Android application controlling Bitalino device

3.4 Data Server

Data server acts as a bridge between client applications (web and Android) and storage service (Dropbox). It has two main responsibilities : First, collecting streamed sensor data and uploading it along with test metadata for future analyses. Second, constructing websocket connection between the web application and the Android application.

The implementation is done using NodeJS. It has REST endpoints to create records of tests, update records with sensor data and save them on Dropbox by using its APIs (*Dropbox Javascript APIs*, n.d.). It also provides API to load records from Dropbox, which is used for visual graphs.

4 EXPERIMENTS AND RESULTS

For the purpose of understanding what stress changes in human physiology, several tests should be performed and results should be analyzed to see the effects of them and make inferences accordingly. There are a lot of conditions that can affect the results of experiments (i.e. devices and sensors used, the points of body to collect data, the test to perform, the psychology of the subject, the time that the test is performed on and the health of the collected data would all make difference in the resulting analyses). Keeping in mind these factors, in order to understand the relation between physiological parameters and stress, we performed the following tests using Empatica E3 wristband and Bitalino setup : Stroop color word test, image test and combination of word/symbol and math tests.

4.1 Tests using Empatica E3

According to the literature review we performed, EDA and HR parameters are the most associated parameters with stress. So, in the analyses of collected dataset with Empatica E3, we worked on these parameters. Besides built-in IBI data provided by E3, average heart rate, standard deviation of heart rate and root mean square of EDA are calculated. Along with these EDA and HR parameters, average temperature is also taken into account.

With Empatica E3 device, a total of 14 image tests with different image shift durations (summarized in Table **4.1**) ; 10 stroop color word tests (5 tests twice for color and word selections as valid choices) ; and 5 word & math tests (summarized in Table **4.2**)are performed, to participants with ages between 25-49. For stroop color word test, delay for answer and correctness of the answer are also recorded ; and for word & math test, audio and video recordings are made according to the participants' permissions where audios and videos are analyzed manually to see if the participants show any stress sign in their verbal or facial reactions.

As a result, we have also found correlation between these signals and stress in our experiments. But not all the experiments showed the same correlation such as while EDA showed an increasing trend during stress test, there was not a clear sign in HR values. In such uncertainties, it revealed that we needed further signs or parameters to take into account, or a questionnaire would provide extra information, so that we could make the decision of stress or no stress state for the data.

Table **4.1**: Image test experiments with E3

Gender	Age	Image Shift (ms)	Count
F	26	1500	1
F	25	1500	1
M	27	1000	1
M	30	1000	2
M	25	1000	2
M	25	2000	1
M	25	5000	6

Table **4.2**: Word and math test experiments with E3

Gender	Age	Audio	Video
F	29	✓	✗
F	29	✗	✗
F	49	✓	✗
M	32	✓	✓
M	32	✗	✗

4.2 Tests using Bitalino Setup

After experiencing E3 wristband, we continued tests with the Bitalino setup, and aimed at adding new sensors to our study but again started with well known sensors for stress detection, where we used EDA and ECG sensors. From the ECG sensor's raw data, filtered ECG, heart rate, systole, RR intervals; ultra low, very low, low, high and very high frequency (ULF, VLF, LF, HF and VHF, respectively) analyses of HRV are extracted (Figure **4.1**). RR interval can be simply described as interval between heart

beats. Phasic and tonic components; onsets, recoveries and peaks in skin conductance response (SCR) with amplitude of peaks are extracted from raw EDA signal. SCR is described as phasic change in electrical conductivity of skin. Tonic level of electrical conductivity of skin is also known as skin conductance level (SCL). Sample EDA raw signal with SCR and SCL parameters calculated for a SDMT & Math test can be seen

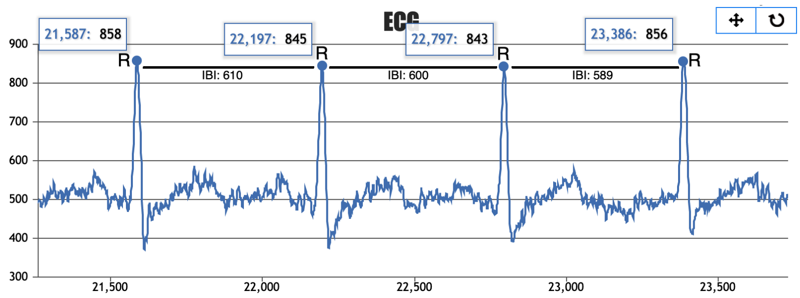

Figure **4.1**: R-R segment and IBI plotted on the ECG signal.

in Figure **4.2**. A typical SCR signal and its properties are shown in Figure **4.3**. The

parameters are extracted by using the open source NeuroKit python library (Makowski, 2016). Answer latency, HR mean, HRV LF mean, HRV HF mean, EDA mean, EDA standard deviation and EDA RMS are calculated manually for each segment. Other than these parameters, age and gender information are also used in analyses.

Recordings with Bitalino are made from two hands, via EDA sensors on fingers and ECG sensors on hands. Data are collected from 12 participants (6 females with ages between 30-36 and 6 males with ages between 22-33). Since the sensors used with Bitalino setup are pluggable and free to move with their cables unlike the Empatica E3, further tests with recordings from different body parts would give more information to compare the results, which is left as future work.

A total of 17 image tests with 2000 ms image shift interval; 45 stroop tests (including both color and word selections as the valid choices); and 19 SDMT & math tests with audio and video recordings are performed. Analyses of audio and video recordings are left as future work.

The extracted parameters are fed into classification algorithms using WEKA software (Frank et al., 2016) and analyses are performed on the test type basis and results are given with kappa statistics with 10-fold cross validation. Kappa statistics represents the reliability of the model that the classification algorithm has generated, derived from k-fold cross validation. K-fold cross validation is a technique to verify the model by dividing dataset into sections (folds), creating model for each fold and using the other folds (the folds except the fold that created the model) to evaluate the model. Used techniques and analysis results are detailed in the following sections.

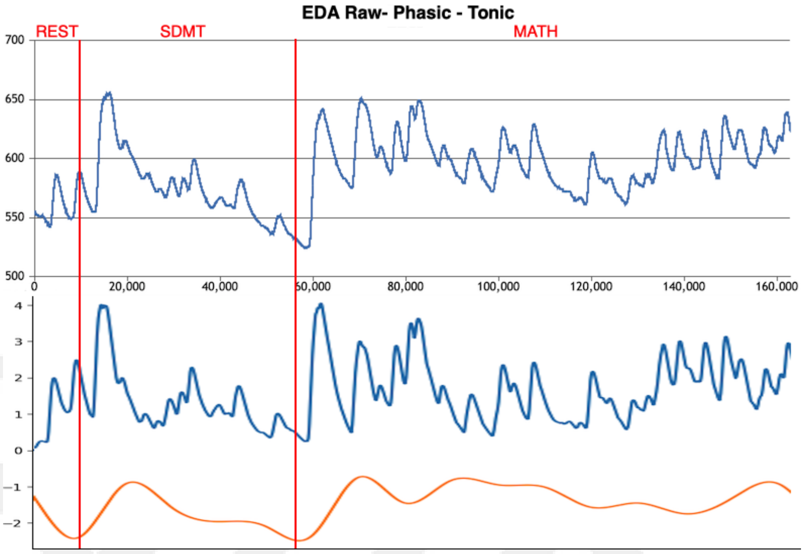

Figure **4.2**: Raw EDA, SCR and SCL plotted for SDMT & Math test with test sections

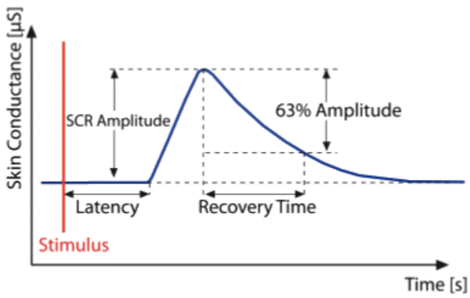

Figure **4.3**: Skin Conductance Response

4.2.1 Image Test Analysis

In image test analyses, some categories of images like fear, sexual, violence and weird are labeled as stressed, and many combinations are tried to get better classification results. After removing falsy data, a total of 70 rest, 220 stressed segments are used in analyses. Analyses results are summarized in Table **4.3**, where KStar algorithm gave the best results ; Random Forest and K-Nearest Neighbour follow.

In order to reveal the effect of gender toward affective stressors, classification analyses with only males' data and only females' data are also studied. While males' results were similar to mix analysis results, females' accuracy was about 3% better with KStar algorithm and similar with other algorithms. However, because of the fact that the most of the falsy data was of females', with low number of samples, it is hard to say that gender plays important role toward affective stressors.

Table **4.3**: Image Test Analyses Results (Rest vs Stressed)

Algorithm	Accuracy (%)	Kappa statistics
KStar	83.15	0.5007
Random Forest	82.80	0.4012
K-Nearest Neighbour	81.75	0.4708

4.2.2 Stroop Test Analysis

The collected data from stroop tests is separated into segments of rest, stressed and non-stressed sections, assuming incompatible word and ink color constructs a stressed section. A full segment of data recorded during color word stroop test is shown in Figure **4.4**, where the first 10000 samples (10 seconds with sampling rate of 1000 samples/sec) are of rest interval. Rest data is divided into 2-second segments before using in analyses. After removing falsy data, a total of 130 rest, 440 stressed and 54 non-stressed segments are used in analyses.

Analysis results on binary classification of rest vs stressed in stroop test are summarized in Table **4.4**, in which Random Forest, k-Nearest Neighbour and K-Star algorithms gave top 3 accurate results, respectively. K-Star algorithm uses entropy based distance

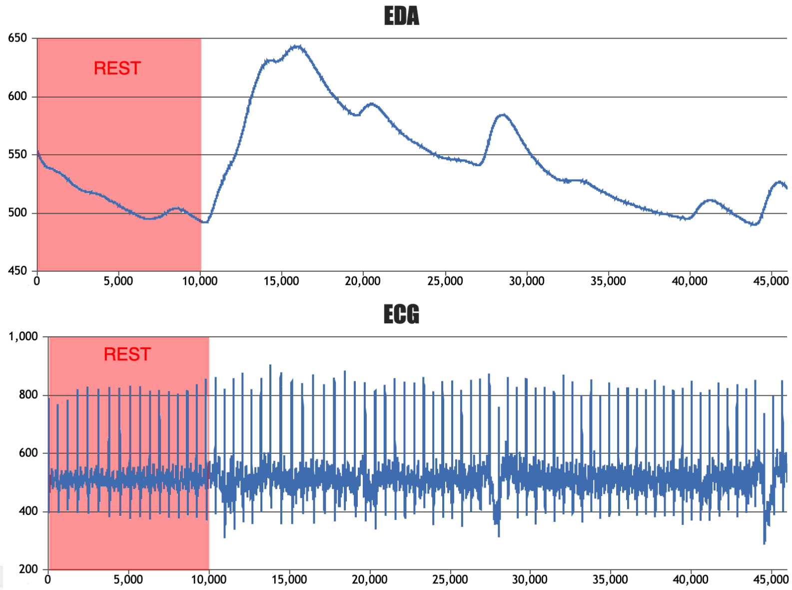

Figure **4.4**: EDA and ECG signals recorded during stroop test

function for classification while we used k=3 for KNN algorithm. Default settings with 100 iterations to construct random trees are used for Random Forest algorithm.

Classification results of rest vs non-stressed vs stressed are summarized in Table **4.5**, in which Random Forest, K-Nearest Neighbour and Multilayer Perceptron algorithms again gave top 3 accurate results, respectively. But the accuracy of this classification is lower than the rest vs stressed analysis. This is because of the fact that we had low number of samples of non-stressed sections. It would give better results if we had similar number of samples for each section.

Table **4.4**: CWT Analyses Results (Rest vs Stressed)

Algorithm	Accuracy (%)	Kappa statistics
Random Forest	83.09	0.3825
K-Nearest Neighbour	82.04	0.4105
K-Star	79.22	0.3431

Table **4.5**: CWT Analyses Results (Rest vs Non-Stressed vs Stressed)

Algorithm	Accuracy (%)	Kappa statistics
Random Forest	73.79	0.2119
K-Nearest Neighbour	72.34	0.2985
Multilayer Perceptron	71.54	0.2276

4.2.3 SDMT and Math Test Analysis

In SDMT & math test analyses, four different classifications are studied : "Rest vs Stressed", "Low vs High stress", "Rest vs Low stress" and "Rest vs High stress". The first part of the test, in which the participant selects the correct digit, is assumed to be low stressful; and the second part, in which calculation is needed, is assumed to be high stressful. A total of 95 rest, 300 low-stress and 300 high-stress segments are used in analyses. Analyses results are summarized in tables **4.6**, **4.7**, **4.8** and **4.9**. "Rest vs Low stress" and "Rest vs High stress" classifications gave better kappa statistics than "Rest vs Stressed" classification and better results than "Low vs High stress" classification.

Table **4.6**: SDMT Analyses Results (Rest vs Stressed)

Algorithm	Accuracy (%)	Kappa statistics
Random Forest	89.61	0.1331
KStar	89.06	0.3549
K-Nearest Neighbour	86.74	0.3316

Table **4.7**: SDMT Analyses Results (Low vs High stress)

Algorithm	Accuracy (%)	Kappa statistics
Random Forest	77.46	0.5493
K-Nearest Neighbour	77.46	0.5492
KStar	75.12	0.5025

Table **4.8**: SDMT Analyses Results (Rest vs Low stress)

Algorithm	Accuracy (%)	Kappa statistics
Random Forest	88.35	0.6368
K-Nearest Neighbour	84.30	0.5609
KStar	83.54	0.5379

Table **4.9**: SDMT Analyses Results (Rest vs High stress)

Algorithm	Accuracy (%)	Kappa statistics
K-Nearest Neighbour	85.78	0.5716
Random Forest	85.27	0.5598
KStar	84.26	0.5438

4.3 Discussion

According to the first experiments conducted by Empatica E3, some improvements are made for tests like getting correctness and latency of the answer in math test which was absent in Word & Math test with E3. When answer latency is taken into account, about 2% accuracy improvement is achieved. Yet, after the analyses, we noticed other improvements that could be done. For example, if the distribution of the number of the stressed and non-stressed images were even, we could obtain better results in differentiating stressed and non-stressed sections. Or, for image test, separation of stressed images and non-stressed images, like displaying 10 non-stressed images first and then displaying 10 stressed images instead of displaying random 20 images, would give better results. Also, this approach can be applied to stroop test (where random images are displayed) like color-word matching images would be displayed first and then unmatching images would be displayed. Of all stressors, SDMT & Math test has shown the best classification results which is most probably the result of having clear separate sections (i.e. Rest then Low stress then High stress sections). Analysis results of "Low vs High Stress" classification in SDMT & Math test reveals that Low and High sections have different characteristics which is also validated by the results of "Rest vs Low stress" and "Rest vs High Stress" classifications. Also, these two classifications have better results than "Rest vs Stressed" classification result, which also supports this idea since "Stressed" section have both low and high stress sections. Separation of these sections gives better results.

In all three test type analyses, Random Forest, K-Nearest Neighbour and KStar algorithms gave top three results with similar accuracies and the most discriminative parameters detected by principal component analysis and by Random Tree algorithm has been SCR amplitude, HRV LF, EDA mean and EDA RMS.

5 CONCLUSION

Hardwares used in sensors may differ from device to device. For example, higher sampling rates used in Bitalino setup can provide more resolution compared to Empatica E3 but the plugged sensors may not be capable of sampling data. These differences can change the results one can conclude according to the weight of them in the algorithm used. It can also affect the parameters extracted. So, the accuracy of the stress detection algorithms becomes heavily dependent on hardware used.

Combination of parameters extracted from physiological signals, gestures and movements monitored in videos, and changes in voice parameters extracted from audios would provide better results compared to results derived from a single or a few parameters. Since the combination of such parameters comes with very huge variations, machine learning algorithms comes into play. As more studies are done using machine learning techniques accounted with analyses of test metadata like answer delays, age and sex differences and analyses of questionnaires, even psychological effects can be revealed.

In order to eliminate the effects that can cause in falsy results, data with non-stressful states should be provided to algorithms like data that can be recorded before and after stressful event, which can also be supported with questionnaires during tests. Other than that, control groups can be constructed like those performing easy calculation tasks or without time pressure, which would guide to understand the effect of stressors.

Finally, Table **5.1** summarizes the classification results from literature and our experiments, where our results also provides good results compared to previous studies which shows that our platform can be a used as a base platform for stressor needs in future studies related with stress. # column represents the number of participants in the study and reliability measures represent how the results are validated. Classification correctness can be measured by using true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates. Sensitivity or recall is the ability to correctly identify the positive sections, calculated as true positive rate (TP / TP+FN).

Specificity represents the true negative rate (TN / TN+FP) that is correctly identified negatives' rate. Precision is positive prediction value measured as (TP/TP+FP). By using the combination of precision and recall, f-measure provides a more general reliability value calculated as 2 * Precision * Recall / (Precision + Recall) with a value between 0 and 1. Leave one person out method is like k-fold cross validation technique, using number of participants as k value and creating models for each person's records.

Table **5.1**: Summary of Results

Article	Stressor	Parameters	Algorithm	#	Accuracy (%)	Classes	Reliability
(Palanisamy et al., 2013)	MAST	HRV	KNN, PNN	40	93.75	2	Specificity :88.89, Sensitivity :100
(Frank et al., 2013)	CWT	Heart Rate, Breathing Rate, Temperature and RR Interval	Bayesian Network	15	90.35	5	Precision : 97.1, Recall : 97.7
(Zhang et al., 2014)	Cognitive workload experiments	GSR, EEG, HR	Large Margin Distribution Machine (LDM)	16	62.5(HRV), 75(GSR), 87.5(EEG)	3	-
(Shi et al., 2010)	Public speaking, MAST, CPT	26 features of HR, ECG, Respiration, SC, GSR and Temperature	SVM	22	68 ±7.3	2	Recall :80
(Setz et al., 2010)	MIST	16 features of EDA	LDA	32	82.8	2	Leave one person out cross validation
(McDuff et al., 2014)	MAST	Breathing rate, HRV LF/HF ratio	SVM	10	85	2	Leave one person out cross validation
(Wijsman et al., 2011)	Calculation, Logical puzzle, Memory	7 features of ECG, Respiration, SC, EMG	Linear Bayes Normal, Quadratic Bayes Normal, K-Nearest Neighbor, Fisher's Least Square	30	80	2	5-fold cross-validation, Error Rate : 0.2
(Ollander, 2015)	TSST, SECPT, d2	Mean of HR, Mean of GSR, Mean of absolute derivation of GSR	KNN	9	99.5 ±0.6	2	Margin of error at 95% confidence interval
(Jing Zhai et al., 2005)	Stroop	10 features of BVP, GSR and PD	SVM	6	80	2	Jackknife test
Ours	Image	HR mean, HRV LF&HF, EDA mean & SD & RMS, SCR amplitude, answer latency	K-Star	12	83.15	Rest vs Stressed	Kappa :0.5007, Precision : 0.823, Recall : 0.832, F-measure : 0.826
"	Stroop	"	Random Forest	12	83.09	Rest vs Stressed	Kappa :0.3825, Precision : 0.829, Recall : 0.831, F-measure : 0.801
"	SDMT & Math	"	K-Star	12	89.06	Rest vs Stressed	Kappa :0.3549, Precision :0.880, Recall : 0.891, F-measure : 0.885
"	"	"	Random Forest	12	77.46	Low vs High Stress	Kappa :0.5493, Precision : 0.775, Recall : 0.775, F-measure : 0.775
"	"	"	Random Forest	12	88.35	Rest vs Low Stress	Kappa :0.6368, Precision : 0.885, Recall : 0.884, F-measure : 0.874
"	"	"	KNN	12	85.78	Rest vs High Stress	Kappa :0.5716, Precision : 0.852, Recall : 0.858, F-measure : 0.850

5.1 Thesis Contribution

Within this thesis, a summary of literature review on stress detection techniques and comparison with our results are given. A controlled environment that has similar tests with those used in literature is developed. These tests include image, stroop color word, SDMT & Math and an experimental image test in VR environment which provides a test environment to study the effect of VR environment. Tests can be reached online from (*Web Application*, n.d.) and (*VR Application*, n.d.) ; and the Android application file can be downloaded from (*Android Applicaiton Repository*, n.d.). Using this platform, a test can be performed easily with Bitalino setup, even by non-developers.

The collected data to perform stress detection analyses construct an experimental dataset, which may be used in further analyses and comparisons with other studies. Also, this dataset can be grown by using the online platform ; or a private dataset can be created by providing Dropbox API key when running data server which is available at (*Data Server Repository*, n.d.) as opensource. The web application's codes are also available as opensource at (*Web Application Repository*, n.d.) which can be modified and used freely.

The analyses conducted in this study, also revealed the relationship of EDA and hearth activity with stress. It is experienced that machine learning algorithms ease analyses while working with many combinations of attributes of EDA and ECG. Also, besides the physical activities, the importance of non-physical measurements has revealed, where 2% improvement is achieved with answer latency in SDMT & Math test analyses.

5.2 Limitations and Future Work

5.2.1 Limitations

Since the results are hardware dependent, the devices used in this paper and collected dataset basically forms the results achieved and the test environment limits the scope.

5.2.2 Future Work

The more parameters analyzed the more relevant parameters can be figured out to detect stress. So, extraction of further parameters from several signal data can be studied in future works.

Beside the provided four types of tests, further test types can be added which would make it possible to compare test environments. For that, an experimental image test in VR environment is also developed which is left as future work. In this test, using WebVR polyfill (*WebVR-polyfill*, n.d.), javascript implementation of WebVR specification (*WebVR*, n.d.), images are shown in virtual reality environment in mobile website. The same application of image test in VR environment would make it possible to analyze the differences the virtual environment produces. While VR environment itself can affect how people respond to such an environment, also different virtual environments can affect differently. In order to make it possible to study on different environments, VR background setting is provided at settings menu which has sky, sky-night, beach and forest options. Besides new test types, even modification of the tests, like using different images in image tests such as that International Affective Picture System (IAPS) provides, can make it possible to compare effects of test contents.

Further tests using these test environments would increase dataset volume and provide better information to understand stress and prevent it. As Bitalino setup gives freeness to attach sensors to different parts of the body, the effect of sampling points of signals can be further analyzed to detect the most accurate points to gather signals.

Furthermore, by using the online stress test web application, online stress analysis and recommendation services can be developed for end users having devices, which is getting ubiquitous with the advent and improvements of smart devices like wristbands. Also, stress level can be monitored in real time if the classification algorithms are run simultaneously instead of processing externally.

REFERENCES

Android Applicaiton Repository (n.d.). `https://github.com/kemalceng/bitalino-android`. Accessed : 2019-05-15.

Bitalino Plugged Kit BLE (n.d.). `https://bitalino.com/en/plugged-kit-ble`. Accessed : 2019-05-15.

Chen, F. S., Schmitz, J., Domes, G., Tuschen-Caffier, B. and Heinrichs, M. (2014). Effects of acute social stress on emotion processing in children, *Psychoneuroendocrinology* **40** : 91 – 95.
URL: *http ://www.sciencedirect.com/science/article/pii/S030645301300406X*

Costin, R., Rotariu, C. and Pasarica, A. (2012). Mental stress detection using heart rate variability and morphologic variability of eeg signals, *2012 International Conference and Exposition on Electrical and Power Engineering*, pp. 591–596.

Data Server Repository (n.d.). `https://github.com/kemalceng/analyser-server`. Accessed : 2019-05-15.

de Santos Sierra, A., Sánchez Ávila, C., Bailador del Pozo, G. and Guerra Casanova, J. (2011). Stress detection by means of stress physiological template, *2011 Third World Congress on Nature and Biologically Inspired Computing*, pp. 131–136.

Dropbox Javascript APIs (n.d.). `https://www.dropbox.com/developers/documentation/javascript`. Accessed : 2019-05-15.

Flaa, A., H Mundal, H., Eide, I., Kjeldsen, S. and Rostrup, M. (2006). Sympathetic activity and cardiovascular risk factors in young men in the low, normal, and high blood pressure ranges, *Hypertension* **47** : 396–402.

Frank, E., Hall, M. A. and Witten, I. H. (2016). The weka workbench. online appendix for "data mining : Practical machine learning tools and techniques", morgan kaufmann, fourth edition.

Frank, K., Robertson, P., Gross, M. and Wiesner, K. (2013). Sensor-based identification of human stress levels, *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 127–132.

Garbarino, M., Lai, M., Bender, D., Picard, R. W. and Tognetti, S. (2014). Empatica e3 — a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition, *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, pp. 39–42.

Hassellund, S. S., Flaa, A., Sandvik, L., Kjeldsen, S. E. and Rostrup, M. (2010). Long-term stability of cardiovascular and catecholamine responses to stress tests, *Hypertension* **55**(1) : 131–136.
URL: *https ://www.ahajournals.org/doi/abs/10.1161/HYPERTENSIONAHA.109.143164*

Holzinger, A., Bruschi, M. and Eder, W. (2013). On interactive data visualization of physiological low-cost-sensor data with focus on mental stress, pp. 469–480.

Jentsch, V., Het, S. and Wolf, O. (2014). Emotion regulation : Exploring the impact of stress and sex front, *Behav Neurosci* **8**.

Jing Zhai, Barreto, A. B., Chin, C. and Chao Li (2005). Realization of stress detection using psychophysiological signals for improvement of human-computer interactions, *Proceedings. IEEE SoutheastCon, 2005.*, pp. 415–420.

Kirschbaum, C., Pirke, K. and Hellhammer, D. (1993a). The 'trier social stress test' – a tool for investigating psychobiological stress responses in a laboratory setting, *Neuropsychobiology* **28** : 76–81.

Kirschbaum, C., Pirke, K. and Hellhammer, D. (1993b). The 'trier social stress test' – a tool for investigating psychobiological stress responses in a laboratory setting, *Neuropsychobiology* **28** : 76–81.

Makowski, D. (2016). Neurokit : A python toolbox for statistics and neurophysiological signal processing (eeg, eda, ecg, emg...). Memory and Cognition Lab' Day, 01 November, Paris, France.
URL: *https ://github.com/neuropsychology/NeuroKit.py*

McDuff, D., Gontarek, S. and Picard, R. (2014). Remote measurement of cognitive stress via heart rate variability, *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2957–2960.

Minkley, N., Schröder, T. P., Wolf, O. T. and Kirchner, W. H. (2014). The socially evaluated cold-pressor test (secpt) for groups : Effects of repeated administration of a combined physiological and psychological stressor, *Psychoneuroendocrinology*

45 : 119 – 127.
URL: *http ://www.sciencedirect.com/science/article/pii/S0306453014001188*

Ollander, S. (2015). *Wearable sensor data fusion for human stress estimation*, Master's thesis, Linköping University, Automatic Control.

Palanisamy, K., M, M. and Yaacob, S. (2013). Multiple physiological signal-based human stress identification using non-linear classifiers, *Electronics and Electrical Engineering* **19**.

Plux ECG Sensor (n.d.). `https://plux.info/sensors/277-electrocardiogram-ecg-820201203.html`. Accessed : 2019-05-15.

Plux EDA Sensor (n.d.). `https://plux.info/sensors/280-electrodermal-activity-eda-820201202.html`. Accessed : 2019-05-15.

R Saslow, L., McCoy, S., Löwe, I., Cosley, B., Vartan, A., Oveis, C., Keltner, D., Moskowitz, J. and S Epel, E. (2013). Speaking under pressure : Low linguistic complexity is linked to high physiological and emotional stress reactivity, *Psychophysiology* **51**.

Ren, P., Barreto, A., Huang, J., Gao, Y., Ortega, F. and Adjouadi, M. (2013). Off-line and on-line stress detection through processing of the pupil diameter signal, *Annals of biomedical engineering* **42**.

Riera, A., Soria-Frisch, A., Albajes-Eizagirre, A., Cipresso, P., Grau, C., Dunne, S. and Ruffini, G. (2012). Electro-physiological data fusion for stress detection, *Studies in health technology and informatics* **181** : 228–32.

Salafi, T. and Kah, J. C. Y. (2015). Design of unobtrusive wearable mental stress monitoring device using physiological sensor, pp. 11–14.

Salai, M., Vassányi, I. and Kósa, I. (2016). Stress detection using low cost heart rate sensors, *Journal of Healthcare Engineering* **2016** : 1–13.

Schwabe, L., Haddad, L. and Schachinger, H. (2008). Hpa axis activation by a socially evaluated cold-pressor test, *Psychoneuroendocrinology* **33** : 890–5.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G. and Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable eda device, *IEEE Transactions on Information Technology in Biomedicine* **14**(2) : 410–417.

Sharma, N. and Gedeon, T. (2012). Gedeon, t. : Objective measures, sensors and computational techniques for stress recognition and classification : A survey. computer methods and programs in biomedicine 108(3), 1287-1301, *Computer methods and programs in biomedicine* **108**.

Shi, Y., Hoai Nguyen, M., Blitz, P., French, B., Fisk, S., De la Torre, F., Smailagic, A., P Siewiorek, D., Al, M., al'Absi, M., Ertin, E., Kamarck, T. and Kumar, S. (2010). Personalized stress detection from physiological measurements.

Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., Thoma, M. V., Spoerri, C., Ehlert, U. and Nater, U. M. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity, *Psychoneuroendocrinology* **51** : 227 – 236. This issue includes a Special Section on Biomarkers in the Military - New Findings from Prospective Studies.
URL: *http ://www.sciencedirect.com/science/article/pii/S0306453014003849*

Tulen, J., Moleman, P., van Steenis, H. and Boomsma, F. (1989). Characterization of stress reactions to the stroop color word test, *Pharmacology, biochemistry, and behavior* **32** : 9–15.

VR Application (n.d.). `https://teststres-vr.herokuapp.com`. Accessed : 2019-05-15.

Wassenberg, R., Hendriksen, J., Hurks, P., Feron, F., Keulers, E., S H Vles, J. and Jolles, J. (2008). Development of inattention, impulsivity, and processing speed as measured by the d2 test : Results of a large cross-sectional study in children aged 7–13, *Child neuropsychology : a journal on normal and abnormal development in childhood and adolescence* **14** : 195–210.

Web Application (n.d.). `https://teststres.herokuapp.com`. Accessed : 2019-05-15.

Web Application Repository (n.d.). `https://github.com/kemalceng/StressTest`. Accessed : 2019-05-15.

WebVR (n.d.). `https://webvr.info`. Accessed : 2019-05-15.

WebVR-polyfill (n.d.). `https://github.com/immersive-web/webvr-polyfill`. Accessed : 2019-05-15.

Wenhui Liao, Weihong Zhang, Zhiwei Zhu and Qiang Ji (2005). A real-time human stress monitoring system using dynamic bayesian network, *2005 IEEE Computer So-*

ciety Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, pp. 70–70.

Wijsman, J., Grundlehner, B., Liu, H., Hermens, H. and Penders, J. (2011). Towards mental stress detection using wearable physiological sensors, *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1798–1801.

Wilson, C., Barnes-Holmes, Y. and Barnes-Holmes, D. (2014). The effect of emotion regulation strategies on physiological and self-report measures of anxiety during a stress-inducing academic task, *International Journal of Psychology and Psychological Therapy* **14** : 1–15.

Yamaoka, L. K. (2014). The effects of adherence to asian values and extraversion on cardiovascular reactivity : A comparison between asian and european americans.

Zhai, J. and Barreto, A. (2006a). Stress detection in computer users based on digital signal processing of noninvasive physiological variables, *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1355–1358.

Zhai, J. and Barreto, A. (2006b). Stress recognition using non-invasive technology., pp. 395–401.

Zhang, H., Zhu, Y., Maniyeri, J. and Guan, C. (2014). Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine, *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2985–2988.

Zubair, M., Yoon, C., Kim, H., Kim, J. and Kim, J. (2015). Smart wearable band for stress detection, *2015 5th International Conference on IT Convergence and Security (ICITCS)*, pp. 1–4.

APPENDIX A IMAGE TEST PROCEDURE WITH BITALINO SETUP

1. Connect sensors to the participant.
2. Turn on the Bitalino device.
3. Turn on Bluetooth and open the Android application on smartphone.
4. Wait until the Bitalino device gets connected to the Android application.
5. Start streaming signal data by turning on "Recording" button on the Android application.
6. Enter age, sex and pre-rest interval settings from the homepage of the web application.
7. Select "Image" test on the web application.
8. Select "Test Type" as "Image" on the Android application.
9. Enter "Client ID" displayed on the web application into the "Client ID" field in the Android application. (Refresh page if client id does not appear)
10. Enter image shift duration in miliseconds in "Interval" setting.
11. Enter number of images to be displayed in "Count" setting.
12. Inform the participant about the test that he/she only will only watch the pictures displayed on the screen, and tell not to move and talk during the test.
13. Tell the participant to close his/her eyes to relax until warning sound is heard.
14. Start test by clicking "Start Test" button on the Android application.
15. At the end of the test stop recording by turning off "Recording" button on the Android application.

When test is started, the generated record id is set on the Dropbox Id field on the web page. The recordings can be reviewed on graphical interfaces by clicking Load button next to the Dropbox id. Images displayed at an instance can be viewed by clicking the points on the graphs.

APPENDIX B IMAGE TEST SAMPLE METADATA


```
{
  "age": "33",
  "gender": "M",
  "interval": "2000",
  "preRestInterval": "10000",
  "imagesShownTimeList": {
    "times": [
      {
        "imageFile": "https://teststres.herokuapp.com/resources/images/categories/
        "date": 1568055065580
      },
      ...
      {
        "imageFile": "https://teststres.herokuapp.com/resources/images/categories/
        "date": 1568055163582
      }
    ]
  }
}
```

APPENDIX C STROOP TEST PROCEDURE WITH BITALINO SETUP

1. Connect sensors to the participant.
2. Turn on the Bitalino device.
3. Turn on Bluetooth and open the Android application on smartphone.
4. Wait until the Bitalino device gets connected to the Android application.
5. Start streaming signal data by turning on "Recording" button on the Android application.
6. Enter age, sex and pre-rest interval settings from the homepage of the web application.
7. Select "Stroop" test on the web application.
8. Select "Test Type" as "Stroop" on the Android application.
9. Enter "Client ID" displayed on the web application into the "Client ID" field in the Android application. (Refresh page if client id does not appear)
10. Select either "Color" or "Word" as correct answer type from the web application.
11. Enter number of images to be displayed in "Count" setting.
12. Inform the participant about the test that he/she will give answers using the buttons at the bottom of the image according to the selected correct answer type, and tell not to move and talk during the test. Also inform that answers can be given by using mouse or 1-2 and 8-9 keys on keyboard.
13. Tell the participant to close his/her eyes to relax until warning sound is heard.
14. Start test by clicking "Start Test" button on the Android application.
15. At the end of the test stop recording by turning off "Recording" button on the Android application.

APPENDIX D STROOP TEST SAMPLE METADATA


```
{
  "age": "28",
  "gender": "F",
  "preRestInterval": "10000",
  "typeOfRecord": "color",
  "imagesShownTimeList": {
    "times": [
      {
        "imageFile": "https://teststres.herokuapp.com/resources/images/stroopTest/
        "date": 1561608816639,
        "dateAnswered": 1561608819879,
        "correct": true
      },
      ...
      {
        "imageFile": "https://teststres.herokuapp.com/resources/images/stroopTest/
        "date": 1561608845816,
        "dateAnswered": 1561608851053,
        "correct": true
      }
    ]
  }
}
```

APPENDIX E SDMT AND MATH TEST PROCEDURE WITH BITALINO SETUP

1. Connect sensors to the participant.
2. Turn on the Bitalino device.
3. Turn on Bluetooth and open the Android application on smartphone.
4. Wait until the Bitalino device gets connected to the Android application.
5. Start streaming signal data by turning on "Recording" button on the Android application.
6. Enter age, sex and pre-rest interval settings from the homepage of the web application.
7. Select "SDMT" test on the web application.
8. Select "Test Type" as "SDMT" on the Android application.
9. Enter "Client ID" displayed on the web application into the "Client ID" field in the Android application. (Refresh page if client id does not appear)
10. Enter number of images to be displayed in "Count" setting.
11. Inform the participant about the test. Tell that each symbol has corresponding value shown at the bottom of the symbol which will be always displayed during the test on top. The test has two parts; in the first part, symbols will be displayed and asked for corresponding value which the answers will be given using the buttons at the bottom using mouse or by using 1 to 9 keys on keyboard; in the second part plus sign (+) followed by a symbol will be displayed and the sum with the last symbol will be entered as answer. The answers are submitted using Enter key. While the answer is correct, plus sign will be displayed, so keep adding to the sum; otherwise plus sign will disappear and the sum will be reset to zero, so start over with the last shown symbol.
12. Tell the participant to close his/her eyes to relax until warning sound is heard.
13. Start test by clicking "Start Test" button on the Android application.
14. At the end of the test stop recording by turning off "Recording" button on the Android application.

APPENDIX F SDMT & MATH TEST SAMPLE METADATA


```
{
  "age": "30",
  "gender": "M",
  "preRestInterval": "10000",
  "imagesShownTimeList": {
    "times": [
      {
        "imageFile": "https://teststres.herokuapp.com/resources/images/sdmt/1.png"
        "date": 1561608816639,
        "dateAnswered": 1561608819879,
        "correct": true
      },
      ...
      {
        "imageFile": "https://teststres.herokuapp.com/resources/images/sdmt/7.png"
        "date": 1561608845816,
        "dateAnswered": 1561608851053,
        "correct": true
      }
    ]
  }
}
```

BIOGRAPHICAL SKETCH

Born in 1986 in Afyonkarahisar, Turkey, Kemal has graduated from Afyon Science High School with first degree and completed bachelor of computer engineering program at Middle East Technical University. With hands on experience in software development, he had worked in various sectors including research and development, software security, banking, real estate, crypto trading; while developing low level applications to mobile, web and desktop applications for customers from end-users to government and military institutions.