FUSION OF PUPIL DILATION AND FACIAL TEMPERATURE FEATURES FOR DETECTION OF STRESS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

SERDAR BALTACI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
MEDICAL INFORMATICS

NOVEMBER 2016

FUSION OF PUPIL DILATION AND FACIAL TEMPERATURE FEATURES FOR DETECTION OF STRESS

Submitted by **SERDAR BALTACI** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in The Department of Medical Informatics Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin _____
Director, **Graduate School of Informatics**

Assoc. Prof. Dr. Yeşim Aydın Son _____
Head of Department, **Medical Informatics**

Assist. Prof. Dr. Didem Gökçay _____
Supervisor, **Medical Informatics**

Examining Committee Members:

Prof. Dr. Aydan Erkmen _____
Electrical and Electronics Engineering, METU

Assist. Prof. Dr. Didem Gökçay _____
Medical Informatics, METU

Prof. Dr. Osman Eroğul
Biomedical Engineering, TOBB _____
University of Economics and Technology

Prof. Dr. Hasan Oğul _____
Computer Engineering, Başkent University

Assoc. Prof. Dr. Yeşim Aydın Son _____
Medical Informatics, METU

Date: **24.11.2016**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this wok.

Name, Last name: SERDAR BALTACI

Signature : _____

ABSTRACT

FUSION OF PUPIL DILATION AND FACIAL TEMPERATURE FEATURES FOR DETECTION OF STRESS

Baltacı, Serdar
Ph.D., Department of Medical Informatics
Supervisor: Assist. Prof. Dr. Didem Gökçay

November 2016, 92 pages

Stress has several negative physiological/physical impacts in our lives. Hence, it is important to recognize stress during daily activities. The relationship between stress and physiological or physical signals has been studied for a long time. The aim of this dissertation is to detect stress remotely using pupil diameter and facial temperature analysis. For this purpose, we developed a stress triggering experiment in order to generate physiological/physical effects. Our experiment consists of 2 parts which are applied consecutively. The first part was used as a baseline for neutral emotion, in which neutral pictures of International Affective Picture System (IAPS) were utilized. In the second part, to generate stress, negative pictures of IAPS were used.

To detect emotional state of the participants, pupillary and facial thermal responses were measured using a TOBII TX300 eye tracker and a FLIR SC620 thermal camera. Entropy in a sliding window was used to accommodate the time differences in the physiological rise and fall profiles of pupil and thermal data. Pupil and thermal features derived from the measured signals and the entropy based values were fused at the feature level. Finally, classification accuracy of stress was enhanced with machine learning techniques. We were able to identify stressful responses from the participants with an accuracy of 83.8% using AdaBoost and Bagging classification methods. Results also show that the experimental protocol we suggested for stress detection is highly applicable based on pupil diameter and facial temperature.

Keywords: Stress Detection, Pupil Dilation, Facial Thermal Changes, Feature Level Fusion, Shannon Entropy

ÖZ

GÖZBEBEĞİ AÇILIMI VE YÜZE AİT SICAKLIK ÖZNİTELİKLERİNİN STRES TESPİTİ İÇİN BİRLEŞTİRİLMESİ

Baltacı, Serdar
Doktora, Tıp Bilişimi Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Didem Gökçay

Kasım 2016, 92 sayfa

Stres, günümüzde insanlığın önemli sorunlarından birisi olarak karşımıza çıkmakta olup stresin hayatımıza etki eden birçok negatif fizyolojik/fiziksel etkileri bulunmaktadır. Stres ile fizyolojik ya da fiziksel sinyaller arasındaki ilişki, günlük aktiviteler sırasında stresi fark etmenin öneminden dolayı uzun zamandır irdelenmektedir. Bu tezin amacı gözbebeği çapı ve yüzdeki sıcaklık verisinin analiziyle stresi uzaktan tespit etmektir. Bu amaçla, fizyolojik/fiziksel etki oluşturmak için stresi tetikleyen ve arka arkaya uygulanan iki bölümden oluşan bir deney geliştirdik. Birinci bölüm, nötr duygu temeli oluşturmak için kullanılmış olup, nötr Uluslararası Duyuşsal Resim Sistemi (IAPS) resimleri içermektedir. İkinci bölümde ise, stres oluşturmak için, negatif IAPS resimleri kullanılmıştır.

Katılımcıların duygusal durumunu tespit etmek amacıyla, gözbebeği ve yüz termal tepkileri TOBII TX300 gözbebeği tarayıcısı ve FLIR SC620 termal kamera kullanılarak ölçülmüştür. Bu çalışmada, göz bebeği ve termal verinin fizyolojik artış ve azalışındaki zaman farkını birbirine uyumlu hale getirmek için kayan pencere içerisinde entropi yöntemi kullanılmıştır. Sinyal ve entropi değerlerinden elde edilen termal ve gözbebeği özniteliklerinin, öznitelik seviyesinde füzyon edilmesi sağlanmıştır. Stres sınıflandırmasının doğruluğu makine öğrenme teknikleri ile iyileştirilmiştir. Bu çalışmada, AdaBoost ve Bagging sınıflandırma yöntemleri kullanılarak 83.8% doğrulukla katılımcıların stres tepkileri tespit edilmiştir. Araştırma sonuçları, gözbebeği çapını ve yüz sıcaklığını kullanmaya dayalı önerilen deneysel protokolün stresin tespiti için son derece uygulanabilir olduğunu göstermektedir.

Anahtar Sözcükler: Stres Tespiti, Göz Bebeği Büyümesi, Yüzdeki Termal Değişiklikler, Öznitelik düzeyli füzyon, Shannon Entropi

To Ayşegül, my amazing wife,

To my children, Betül, Sedat and Selim.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Assist. Prof. Dr. Didem Gökçay for the continuous support of my study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me through my entire research and thesis writing. I am very grateful for her excellent supervision. I can simply not imagine having a better supervisor and mentor for my PhD study.

Besides my supervisor, I would like to thank the rest of my thesis committee: Prof. Dr Aydan Erkmen, Prof. Dr. Osman Eroğul, Prof. Dr. Hasan Oğul, Assoc. Prof. Yesim Aydin Son, for their encouragement, insightful comments, and hard questions. Without their guidance and criticisms this thesis would not be completed.

I would also like to thank all my colleagues, Selgün, Hayriye, Serdar, Aykut, Ayşenur and Ömer Faruk for their friendship, to all faculty members and former faculty members of the institute for their guidance and criticisms during the courses I took.

Special thanks goes to my wife, Ayşegül Baltacı, for her continuous support and understanding. I would also like to thank my parents Ulviye Baltacı, Sabahattin Baltacı and mother-in-law, Ümmü Salık for their endless and limitless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABRF	AdaBoost with random forest
ANS	Autonomic Nervous System
Bagging	Bootstrap Aggregating
BVP	Blood Volume Pulse
EDA	Electrodermal Activity
EEG	Electroencephalography
EEG	Electroencephalography
EO	Electro Optical (Camera)
GSR	Galvanic Skin Response
HCI	Human Computer Interaction
IAPS	International Affective Picture System
IG	Information Gain
mK	Mili Kelvins
MWIR	Mid-Wave Infrared (Camera)
NETD	Noise-Equivalent Temperature Difference
PD	Pupil Dilation
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SDK	Software Development Kit
SMM	Simple Moving Median
SPSS	Statistical Package for the Social Sciences (Software)
ST	Skin Temperature
SVM	Support Vector Machine
TC	Thermal Camera
WEKA	Waikato Environment for Knowledge Analysis (Software)

CHAPTER 1

INTRODUCTION

Stress detection is an important issue in the human computer interaction (HCI) domain. Detecting the stress level of a computer user could possibly improve the computers' ability to respond intelligently and drive the user away from negative emotional behaviors during HCI. In this way, the overall system performance can be enhanced and users'/workers' psychophysical states may become more suitable for performing the task (Czaja & Sharit, 1993; Fujigaki & Mori, 1997). Furthermore, early diagnosis of psychological disorders can be facilitated in clinical settings (Jaimes & Sebe, 2007).

Physiological signals are reliable indicators of emotional states of the subject (Zhai, Barreto, Chin, & Li, 2005). Changes in physiological arousal during stressful conditions are quantifiable through skin conductance (Lang, Greenwald, Bradley, & Hamm, 1993), thermal camera recordings (Pavlidis, 2003; Pavlidis et al., 2007) and pupil dilation (Bradley, Miccoli, Escrig, & Lang, 2008).

Stimuli with positive and negative arousal increase pupil diameter more compared to neutral stimuli. According to Bradley et al. (2008), emotional pupil dilation responses are detectable within 2-3 seconds. When comparing measurements of other physiological signals, the measurement of pupil size has important advantages. One advantage is that it is an unobtrusive method because sensors do not need to be attached to the user. Another important advantage of pupil size measurement is that pupil size variation is an involuntary response of the autonomic nervous system (ANS) (Partala & Surakka, 2003). This means that pupil size variation cannot be controlled voluntarily; therefore, it identifies actual spontaneous response. In contrast, emotions detected through facial expressions are prone to visually observable changes that can be masked, inhibited, exaggerated, and faked (Ekman, 1985; Ekman & Friesen, 1982; Partala & Surakka, 2003; Surakka & Hietanen, 1998). Despite their advantage, pupil size measurements are rarely used in arousal detection due to the interference effects related to lighting conditions and cognitive effort (Bradley et al., 2008). Pupil diameter changes can be utilized for various forms of stimuli, such as images and sounds. For instance, visual (Bradley & Lang, 1994), auditory (Baltaci & Gokcay, 2012; Partala, Jokiniemi, & Surakka, 2000; Partala & Surakka, 2003) and combined visual and auditory stimuli (Morency, Mihalcea, & Doshi, 2011) can be used in order to detect emotion. Pupil data can also be fused with other modalities such as electroencephalography (EEG) at the decision level of classification models to increase the accuracy (Qian et al., 2009).

Thermography is an attractive modality, as it is completely non-contact and unlike visible light imaging systems, unaffected by skin color or ambient lighting conditions. Increase in an individual's stress level produces changes in facial skin temperature that can be reliably detected using thermal imaging (Gane, Power, Kushki, & Chau, 2011). Within the last decade, thermal cameras have been introduced to predict arousal as a result of increased blood flow in the face (Shastri, Merla, Tsiamyrtzis, & Pavlidis, 2009; Yun, Shastri, Pavlidis, & Deng, 2009). During emotion states with high arousal, the blood flow in the supraorbital and periorbital vessels increases. Consequently, such increases in blood flow raise the periorbital region temperature, which can be captured through a highly sensitive thermal camera. A recent study (Nhan & Chau, 2010) found that, emotional states induced by the viewing of images from IAPS (Lang, Bradley, & Cuthbert, 2008), could be distinguished from a baseline (i.e. neutral) emotional state using thermal video of the face with accuracies ranging from 70% to 80%.

The most commonly used methods of emotion detection involve skin conductance response (SCR), electroencephalography (EEG). SCR shows the continuous variation in the conductivity of a person's skin. SCR is not so viable method since range is subjective to participants. External factors such as temperature, humidity and internal factors (e.g. medications) can change SCR measurements and lead to inconsistent results with the same stimulus level. And also it is not comfortable for participants. On the other hand, EEG measures electrical potentials of the brain. EEG is not feasible also because it is intrusive and validation of signal is difficult. These two methods (SCR & EEG) do not use remote sensors, hence they are not feasible. Among emotion detection methods, pupil dilation and thermal recordings are non-invasive and non-intrusive, besides they do not require direct contact with the participants. These methods are more viable than the first two, but they have one downside which is "inevitable data loss". The main reasons for data loss are; head movements, environmental heat and light conditions.

The Aim of the Thesis, Research Questions and Hypotheses

Recent advances in HCI have demonstrated significant stress detection capability with thermal or pupil data separately. In order to increase the accuracy of stress detection non-invasively in near real-time, new studies aiming to use pupillary and facial thermal changes are needed. Our aim is to use pupil and facial thermal data in order to predict stress remotely in a robust, non-invasive and subject independent way. The focus is to provide a test bed using pupillary and thermal signals for this purpose.

As explained before, stress can be triggered by visual and behavioral stimuli. In this thesis, we proposed a new system to separate stressful states of a participant from his/her neutral states. In this system, we exploited the complementary natures of one fast and one slow physiological signal, namely pupil dilation and skin temperature respectively. In the experiment, IAPS pictures that vary in valence and arousal axes were used as stimuli.

Research Question 1. Can visual and behavioral stimuli affect participants' stress and can this stress be measured by physiological changes?

Hypothesis 1. Pupil size variation can be used in order to detect user's stress. At stressful times during our experiment, participants' pupil diameter will increase in comparison to other times.
Hypothesis 2. Stress can be evaluated by measuring thermal changes of the facial area. At stressful times during our experiment, participants' facial temperature will increase in comparison to other times.

Research Question 2. Is it possible to increase the success of stress detection by fusion of pupil, thermal data and entropy based features?

Hypothesis 3. Fusing pupil and thermal features will increase the success of stress detection.
Hypothesis 4. Using features that capture the difference of the rise and fall profiles in thermal and pupil signals will increase the success of stress detection.

In this thesis, the research questions presented above are handled through an experiment we conducted. **The first chapter** is a general introduction of the study. Following the introduction, this chapter is concluded by presenting the motivation, research questions and hypotheses of the study. **Second chapter** gives an overview of the stress, emotions, pupillary and facial thermal responses. This chapter comprises of a brief summary of the pupillary and thermal response physiology, their significance on scientific research and measurement techniques. **Third chapter** consists of the experimental setup and the method that is executed to verify the hypothesis. **Fourth chapter** presents data processing, proposed feature extraction and classification algorithms. **Fifth chapter** consists of the results of the experiment. Analyses of these results and all statistical tests are given in detail. **Sixth chapter** involves discussion about the results and the hypotheses of the study. **Seventh chapter** includes possible implications and ideas as a future work and conclusion that is provided as a summary of the findings.

Findings of this research was published in International Journal of Human Computer Interactions with the title "Stress Detection in Human Computer Interaction: Fusion of Pupil Dilation and Facial Temperature Features" in 2016 (Baltaci & Gokcay, 2016).

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND

This chapter comprises of four sections in which relevant aspects of the literature in stress, pupillary response, thermal response and related works are presented. In the first section, developmental, comparative and cognitive aspects of emotion and stress are reviewed. In the second section, eye tracking is discussed specifically for pupillary response. This section is followed by representation of studies focusing on thermal responses and facial temperature changes. In the fourth section, related pupil and thermal works in stress detection are discussed.

2.1. Stress and Emotions

Stress is defined as a temporarily-induced physiological or psychological imbalance caused by an action or a situation (stressor) which can be regarded as a possible danger or threat. An emotional (mental) stressor is one in which only information reaches the brain with no direct physical impact on the body. This information may place demands on either the cognitive systems (thought processes) or the emotional system (feeling reactions, like anger or fear) in the brain (Yuen et al., 2009). A physical stressor is one that there is a direct effect on the human body. This may be an external condition (heat, cold and noise) or due to the internal physical/ physiological demands of the human body (physical exercise).

Due to its impact on the quality of life, human stress analysis has received special attention in the recent decades. Besides social factors, financial, economic, political, chemical, biological and physical factors can be sources of stress. These factors can lead to various disorders, diseases, low performance and depression (Sharawi, Shibli, & Sharawi, 2008). In addition to negative effects on human health, stress plays a crucial role in various cognitive tasks including rational decision making, perception and learning (Jing Zhai et al., 2005).

Detection and characterization of human stress by computers are ongoing research areas in the field of HCI that are still attracting a lot of attention due to their importance in our daily lives. HCI involves a two-way exchange where each participant should be aware of the other party. Therefore, to make machines be aware of the stress level of the user would result in a more natural HCI (Jing Zhai et al., 2005).

Emotion can be described as a response to an environmental event that is considered as purposeful behavior in the adaptation of the organism to dynamic environmental demands (Gokcay & Yildirim, 2011). Cognitive, affective, behavioral, and autonomic sub-systems are involved in this response. When emotional stressor is in charge, only information reaches the brain with no direct physical impact on the body. Cognitive systems (thought processes) or the emotional systems (feeling responses, such as anger or fear) in the brain may be triggered by this information (Yuen et al., 2009). Our research attempts to visualize and evaluate the emotional state identified as 'stress' of the computer user.

Subjective experience, emotional expression, and physical sensation are three distinct but complementary components of emotions (Erdem & Karaismailoglu, 2010). While subjective experience or in other words, personal feelings of current emotions are hard to investigate, emotional expression can be studied instead through facial expressions. The third component, physical sensation is manifested through the autonomous nervous system, which in turn modifies physiological arousal. Changes in physiological arousal are quantifiable through pupil dilation or thermal camera recordings (Gokcay, Baltaci, Karahan, & Turkay, 2011).

Once captured, emotions can be categorized or quantized in two different ways:

- Using distinct emotional classes such as happy, angry, fearful, surprised (Circumplex model)

- Through continuous values along two orthogonal axes, valence and arousal (Dimensional model)

In order to account for emotional states, most recently, the dimensional model has gained impetus. According to this model, emotions consist of two measures: **valence** (refers to how positive or negative an event is), and **arousal** (reflects whether an event is exciting/agitating or calming/soothing). While arousal can be predicted from physiological features such as perspiration or blood flow, valence can be predicted from facial expressions or gestures.

Valence dimension ranges from highly negative to highly positive on a scale of 9. Moreover, arousal dimension ranges from calming to exciting on a scale of 9 (Figure 1). Therefore, stimuli can be highly positive and exciting (e.g. miracle), highly positive and calming (e.g. relaxed), highly negative and exciting (e.g. slaughter), highly negative and calming (e.g. fatigued).

Happiness, sadness, anger, fear, disgust, and surprise are six basic emotions that are widely accepted. These basic emotions are intuitive and innate. Independent from cultures, they are composed of organized automatic and stereotypical behaviors and they are necessary for survival. Aforementioned emotions can be divided into two main categories which are pleasant (happiness, surprise) and unpleasant (sadness, anger, fear, disgust). The pleasant ones result in positive affect while unpleasant ones comprise

negative affect (Gökçay, 2011; Izard, 2009). Among these categories, neutral and stressful (unpleasant emotion) stimuli fall within the scope of this thesis.

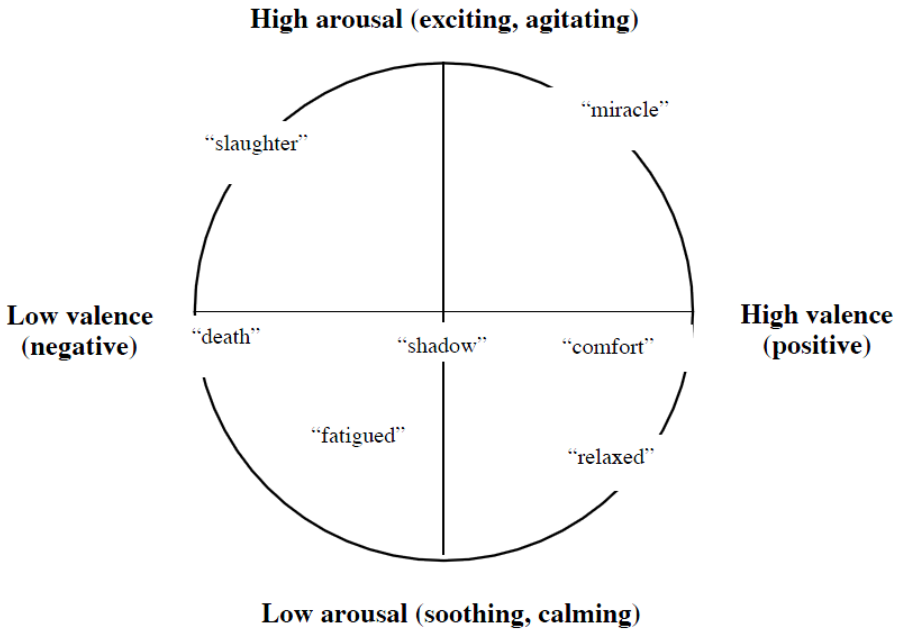


Figure 1. Two dimensions of Emotional Model (Kensinger, 2004)

Emotion recognition techniques can be systematized according to the modalities or channels such as face, voice and text. In terms of feasibility each modality has advantages and disadvantages and there are several factors that affect significance of every modality (Calvo & D'Mello, 2010). Some of these factors can be found below.

- Signal validity as a natural strategy for recognizing an affective state
- Signal reliability in real world environments,
- Signal time resolution as it relates to the specific needs of the application,
- Intrusiveness and cost for the user

The current thesis focuses mainly on two affective input signals: pupil diameter and facial temperature changes. The main contribution of this study is the design of an affect estimator in order to estimate negative arousal (stress) and neutral states robustly using these signals. Suppressing such emotions or social masking of these physiological signals is impossible (Jonghwa Kim & Andre, 2008; Kim, Bang, & Kim, 2004) as these emotions originate from the activity of the ANS, hence they cannot be triggered by any conscious or intentional control.

Subjective and complex nature of physiological signals, the inability to visually perceive emotions and sensitivity to movement artefacts from the data make it difficult for annotating and obtaining the ground truth from the raw physiological data (Kim et al., 2004). In order to achieve high classification rate for the a system, despite individual

variability brought on by the users, efficient emotion induction method(s), larger data samples and intelligent signal processing techniques are essential (Jerritta, Murugappan, Nagarajan, & Wan, 2011).

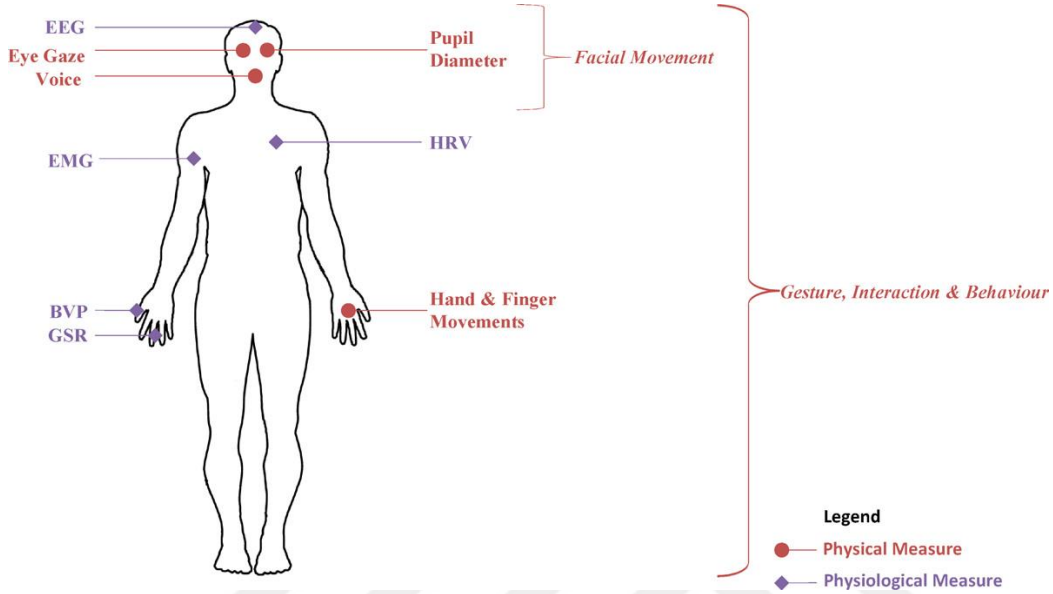


Figure 2. Physical and physiological signals that were investigated in stress detection (Sharma & Gedeon, 2012)

Physical features are properties that can be seen without the need for equipment and tools, but tools are necessary in order to detect physiological features. Figure 2 shows signals that can be investigated for the stress detection. While EMG, EEG, BVP, GSR are physiological signals, facial expression, eye gaze, blinks, pupil dilation, and voice are physical signals that are sensitive to stress (Sharma & Gedeon, 2012).

2.2. Pupil Size Variation

The pupil is a hole located in the center of the eye's iris that allows light to enter the retina. Iris has sphincter pupillae (circular) and dilator pupillae (radial) muscles to control the constriction (miosis) and the dilation (mydriasis) of the pupil, respectively (Figure 3). It is known that all sensory stimuli (visual, tactile, auditory, gustatory, olfactory) may result in pupillary responses (Beatty & Lucero-Wagoner, 2000).

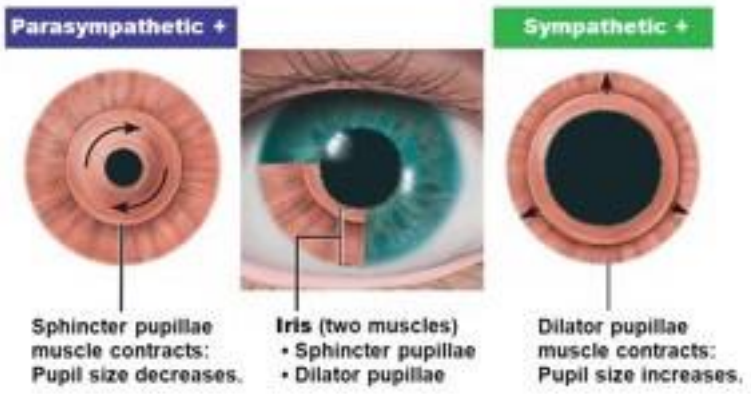


Figure 3. Muscles of the iris, pupil constriction and dilation (2013 Pearson Education Inc.)

Pupillary response is a physiological response that varies the size of the pupil, via the optic and oculomotor cranial nerve. Both sympathetic and parasympathetic pathways of ANS control pupillary responses. Thus, pupil dilation is caused by not only sympathetic system but also parasympathetic system. When sympathetic activity increases, dilator muscles' activity also increases. Alternatingly, inhibition of parasympathetic system minimizes the activity in sphincter muscle and causes dilation. In other words, response to the changes in both divisions of ANS may lead to changes in pupil diameter.

Under normal conditions, light and accommodation reflexes result in pupil constriction and dilation (Andreassi, 2006). Pupils constrict in intense light whereas they dilate in dim light. Pupil diameter of human ranges between 1.5 mm and 8-9 mm and pupils' initial reaction to light and stimuli occurs in 1-1.5 seconds and peaks around 2 seconds (Beatty & Lucero-Wagoner, 2000).

Moreover, in stable lighting conditions, if pupil dilation is less than 0.5 mm it can be an indicator of cognitive processing (Beatty, 1982; Beatty & Lucero-Wagoner, 2000) and named as task-evoked pupillary response (TEPR). In terms of cognitive load, arousal and interest indication change in the pupil size is a reliable measure for some tasks which involve attention, memory, problem solving and decision-making (Beatty, 1982; Beatty & Lucero-Wagoner, 2000).

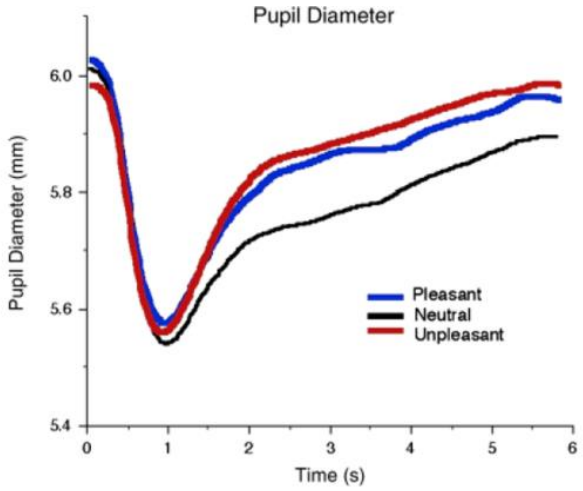


Figure 4. Pupillary response when viewing IAPS Images (Bradley et al., 2008)

During a task, latencies and peaks of pupillary responses depend on the task type. For example, during viewing visual and emotional stimuli such as a laughing man and a crying baby, dilation occurs after 2-7 seconds. While listening to auditory and emotional stimuli (laughing, crying sounds, etc.) dilation occurs after only 2-3 seconds (Bradley et al., 2008; Partala & Surakka, 2003). According to Bradley, pupillary variations were greater when viewing emotional pictures (see Figure 4).

Measurement Techniques

For reading researches, eye tracking technology is demonstrated over 100 years ago (Rayner, K., Pollatsek, A., Ashby, J., & Clifton, 2011). Electro-oculography (EOG), scleral search coils, photo-video oculography (POG-VOG) and pupil/corneal reflections (dual-Purkinje method) are some of the techniques to track eye movements.

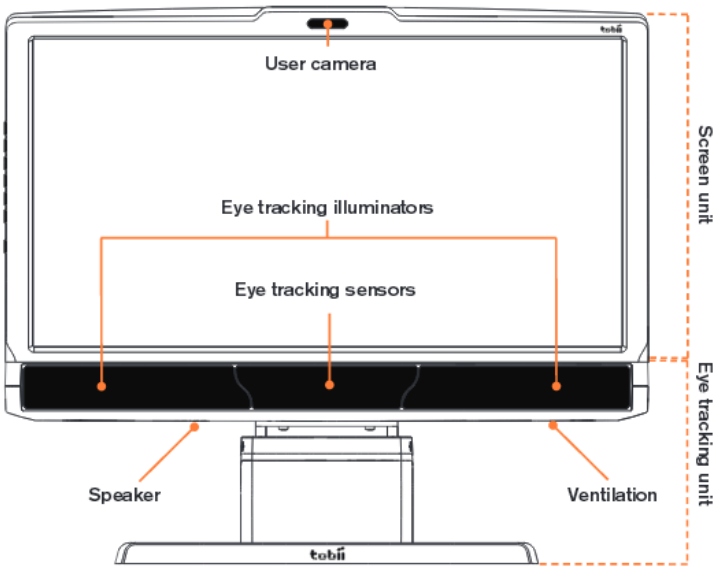


Figure 5. Eye Tracking System (TOBII TX300)

Pupil/corneal reflections (dual-Purkinje) is one of the most recently utilized eye tracking method which comprises of a desktop computer with an infrared camera under the monitor and dedicated software (Figure 5). In this method, to create reflections, camera emits infrared light to the eye. Light enters the retina and great amount of the light is reflected back and results in bright pupil effect for detection. As a small glint, infrared light generates the corneal reflection. When eye tracking software recognizes the center of the pupil and the corneal reflection, then the distance between them is measured and point of fixation can be found (Duchowski, 2007).

Eye trackers can measure pupillary responses; furthermore, fixations and eye movements. By the help of latest eye trackers like TOBII TX300, pupillary activity can be measured by a particular pixel counting method in which, pupil size is measured by counting the number of pixels in the pupillary area.

Calibration process is required in all video-based eye trackers, including pupil/corneal reflection method. In the calibration process, dots at different locations on the screen are presented to participants upon which participants have to fixate repeatedly several times in order to excess a limited threshold (Wang, 2011). TOBII software development kit (SDK) includes interfaces and a platform for designing experiments.

2.3.Facial Temperature Change

Under conditions of stress or physical activity, the body is prepared for a rapid defense reaction by the sympathetic division of ANS. Because of stimuli that involves emotional excitement, injury, stress, or exercise, hypothalamus stimulates the adrenal medulla for an increase of epinephrine and norepinephrine secretion. These hormones enable "fight or flight" response in target tissue and they arrive at their target tissue by cardiovascular system. Increased heart rate and contractile force, dilation of blood vessels in skeletal and cardiac muscles, and constriction of blood vessels in internal organs are major responses that are triggered by these hormones. These responses energize the muscles, brain, and heart for physical activity but conserve energy by slowing the functions of internal organs and the gastro-intestinal system(Seeley et al., 2008)

The core body temperature rises above the constant homeostatic range, as the metabolic activity of skeletal muscles increases. In order to promote methods of heat loss, the hypothalamus receives input from thermoreceptors. Heat is transferred from the body core to its surface by dilation of the blood vessels in the skin, and then heat is released to environment with three types of heat transfer. First, conduction from the blood to skin transfers heat. Second, convection as air passes over or sweat evaporates from the skin transfers heat. Third, heat can be transferred from the skin to the environment by radiation. (Cross, 2013)

Measurement Techniques

In old times physicians measured body temperature by just touching to patients with bare hands in order to assess patients' physical conditions. In early 18^{th} century, quantitative measurement of body temperature was made possible. Nowadays, several

advanced and easy body temperature measurement methods are available. For example, mercury in glass, sterile thermocouples, radiometers and liquid crystal can be used for body temperature measurements. They are cheap, precise and easy to use but they need direct contact with target. However, detection and quantification of natural radiation is the only method known today for non-contact body temperature measurement. Modern thermal infrared imaging methods depend on the radiation measurement techniques. (Khan, Ward, & Ingleby, 2006; Ring, 1998)

Because of its composition and structure, the human body surface is an efficient radiator. By using some of the well-known non-invasive radiation detection methods, it is easy to measure infrared emissions from the skin surface (Khan et al., 2006; Ring, 1998). In addition, sophisticated infrared cameras which are generally inexpensive are widely used for analysis of patterns of skin temperature variations (Fujimas, 1998; Khan et al., 2006).

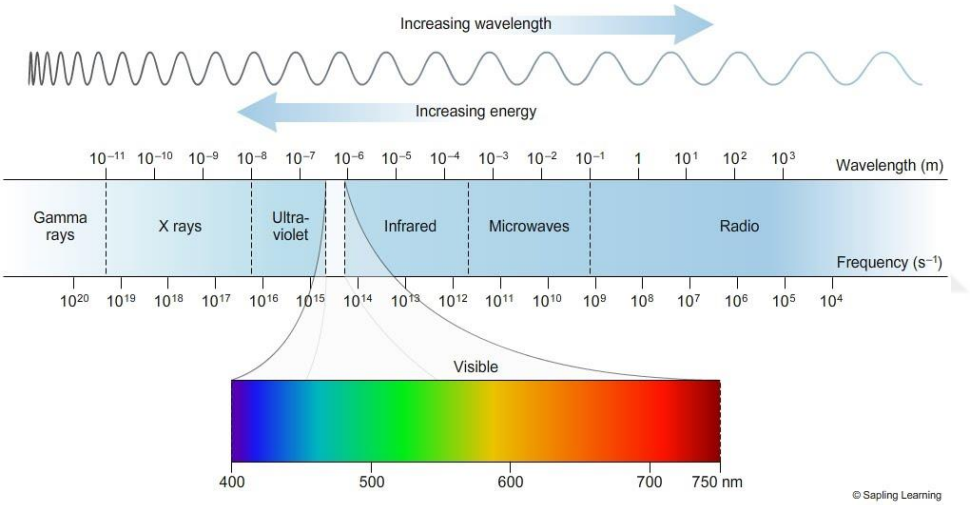


Figure 6. Electromagnetic Spectrum (Sapling Learning)

Thermal sensitivity, or Noise-Equivalent Temperature Difference (NETD), measures the smallest temperature difference that a thermal imaging camera can detect in the presence of electronic circuit noise. Cameras with a low NETD will detect smaller temperature differences and provide higher resolution images with increased accuracy. MiliKelvins (mK) is the measurement for thermal sensitivity.

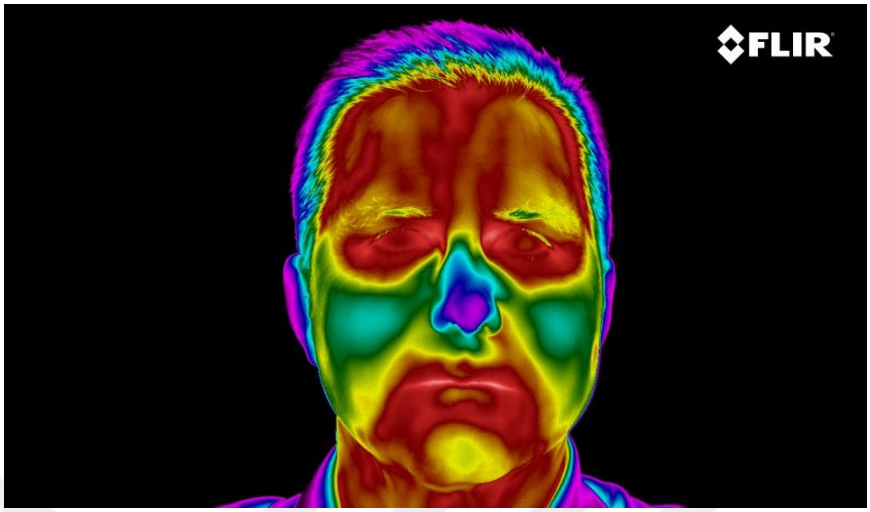


Figure 7. Sample FLIR SC620 IR Image

In our experiment, IR images were captured by a model SC620 FLIR thermal camera. A sample IR image taken by this camera is shown in Figure 7. SC620 FLIR thermal camera has sensitivity of less than 0.04 °C range (40mK) and captures standard −40°C to 500°C. In Figure 8, vascular representation of major vessels affecting the temperature of the face and thermal representation is presented. According to reaction to stimulus, temperature increases and decreases (Berkovitz, Kirsch, Moxham, Alusi, & Cheesman, 2013).

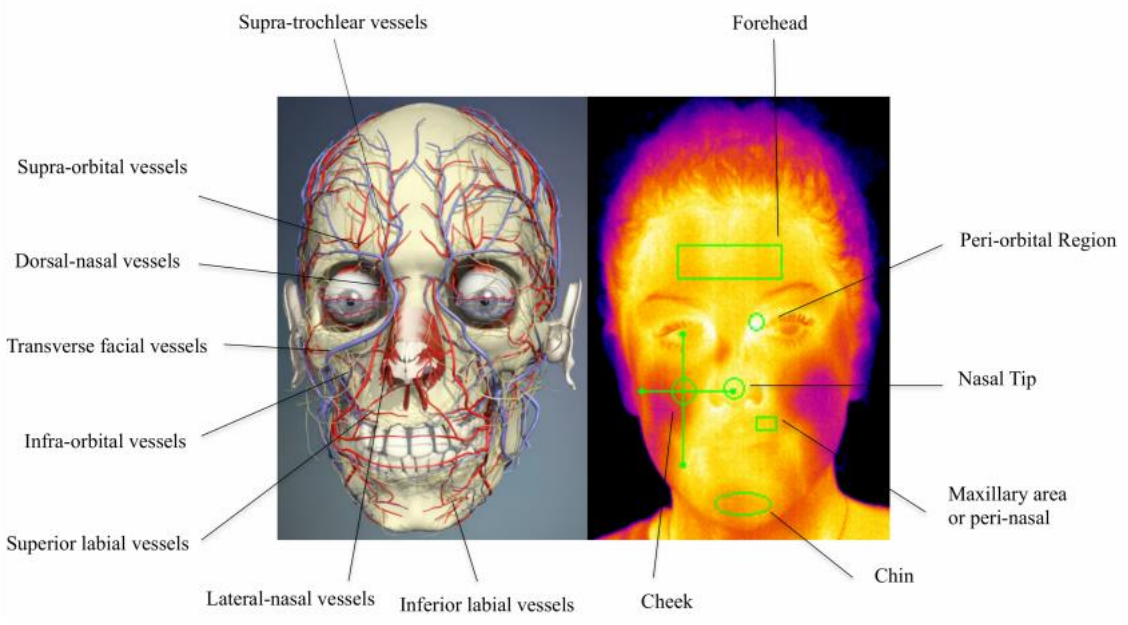


Figure 8. Vascular representation of major vessels affecting the temperature of the face and thermal representation (Berkovitz, Kirsch, Moxham, Alusi, & Cheesman, 2013)

2.4.Related Works in Stress Detection

Stress, emotion, pupil dilation and thermal imaging concepts were discussed in previous sections. The following literature survey in this section provide a brief view of relation between emotion/stress and physical/physiological signals.

Pedrotti et al. (2014) proposed a method to detect psychological stress from pupillary behavior. In their method, pupil diameter and electrodermal activity were recorded during a simulated driving task. The experiment consisted of one baseline run and three stress runs where subjects performed the driving task along with sound alerts. According to Pedrotti, pupil diameter indexed stress manipulation. In the study, a neural network algorithm was used as a classification method.

In the study of Pavlidis (et al., 2007), a system that incorporates physiological monitoring as a part of HCI was proposed. According to their bio-heat modelling of facial imagery, localized blood flow, cardiac pulse and breath rate could be extracted. In the study, in order to monitor stress, heartbeat irregularities, sleep apnea, localized blood flow, cardiac pulse, and breath rate signals were used. Experiment in the study showed that instantaneous stress brings an increase in the periorbital blood flow.

Effects of adrenaline in the body which is induced from stress cannot be suppressed by training. Dilation of pupil and increased feed of blood to muscles are examples for the adrenaline effects. A system for detection of stress (emotional or physical) remotely was developed by Yuen et al.(2009) based on Electro-Optics (EO) technologies such as thermal and hyperspectral imaging. Their results showed that areas such as periorbital areas, forehead, neck and cheek exhibited alleviated skin temperatures dependent on the types of stressors which were emotional or physical.

In the study of Zhai & Barreto (2006), a stress detection system was developed based on the physiological signals (Galvanic Skin Response (GSR), Blood Volume Pulse (BVP), Pupil Diameter (PD) and Skin Temperature (ST)) which were non-invasive and non-intrusive. These signals were monitored, analyzed and fused to detect emotional states between “stress” and “relaxed”. Their experiment system involved three stages. First one was an experiment setup for physiological sensing. Second was a signal pre-processing module for the extraction of affective features. Third was an affective classification (SVM method used) procedure. Correlation with monitored physiological signals and emotional state of their experimental subjects were noted. Moreover, pupil diameter was identified as the most significant emotional indicator.

De Santos Sierra (Sánchez Ávila, Guerra Casanova, & Bailador Del Pozo, 2011) proposed a stress detection system based on physiological signals which were GSR and HR. A specific psychological experiment on subjects was generated to acquire a database for training, validating, and testing the proposed system. The system was based on fuzzy logic and described the behavior of an individual under stressing stimuli in terms of HR and GSR.

Physiological responses caused by mental stress can be masked by variations due to physical activity. An activity-aware mental stress detection scheme was proposed by Sun et al. (2012). ECG, GSR, and accelerometer data were gathered from participants across three activities: sitting, standing, and walking. For each activity, physiological baseline was gathered while users were subjected to mental stressors.

High-resolution electro-optical and mid-wave infrared (MWIR) cameras and millimeter-wave radar systems can also be incorporated to identify stressed individuals.

In the study by Carl B. Cross (Skipper, & Petkie, 2013), a multimodal sensor platform was developed. In their experiment, while subjects were performing mental and physical tasks, registered image and sensor data were collected. Face was segmented into 29 non-overlapping segments based on fiducial points automatically in the images outputted by facial feature tracker. Chest displacement which was extracted from the radar signal and temperature fluctuations at the nose tip and regions near superficial arteries detected respiration and heart rates, respectively extracted from the MWIR image. All these extracted signals were fused in order to detect stress.

Psychological stress detection is also possible using subject-dependent bio signal features. According to Giakoumis (Tzovaras, & Hassapis, 2013), SC and ECG signals were analyzed and "rest signatures" were calculated from each subject's baseline recordings. These signatures were bio signal transformations capable to express each individual's baseline deviation from signal templates. According to Giakoumis, automatic stress detection accuracy was increased by their subject-dependent features extracted from SC and ECG signals.

In Sharma (Gedeon, 2014)'s work, an individual's response to real-life events was investigated. A computational model was developed to recognize observer stress using physiological and physical response sensor signals in real-life settings. Individual-independent SVM based model classifier was used to recognize stress patterns from observer response signals.

Use of non-invasive and unobtrusive sensors to measure and model stress is very common. According to Sharma (Gedeon, 2012), similar techniques for modelling stress were discussed. Sensors which could be usable in everyday activities were the focus of the study. Computational methods have the capability to determine optimal fusion and automate data analysis for stress recognition and classification. Several techniques have been developed to model stress based on Bayesian networks, artificial neural networks, and SVMs.

The studies discussed in this section, are summarized in Table 1. After we introduce the proposed method in the following chapters, we will compare these methods with our method in the discussion chapter.

Table 1. Emotion & stress detection studies.

Reference (Citation)	Stimulus	Measurements
(Ren, Barreto, Gao, & Adjouadi, 2013)	Stroop Color Word Test (Mental Stress)	PD, GSR
(Zhai & Barreto, 2006)	Paced Stroop Test (Mental Stress)	PD, BVP, GSR, ST
(Pedrotti et al., 2014)	Simple Driving Test with external stressful stimuli added (Psychological Stress)	PD, EDA
(Nhan & Chau, 2010)	Visual stimuli (IAPS) with varying arousal and valence content (Psychological Stress)	ST (facial)
(Yuen et al., 2009)	running exercise (physical stress) and Quiz (emotional stress)	ST (Facial)
(Giakoumis et al., 2013)	Video-game competition, arithmetic questions (Psychological Stress)	GSR, ECG
(Carl B. Cross et al., 2013)	Computerized version of the Stroop Color-Word Interference Test (Mental Stress),pedaling a recumbent exercise bicycle (Physical Stress)	ST and EO
(Sharma & Gedeon, 2014)	Interview experiment & Meditation experiment (Observer Stress for an observer of a real-life environment)	EEG, GSR and ST

CHAPTER 3

DATA COLLECTION METHOD AND EXPERIMENT

Stimulus generation steps and experiments are covered in this chapter. During the experiment, specialized IAPS pictures are displayed to participants for generating stress on them. The experiment has two parts, consisting of showing neutral and negative IAPS images respectively. While running first and second parts of the experiment, eye-tracker and infrared thermal camera recorded participant's pupillary and temperature data. Between the parts, Positive and Negative Affect Scale test (Appendix C) is administered. For the last part of the experiment, Debriefing Form (Appendix E) is filled out.

3.1 Signal Acquisition

In this study, pupil and thermal signals were collected using a TOBII TX300 Eye-Tracker and a FLIR SC620 IR Camera.

3.1.1. Hardware Setup: The complete instrumental setup of our experiment is shown in Figure 9. A TOBII TX300 eye tracker embedded in a 19" screen and a FLIR 620 IR Camera were used in the experiment. Each participant sat at 0.65 m distance from the TOBII screen and 1.15 m distance from the FLIR camera. Pupil dilation and facial thermal signals were recorded at a rate of 60 Hz and 30 Hz respectively.

Our experiment did not require a head restraint because real time gaze tracking (TOBII Technology, 2010) was available for pupil data and an in-house developed ROI (Region of Interest) tracker was employed during thermal data collection[1].

3.1.2. Software Setup: The experiment was controlled by an in-house desktop application. Multiple threads were generated for rapid data collection. After the experiment, all continuously recorded pupil and thermal values were saved into text files respectively. Our application analyzed the data using the steps illustrated in Figure 14.

[1] Thanks to Doğuş Türkay for the ROI tracker software.

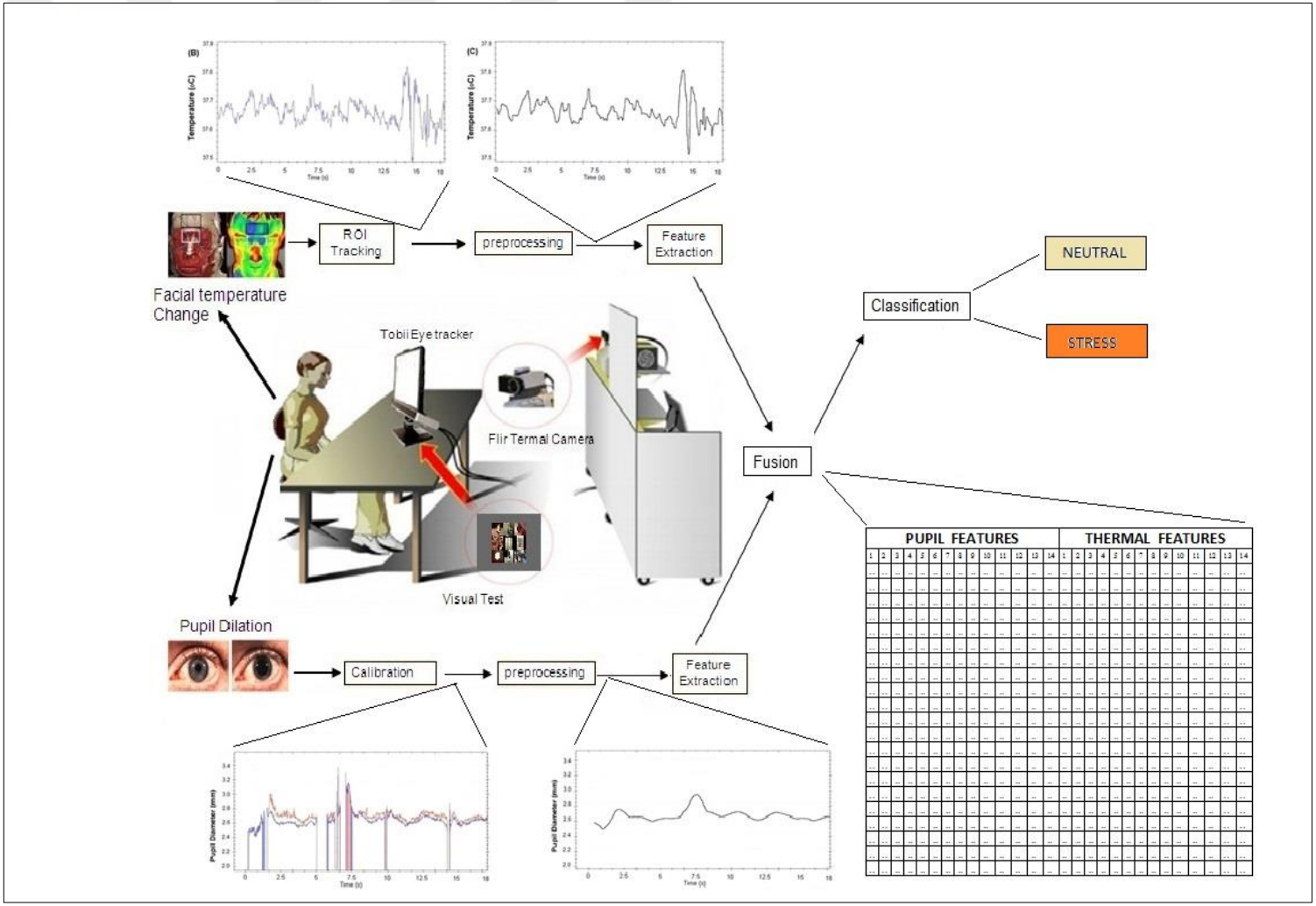


Figure 9. Experimental Setup

3.2 Participants

Participants can be chosen in many ways from the population. Our participants were selected from the circle of our acquaintances (school, work) randomly. Furthermore, individuals' availability and time constraints effected number of participants to be tested feasibly.

Eleven healthy subjects consented to participate in the experiment (age range: 29-40, 33±3.464, 2 females, and 9 males). The subjects had no pathological condition with their eyes or faces.

Informed consent was read and signed by the participants (Appendix F). The study was approved by the METU Ethics Committee (Appendix G) in conjunction with another data collection project which is an extended version of this thesis.

3.3 Experimental Procedure

Environment: Pupil size was taken in the stable light conditions to provide an accurate measurement. Moreover, for the thermal infrared image acquisition, room temperature was set to 19-22 ºC. There were humidity controller and an air recycling system in the building which experiment took place.

Experiment: The experimental design consisted of two parts. The first part was constructed as a baseline to measure neutral sentiment while the second part was designed to generate stress. Total number of trials was 20 for each part. Each trial consisted of a rest period of 12 seconds and a stimulus display of 6 seconds as illustrated in Figure 10.

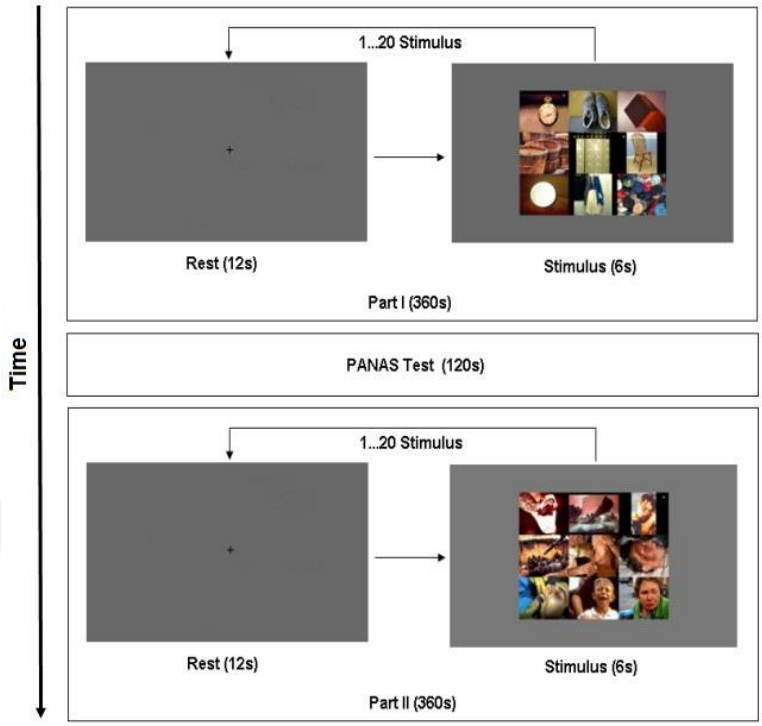


Figure 10. Experiment Flow

Part-I: Twelve neutral IAPS pictures were selected with low arousal values (2.77±1.896) and neutral valence values (4.949±1.185). Each stimulus consisted of a 3x3 grid of images randomly arranged from these set of 12 neutral images. Arrow symbols were embedded into some of the images in the grid randomly. Every stimulus was edited to contain a total of 3-6 arrows arbitrarily. Subjects were asked to identify the total number of arrows in the given display. Verbal responses were collected from the subjects during the rest period, which followed the picture display. The subjects were given immediate feedback for their response. If their prediction was correct, the experimenter said "Correct", otherwise "Wrong, the number of arrows is x". Sample stimulus image from Part I can be seen in Figure 11. All neutral stimuli including IAPS pictures are given in Appendix A.

Part-II: The experimental procedure was similar to Part I, except for arousal/valence of IAPS pictures, number of arrows, and verbal feedback. The chosen IAPS pictures had high arousal (6.12±2.02) and negative valence values (2.87±1.74). The number of arrows inserted in the display varied in the range of 6-9. The verbal feedback provided to the subject was misleading at a rate of 30%. On some of the trials where the subject's prediction was correct, the subject was deliberately told "Wrong, the number of arrows is x", as if the subject's response was not correct. Sample stimulus image from Part II can be seen in Figure 12. All negative stimuli including IAPS pictures are given in Appendix B.

Figure 11. Sample Stimulus Image from Experiment (Part I)

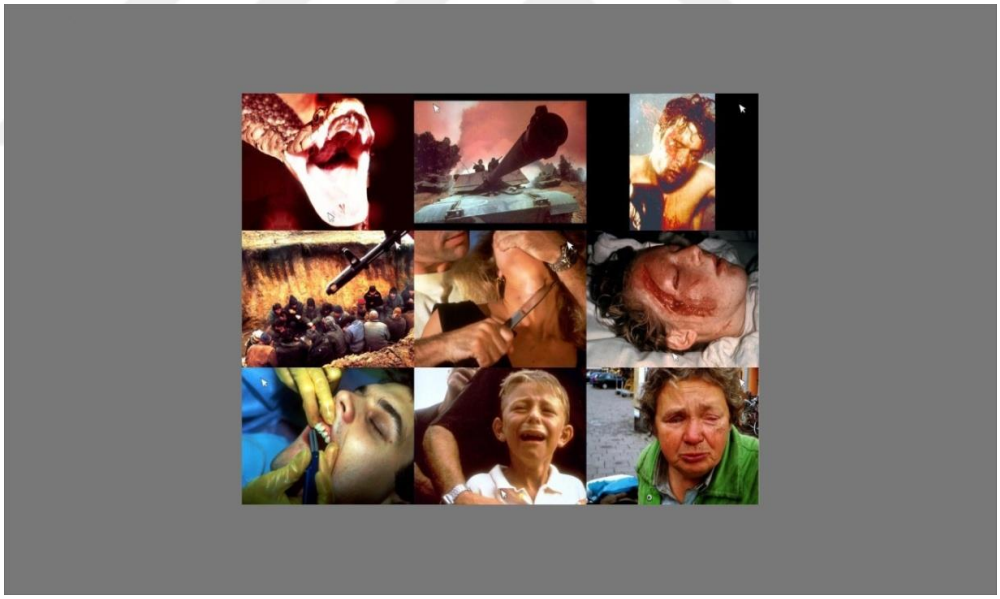


Figure 12. Sample Stimulus Image from Experiment (Part II)

Sensitivity of the method with respect to the size of the IAPS pictures and grid size were also investigated in our study. Only one IAPS picture, versus 3x3, 4x4, 5x5 grids of IAPS images were tested separately as stimuli. Results of pupil and thermal data for 4x4 and 5x5 grid of images were not meaningful across part 1 and part 2 of the experiment. Subjects' comments on the experiment also showed that 4x4 and 5x5 grid of images did not affect participants' emotion. On the other hand, only one IAPS picture as a stimulus affected emotions much less than expected. When the 3x3 grid was administered, the debriefing reports indicated that the subjects felt stress during the second part of the experiment. Therefore a 3x3 grid is chosen.

Before the initiation of the experiment, 9-dot calibration was applied. During the first phase, participants were shown 20 images on the center of grey background with 15° visual angle (see Appendix H for visual angle computation). Specifications of the TOBII eye tracker are summarized below.

Table 2. Specifications for Tobii TX300 eye tracker

Type	**Values**
Accuracy	0.4◦ At ideal conditions, binocular
Accuracy, Large angle	0.5◦ At 30 ◦ gaze angle, binocular
Precision	0.07 ◦ Without filter, binocular
Sample rate	300, 250, 120, 60 Hz
Sample rate variability	0.3%
Total system latency	<10 ms
Head movement Operation distance	37 x 17 cm. Freedom of head movement at a distance of 65 cm 50-80 cm
Max gaze angle	35 ◦

Many commercial eye trackers state an accuracy of about 0.5 degree, for example the eye tracker used in this thesis, Tobii TX300, has a majority of participants with an accuracy distribution of 0.4 degree during ideal conditions. Binocular data is the average of the two eyes. Ideal conditions for the system are when the users head stays in the middle of the eye-tracker, at a distance of 65 cm from Tobii and with an illumination of 300 lux in the room. Precision is stated to be 0.07◦ for binocular data without any filter. Precision is calculated as root mean square of successive samples.

Table 3. Specifications for FLIR SC620 Camera

Type	**Values**
IR resolution	640×480 pixels
Thermal sensitivity/NETD	<40mK @ +30°C
Field of View (FOV)	24° × 18°
Minimum focus distance	0.3 m
Focal Length	38mm
Spatial resolution (IFOV)	0.65 mrad
Image frequency	30 Hz
Spectral range	7.5–13 μm
Detector pitch	25 μm
Object temperature range	–40°C to +500°C
Accuracy	±2°C or ±2%

The FLIR camera was located 115 cm above floor was placed 100 cm away from the subject. In order to capture only the frontal views of face, the height of the camera was adjusted accordingly. Automatic focus of camera was always employed to during image recording. The IR images are 640x480 pixels and recorded on a laptop with the ThermoVision SDK at 30 frames per second. Images acquired with a FLIR SC620 thermal imager had resolution of 640x480 and NETD <40mK. The accuracy of SC620 is quoted as ±2°C or ±2% (whichever is the largest) where within the temperature range of 15-30°C. In Table 2, some of the system specifications are listed.

Changes in the experiment were made to assure that during the regular flow of the experiment, the subjects' stress would increase. This stress could have been due to the content of the IAPS pictures, the inability to count more arrows or the subjective belief that performance was low. All IAPS pictures in Part I and II were normalized for intensity values (96.66±0.279 out of 255) using Adobe Photoshop (Adobe Systems Inc., San Jose, CA; version CS3) software in order to avoid luminosity effects on eyes. Valence and arousal distribution of the pictures for each stimulus are shown in Figure 13 and Table 4.

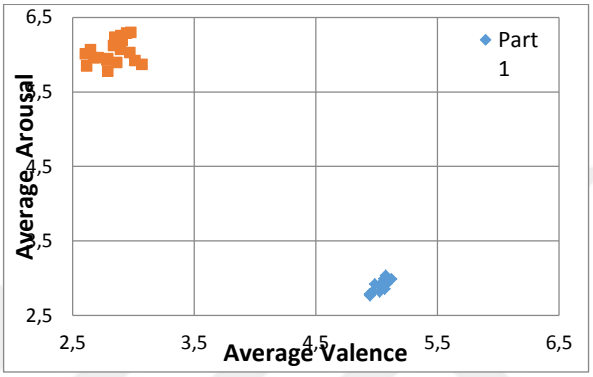


Figure 13. Average arousal and valence values of the IAPS Pictures used as stimuli

Table 4. IAPS Pictures' arousal and valence average values in Arousal & Valence Axes

	PART 1 (Neutral)			PART 2 (Negative)		
# of Stimulus	Arousal Average	Valence Average	Arrow	Arousal Average	Valence Average	Arrow
1	2.988	5.122	6	6.008	2.606	7
2	2.788	4.944	3	5.916	3.015	8
3	2.924	5.064	3	5.864	3.073	6
4	2.986	5.065	5	6.064	2.65	8
5	2.892	5.037	6	5.954	2.676	9
6	2.993	5.061	4	6.296	2.981	8
7	3.037	5.076	5	6.215	2.883	6
8	2.924	5.064	3	5.942	2.791	8
9	2.815	5.026	3	6.067	2.895	9
10	2.856	5.066	5	6.251	2.898	8
11	2.921	4.986	5	5.891	2.865	7
12	2.807	4.963	4	6.286	2.945	6
13	2.927	5.066	5	5.842	2.618	8
14	2.824	4.973	4	6.232	2.848	9
15	2.943	5.084	6	5.955	2.712	7
16	2.965	5.105	6	6.187	2.908	9
17	2.763	4.947	3	5.776	2.792	8
18	2.932	5.041	3	6.12	2.837	9
19	2.897	5.054	4	6.024	2.972	8
20	2.922	5.084	5	5.923	2.783	7
Averages	**2.905**	**5.041**	**4.4**	**6.041**	**2.837**	**7.75**

3.4 Procedure for Collecting Data

A survey (Appendix D) was conducted to get participants' personal information such as age, gender, education, health conditions, marital status, and eye glass status before the study. Especially who had health conditions like flu or cold which could affect the experiment were not allowed to take the experiment.

Between Part I and Part II of the experiment, subjects were provided with a chance to rest for a few minutes. During this period, they completed PANAS test[2] (Appendix C) in order to detect and exclude outlier subjects based on their current mood state. In the PANAS test, subjects read each item and then list how they feel by indicating the number on the scale next to each word. This scale consists of a number of words that describe different feelings and emotions.

One last survey (see Appendix E) was conducted to get participants' opinions about the experiment and to compare declared mental status of the participants with the measured values. Survey questions were designed for evaluating mental status of participants after being exposed to stressors (negative images, plenty of arrows and misguidance of participants)

Volunteering participants were taken one-by-one into the experiment room. The informed consent form (Appendix F) describing the kinds of tasks they would complete during sessions was filled out initially. Subjects were not informed about subsequent testing to prevent intentional learning. In the consent form, no mention was made of any forthcoming stress test. Instead, the subjects were informed about our investigation on positive and negative IAPS pictures.

[2]The Positive and Negative Affect Schedule (PANAS) test (Watson et al., 1988)

CHAPTER 4

DATA PROCESSING, FEATURE EXTRACTION AND CLASSIFICATION

There are three sections in this chapter. In the first section, data processing is explained. In the second section, feature extraction, feature definitions and feature selection procedures are covered. Third section is about the classification algorithms. Decision Tree, AdaBoost, Bagging and Random Forest classification methods are explained.

4.1. Signal Processing

We concatenated on the stimulus period and the rest period following immediately thereafter to assemble the response of each trial. Individual trials, consisting of 6 seconds stimulus and 12 seconds rest are referred in the following as one 'record'.

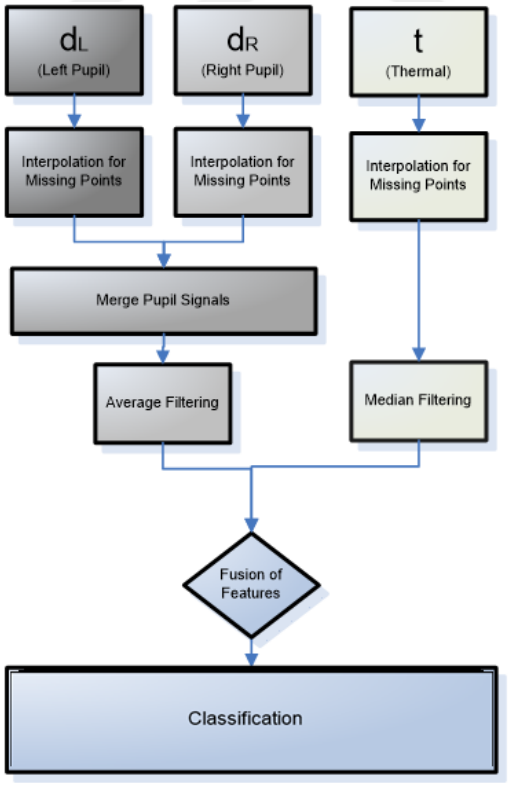


Figure 14. Pre-processing and Data Analysis Steps

In a record, pupil data is collected from the left and right eyes as $d_L=l_1, l_2,\ldots,l_k,\ldots, l_K$ and $d_R=r_1, r_2,\ldots,r_k,\ldots, r_K$ respectively, where k=1...K and K = 18 * 60. Thermal data is collected from the periorbital area as $t=t_1, t_2,\ldots,t_m,\ldots,t_M$ where m=1...M and M = 18 * 30. Total number of records is 20 for each part of the experiment. The pre-processing steps are presented in Figure 14

Pupil Data Processing: Due to eye blinks or insufficient tracking when the subject looks away, the gaze is lost. Therefore, there may be missing data points in pupil measurements. These data are marked as -1 by TOBII Eyetracker. An in-house extrapolation algorithm is used to compensate for such data loss. When there is missing data for only one eye, pupil data from the other eye is used for interpolation; otherwise, linear interpolation is done using the average of the last five samples before the lost data and the average of the three samples after the lost data.

At the very end, pupil diameter measurements are merged after testing whether left and right pupil diameters are highly correlated. For this purpose, the average of left and right pupils are computed and a single pupil data stream, $p_1, p_2, \ldots, p_k, \ldots, p_K$ is obtained. Finally, moving average filter (Witten & Frank, 2005) is used for data denoising. In this filter, small window size is specified and moved from the beginning to the end of a record. A window size of 20 sample points is found to be optimal as explained in chapter 5 and Figure 27. A sample result of this method can be seen in Figure 15.

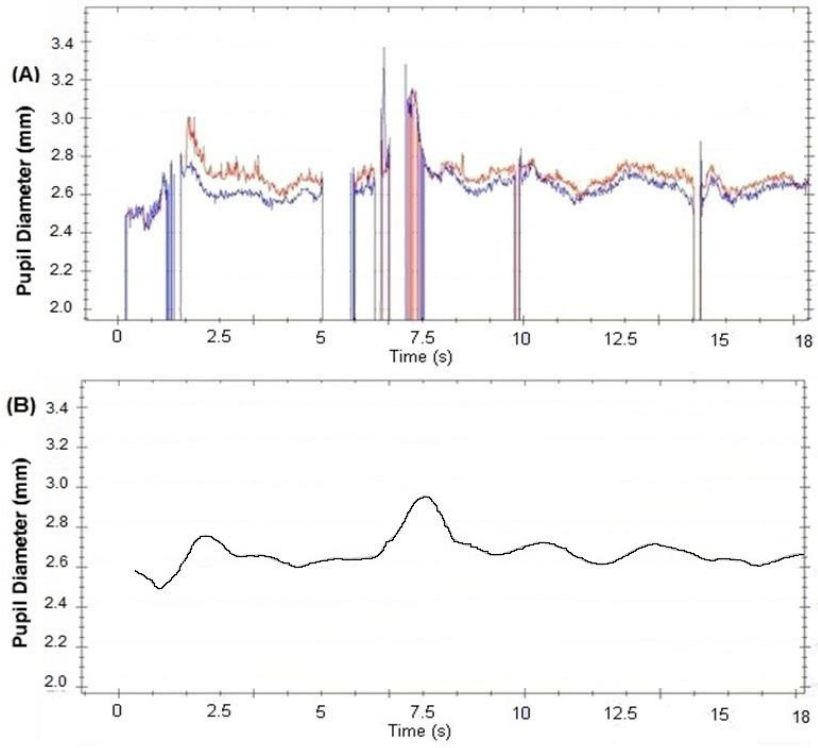


Figure 15. A. Pupil record before pre-processing, dL (blue) and dR (red) B. Pupil record after pre-processing, Interpolated and Merged Pupil Record

Thermal Data Processing: Each thermal data point is obtained from the mean temperature of the 10% hottest pixels within a pre-specified ROI. As suggested by (Pavlidis et al., 2007), before starting the experiment, for each subject we select the periorbital area that includes the vessels between the eyes (see Figure 16A). A tracking algorithm utilizing template matching registers this ROI throughout the course of experiment. Despite subject motion, 10% of the hottest pixels (Pavlidis et al., 2007) within this region are extracted for each frame. Hottest pixels' temperature values are averaged in order to get a single result, depicted by t_m which is used in feature extraction later. In rare events such as large motion, the thermal record is affected because tracking is lost. Hence interpolation is done to recover lost data. The same algorithm described for pupil diameter samples is used for interpolation of the thermal data as well. However, the moving average filter may become insufficient. Hence, to make a more robust estimate of the trend, simple moving median (SMM) is preferred for noise removal. In this method, the median is found by sorting the values inside a window and selecting the value in the middle. A window size of 15 points is found to be optimal as explained in chapter 5 and Figure 27. A sample record after preprocessing is seen in Figure 16.

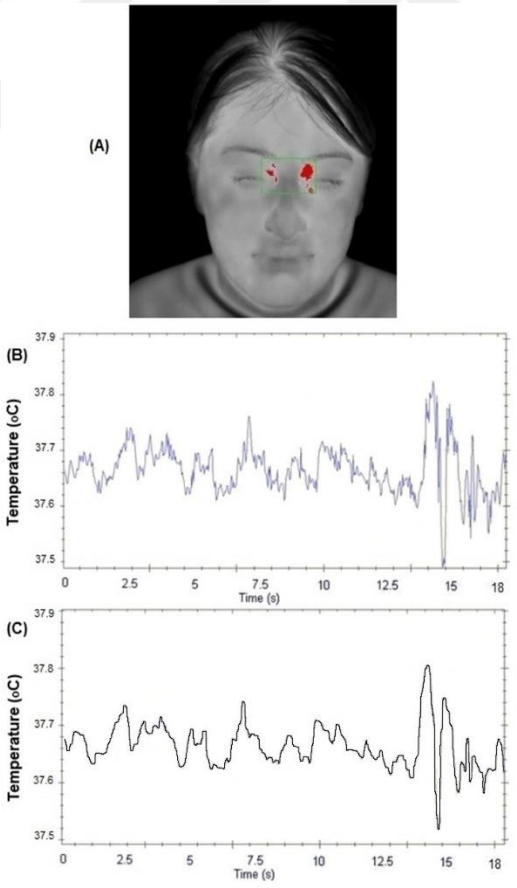


Figure 16. Thermal record after pre-processing

A. Sample Thermal ROI B. Thermal Data C. Processed Thermal Data

Afterwards, a standard outlier removal algorithm with median (Grubbs, 1969) is used in order to delete inappropriate data points. Especially in thermal signal acquisition, thermal tracking could be affected because of subject's abrupt movement. Omitting unwanted data and computing average before and after the outlier part is very important for classification at later steps. If ratio of missing values is higher than 30% in a record in both thermal and pupil data, this record is entirely omitted.

4.2. Feature Extraction

Original data is normalized before feature determination in order to get rid of signal variability between subjects. The signal values in each record are shifted to force all records to start with a value of 0. Overall averages of all records for all subjects are presented in Figure 17A and Figure 17B for pupil and thermal data.

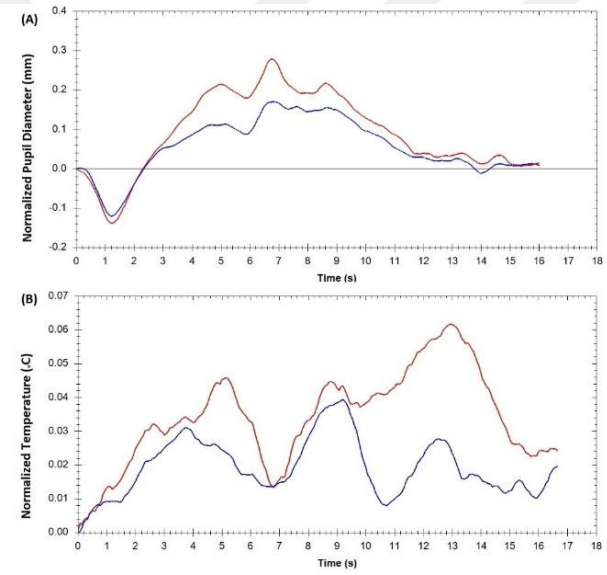


Figure 17. Pupil dilation and temperature record profiles (Blue:phase I; Red:phase II)

A. Average pupil diameters across all records of all subjects

B. Average temperature values across all records of all subjects

In order to generate an entropy-based representation of the pupil dilation or thermal signal in each record, Shannon entropy value is computed in a window of size W, which slides along the pupil and thermal data samples p_k and t_m (de Araujo et al., 2003). To calculate the Shannon entropy of such a signal, p_k and t_m are first discretized into N signal amplitude levels. Let j be the sliding offset of the window W, Shannon entropy of each window is then calculated by (Cover & Thomas, 2006):

$$H_j = -\sum_{n=1}^{N} s_n \log_2(s_n) \quad (1)$$

where j is an index, 1,2,..j,..J, through the entire record and s_n indicates the probability of each discretized signal level n, within the window W. J is different for pupil and

thermal records, since their sampling rates are different[3]. After calculating and combining entropy values, entropy series are created for every record. Sample pupil and thermal records with sliding window and entropy graphs are presented in Figure 18. Note that only the initial 6 seconds of the pupil measurement is extracted for the rest of the analysis. This choice is empirical, after observing that the initial 6 seconds of the pupil record is representative of neutral and stress class differences. This choice is in agreement with Bradley et al. (2008). The entropy graphs run shorter than the original records because of the window application. The amount that is truncated from the end is exactly the size of the chosen window, W.

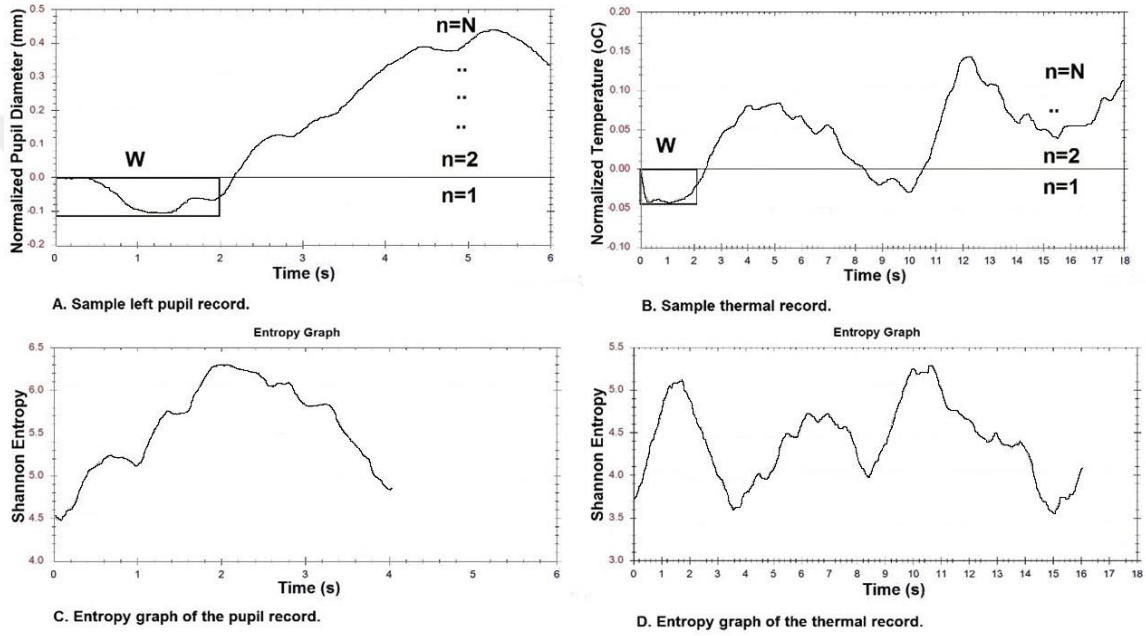


Figure 18. Shannon Entropy graphs of pupil and thermal records

When the window is positioned on an area, which contains a transition between neutral and stressed measurements, the entropy computed within the window will be high. On the contrary, when the window is positioned on an area, which embodies exclusively neutral signals or stressed signals, the entropy in this window will be low. Needless to say, the choice of the window size is crucial for emotion detection. In order to determine whether the transition detected by the entropy value is due to a change from a neutral state to stress or from a stressed state to neutral, the absolute values of the signals should also be considered.

Using the actual measurements and entropy calculations, several features were defined for each record. These features and their explanations are listed in Table 5.

[3]For pupil record, J is K minus length of W, for thermal record, J is M minus length of W

Table 5. List of features

ID. #	Feature Name	Explanation	Entropy (E)	Absolute (A)	Selected Features	
					Pupil	Thermal
1	Min	Minimum value of the record	✓	✓	-	(E)
2	Max	Maximum value of the record	✓	✓	-	(E)
3	Mean	Mean value of the record	✓	✓	(E)	-
4	Median	Median value of the record	✓	✓	(E)	-
5	SD	Standard deviation value of the record	✓	✓	(A)	(A)
6	Kurt	Kurtosis value of the record	✓	✓	-	-
7	Skew	Skewness value of the record	✓	✓	-	-
8	KLD1	Every record was split into 2 pieces. The Kullback-Leibler distance was computed in the first piece as KLD1		✓	-	-
9	KLD2	Every record was split into 2 pieces. The Kullback-Leibler distance was computed in the second piece as KLD2		✓	-	-
10	Slope 1	Every record was split into 2 pieces. Slope of the linear regressor in the first piece was computed as Slope 1.		✓	-	-
11	Slope 2	Every record was split into 2 pieces. Slope of the linear regressor in the second piece was computed as Slope 2.		✓	(A)	(A)
12	CurveCorrelation1	Timeseries of records from phase 1 were used to make one generic mean curve to represent all phase 1 signals. Correlation between each record and the generic phase 1 curve was computed as Curvecorrelation1.		✓	-	-
13	CurveCorrelation2	Timeseries of records from phase 2 were used to make one generic mean curve to represent all phase2 signals. Correlation between each record and the generic phase2 curve was computed as Curvecorrelation2.		✓	(A)	-
14	CurveCorrelationDif	Difference between CurveCorrelation1 and CurveCorrelation2 features of the record. This feature was expected to be positive for phase 1 records, negative for phase 2.		✓	(A)	-

Minimum, maximum, mean, median, SD, kurtosis, skewness are self-explanatory features. KLD, Slope, Curvecorrelation features will be explained in detail.

The **Kullback Leibler Distance (KLD)** is a natural distance function from a "true" probability distribution, P(i), to a "target" probability distribution, Q(i). In Figure 19, a sample pupil record is split into 2 pieces (X_1 and X_2).

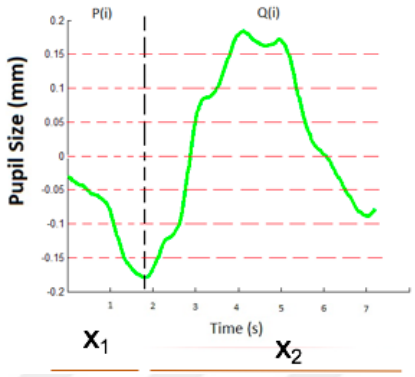


Figure 19. KLD feature of sample pupil record

P(i) indicates the probability of each discretized signal level n in X_1 and Q(i) indicates the probability of each discretized signal level n in X_2. The KLD was computed as;

$$KLD = \sum_{i=1}^{n} \mathrm{P(i)} \log \frac{P(i)}{Q(i)} \qquad (2)$$

Slope is another important feature. In Figure 20, a sample pupil record is split into 2 pieces (X_1 and X_2). Slope of the linear regressor (in the first piece) is computed as;

$$\mathrm{Slope}_1 = y_1/x_1 \qquad (3)$$

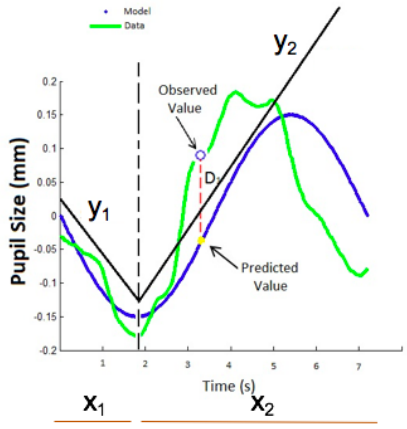


Figure 20. Slope and CurveCorrelation features of sample pupil record

Curvecorrelation is also an important feature. In Figure 20, sample pupil record is shown as green curve (observed data). Timeseries of pupil records from phase 1 are

used to make one generic mean curve to represent all phase 1 signals with Autocorrelation. This generic phase1 curve is shown as blue curve (predicted data) in Figure 20. Correlation values between a record and the generic phase1 curve are computed as Curvecorrelation1. D_1 is the Euclidean Distance between first observed data and first predicted data.

$$CurveCorrelation_1=D_1+D_2+\ldots+D_n \quad (4)$$

Correlation values between a record and the generic phase 2 curve are computed as **Curvecorrelation2**.

To reduce dimensionality of the large number of features, to remove noisy features and select significant ones, the embedded feature selection (CfsSubsetEval evaluator with BestFirst search) method in Weka[4] (Witten & Frank, 2005) is used. As explained in Weka, 'CfsSubsetEval' evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy among the chosen subset. 'BestFirst' searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility. Using this feature selection technique, we distinguish features that were best suited for separating the two classes (i.e. neutral and stressed) and removed all other features that are unproductive for the targeted classification. Ten best features are identified at the "Selected Features" Column for 'Pupil' and 'Thermal' sensors in Table 5. Features are later merged with Weka software to obtain feature level fusion.

4.3. Classification Algorithms

For classification of phase I versus phase II records, C4.5 (Decision Tree), AdaBoost with Random Forest, Bagging and Random Forest classification methods are used which are available in WEKA software[5]. WEKA is a comprehensive suite of Java class libraries that provides implementations of numerous classification, pre-processing, feature selection and prediction algorithms (Witten & Frank, 2005). For each classification algorithm the default parameters are used, as specified in Weka.

Decision trees are predictive models which map input attributes to a target value using simple conditional rules. The most notable and classical examples to decision tree learning are the algorithms ID3 (J. R. Quinlan, 1986) and the C4.5 (Quinlan, 1993). C4.5 constructs a decision tree from a set of data by dividing up the data according to the information gain, IG. It recursively splits the tree by the attribute with the highest IG in the training, yielding a decision tree that can be reused for classification (Quinlan, 1996). J48 method (Weka implementation of C4.5) has been used.

[4] The Waikato Environment for Knowledge Analysis (WEKA) Software

[5] Web Site: http://www.cs.waikato.ac.nz/ml/weka version 3.6

Random Forest is one of the most successful ensemble learning techniques (Breiman, 2001) which have been proven to be very popular and powerful in pattern recognition for high-dimensional classification (Meinshausen, 2006) and skewed problems. Random forest is a very useful method for categorical datasets. In random forest, there is a bunch of decision trees which create a forest structure. Each decision tree in the forest has got a maximum depth and nodes which contain split features. In training part of random forest, each split feature is picked from a random subset of the features. Instead of using the most discriminative thresholds, a random subset of features is used. Because of this randomness, the bias of the forest increases (Breiman, 2001).

AdaBoost is another popular ensemble method. It is used for prediction in classification tasks and reported to present self-rated confidence scores by estimating the reliability of their predictions (Witten & Frank, 2005). It is a learning algorithm used to generate multiple classifiers from which the best classifier is selected (Efron & Tibshirani, 1993). For the combination of AdaBoost with random forest (ABRF) technique, random forest utilized as a weak learner to generate the prediction models with less error rate.

Bagging (Bootstrap Aggregating) is also machine learning ensemble algorithm that is designed to improve stability and accuracy of algorithms used in statistical classification and regression. It helps to avoid over-fitting and reduces variance. Bagging is special case of the model averaging approach. Improvement in the accuracy of one model by using its multiple copies is goal of bagging method. Besides, average of misclassification errors on different data splits gives a better estimate of the predictive ability of a learning method.

In our classification approach, we used binary classifiers in order to discriminate the stress levels in the experiment. Every subject's data (phase I and II) are preprocessed and labeled as either in class 1 (neutral) or in class 2 (stressed). In order to obtain a more accurate and realistic assessment of the classifiers, a 10-fold cross validation method (Efron & Tibshirani, 1993) was used. First, the given data sets are separated into two parts: the training set and a test set, where the labels of the test set are considered unknown (Kohavi, 1995; Witten & Frank, 2005). In our study, the original data were first divided into 10 equal subsets, and one subset was tested using the classifier trained on the remaining nine subsets. This procedure was repeated until every subset had been used once for testing. The overall accuracy for the classifier is based on the average performance over the ten classification runs.

CHAPTER 5

RESULTS

In this chapter, analyses and classification results are reported. This chapter comprises of eight sections. In the first and second sections, statistical and classification results are presented. In the third, ROC results are reviewed. In the fourth section, parameter effects on accuracy, in the fifth section, experimental results are discussed. In the sixth section, similarity results of IAPS pictures' arousal/valence values and features (pupil and thermal) are presented. In the seventh section, analysis of significant features are reviewed and in the last section debriefing results are examined.

Analysis of data focused on the hypotheses which are expressed in the first chapter of study:

Hypothesis 1. Pupil size variation can be used in order to understand user's stress. At stressful times during our experiment, participants' pupil will dilate in comparison to other times.
Hypothesis 2. Stress can be evaluated by measuring thermal changes of facial area. At stressful times during our experiment, participants' thermal measurement will increase in comparison to other times.
Hypothesis 3. Fusing pupil and thermal features will increase success of stress detection.
Hypothesis 4. Using features that capture the difference of the rise and fall profiles in thermal and pupil signals will increase success of stress detection.

We assume that there exist emotional/stressful states of participants in our experimental setup which can be detected. Hence, the purpose of the data analysis was primarily to identify such states.

5.1. Statistical Results

The statistical analysis of pupil and thermal features in Table 6 was performed in SPSS Software. The first step for pupil and thermal datasets included tests for normality (Kolmogorov-Smirnov Test) to specify what type of statistical methods (parametric or nonparametric) should be used. These tests were completed for each stimulus of Part I

(neutral) and Part II (stress) on each dependent measure shown earlier in Table 5. The majority of features were non-normally distributed as shown in Table 7.

Table 6. Normality Results (bold and underlined features are normally distributed)

FEATURES	Normal Parameters		Test Statistic	Asymp. Sig. (2-tailed)
	Mean	Std. Deviation		
Pu_Ent_Min	5.050	0.449	.055	.010
Pu_Ent_Max	**5.940**	**0.383**	**.036**	**.200**
Pu_Ent_Mean	5.528	0.395	.061	.003
Pu_Ent_Median	**5.552**	**0.413**	**.042**	**.200**
Pu_Ent_Std	0.263	0.123	.078	.000
Pu_Ent_Kurt	-0.755	0.824	.129	.000
Pu_Ent_Skew	**-0.189**	**0.543**	**.030**	**.200**
Pu_Abs_Max	**0.285**	**0.181**	**.047**	**.055**
Pu_Abs_Std	0.138	0.046	.053	.016
Pu_Abs_Kurt	-0.744	0.699	.136	.000
Pu_Abs_Skew	**-0.107**	**0.512**	**.037**	**.200**
Pu_Abs_Cross1	1.652	1.912	.095	.000
Pu_Abs_Cross2	21.655	27.349	.210	.000
Pu_Abs_Slope1	-0.002	0.003	.072	.000
Pu_Abs_Slope2	**0.001**	**0.001**	**.039**	**.200**
Pu_KLD1	0.716	0.285	.202	.000
Pu_KLD2	0.729	0.273	.189	.000
Pu_KLD_1_2	**-0.014**	**0.057**	**.040**	**.200**
Th_Ent_Min	3.613	1.189	.143	.000
Th_Ent_Max	4.260	1.104	.153	.000
Th_Ent_Mean	3.952	1.139	.158	.000
Th_Ent_Median	3.962	1.140	.162	.000
Th_Ent_Std	0.182	0.108	.131	.000
Th_Ent_Kurt	-0.557	0.934	.162	.000
Th_Ent_Skew	-0.100	0.655	.046	.064
Th_Abs_Max	0.255	0.341	.251	.000
Th_Abs_Std	0.136	0.472	.396	.000
Th_Abs_Kurt	0.225	1.806	.212	.000
Th_Abs_Skew	0.018	0.751	.070	.000
Th_Abs_Cross1	-2.522	1.051	.058	.006
Th_Abs_Cross2	**161.008**	**52.900**	**.033**	**.200**
Th_Abs_Slope1	0.000	0.001	.208	.000
Th_Abs_Slope2	0.002	0.042	.446	.000
Th_KLD1	0.198	0.223	.052	.022
Th_KLD2	0.257	0.338	.080	.000
Th_KLD_1_2	-0.059	0.350	.049	.035

Since the features had non-normal distributions, we analyzed the data using non-parametric statistics. Using the nonparametric Mann–Whitney *U* test, for each feature

we compared difference between neutral versus stress data. Differences with the probability $\alpha \leq 0.05$ were regarded as significant.

Table 7. Mann–Whitney U test Results (bold and underlined features are significant)

Features	Mann-Whitney U	Z	Asymp. Sig. (2-tailed)
Pu_Ent_Min	**12531.000**	**-3.640**	**.000**
Pu_Ent_Max	**11386.000**	**-4.805**	**.000**
Pu_Ent_Mean	**11528.000**	**-4.660**	**.000**
Pu_Ent_Median	**11362.000**	**-4.829**	**.000**
Pu_Ent_Std	15885.000	-.229	.819
Pu_Ent_Kurt	14813.000	-1.319	.187
Pu_Ent_Skew	15058.000	-1.070	.285
Pu_Abs_Max	15406.000	-.716	.474
Pu_Abs_Std	**13185.000**	**-2.975**	**.003**
Pu_Abs_Kurt	**13942.000**	**-2.205**	**.027**
Pu_Abs_Skew	15518.000	-.602	.547
Pu_Abs_Cross1	**13589.000**	**-2.564**	**.010**
Pu_Abs_Cross2	**13437.000**	**-2.719**	**.007**
Pu_Abs_Slope1	15634.000	-.484	.628
Pu_Abs_Slope2	**9969.000**	**-6.246**	**.000**
Pu_KLD1	**13858.000**	**-2.291**	**.022**
Pu_KLD2	**12611.000**	**-3.559**	**.000**
Pu_KLD_1_2	**12051.000**	**-4.129**	**.000**
Th_Ent_Min	15555.000	-.565	.572
Th_Ent_Max	15704.000	-.413	.680
Th_Ent_Mean	15722.000	-.395	.693
Th_Ent_Median	15750.000	-.366	.714
Th_Ent_Std	14438.000	-1.701	.089
Th_Ent_Kurt	15752.000	-.364	.716
Th_Ent_Skew	15465.000	-.656	.512
Th_Abs_Max	**12997.500**	**-3.166**	**.002**
Th_Abs_Std	**11823.000**	**-4.360**	**.000**
Th_Abs_Kurt	16097.000	-.013	.989
Th_Abs_Skew	16025.000	-.086	.931
Th_Abs_Cross1	15557.000	-.562	.574
Th_Abs_Cross2	16072.000	-.039	.969
Th_Abs_Slope1	16101.000	-.009	.993
Th_Abs_Slope2	14398.000	-1.741	.082
Th_KLD1	15303.000	-.821	.412
Th_KLD2	16049.000	-.062	.951
Th_KLD_1_2	15622.000	-.496	.620

Sample description (n=359) and as p value less than 0.05, concluded that the underlined features provided statistically significant evidence of a difference between Part I and Part II. Because the data is normalized across the participants, we pooled all data from all participants together in this analysis. As shown in Table 7, some pupil and thermal features were obtained as a significant descriptor. With respect to this result, we concluded that the **Hypothesis 1** and **Hypothesis 2** were once again verified.

5.2. Classification Results

Common evaluation measures in classification problems are defined from a matrix with the numbers of examples correctly and incorrectly classified for each class, named **confusion matrix**. The confusion matrix for a binary classification problem (which has only two classes – positive and negative), is shown in Table 8. Performance of classifiers can be evaluated with the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

Table 8. Confusion Matrix

	Predicted Class	
True Class	Stress (Part II)	Neutral (Part I)
Stress (Part II)	TP	FN
Neutral (Part I)	FP	TN

Sensitivity is the percentage chance that the test will correctly identify Part II (stress) records which actually belong to part II.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

Specificity is the percentage chance that the test will correctly identify Part I (neutral) records which the test will identify Part I (neutral)

$$Specificity = \frac{TN}{TN + FP} \times 100\%$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100\%$$

Receiver Operating Characteristic (ROC) curve can be used for evaluating performances of classifiers. All sensitivity namely TP values are shown by a ROC curve. Equivalent values which are (1-Specificity) thresholds are on the y axis. Area Under Curve (AUC) which is the area under the ROC curve is generally thought as a key performance because a single quantity of overall accuracy is provided with this metric (Witten & Frank, 2005).

Several classification tests were conducted to investigate performances with respect to input data type (absolute signal value or entropy), or physiological/physical measurements (pupil or thermal). The outcomes of the classification, which are illustrated in Table 9, Table 10 and Table 11, show that on the average, the fusion of pupil and thermal features produce 10% increase in classification accuracy. Accuracy of 72.7% was reached using only pupil data features, 76.3% was reached using only thermal features and 83.8% was obtained using both pupil and thermal record features. Sensitivity and specificity values were highest (83.9% and 83.8%) with ABRF method using pupil and thermal data.

In Table 9, Table 10 and Table 11, all types of data results are shown according to classification methods. In these tables, SF labels are used in order to point Weka feature selection algorithm is processed before classification step. Best results are obtained via fusing pupil and thermal data features which are selected according to feature selection algorithm.

With respect to these results, pupil size variations and thermal changes were prominent separator for stress and neutral states considering Table 9 and Table 10. Stress detection is successful (accuracy higher than 70%) only as long as adequate features that reflect pupil or thermal data are included. As a result, we concluded that the **Hypothesis 1** and **Hypothesis 2** were verified.

Table 9. Thermal Data Classification Results, SF: Selected Features

	Decision Tree	**Random Forest**	**ABRF**	**Bagging**
Absolute Data Sensitivity	0.35	0.589	0.594	0.606
Entropy Data Sensitivity	0.35	0.517	0.511	0.533
Absolute & Entropy Sensitivity	0.656	0.644	0.717	0.689
Absolute & Entropy (SF) Sensitivity	0.528	0.744	0.772	0.756
Absolute Data Specificity	0.877	0.682	0.648	0.648
Entropy Data Specificity	0.737	0.553	0.542	0.559
Absolute & Entropy Specificity	0.709	0.799	0.754	0.743
Absolute & Entropy (SF) Specificity	0.81	0.754	0.754	0.754
Absolute Data Accuracy	0.612	0.635	0.621	0.626
Entropy Data Accuracy	0.543	0.534	0.526	0.545
Absolute & Entropy Accuracy	0.682	0.721	0.735	0.715
Absolute & Entropy (SF) Accuracy	0.668	0.749	0.763	0.754

Table 10. Pupil Data Classification Results, SF: Selected Features

	Decision Tree	**Random Forest**	**ABRF**	**Bagging**
Absolute Data Sensitivity	0.806	0.683	0.694	0.711
Entropy Data Sensitivity	0.683	0.594	0.617	0.611
Absolute & Entropy Sensitivity	0.844	0.728	0.772	0.756
Absolute & Entropy (SF) Sensitivity	0.817	0.744	0.744	0.761
Absolute Data Specificity	0.436	0.67	0.693	0.642
Entropy Data Specificity	0.413	0.536	0.553	0.564
Absolute & Entropy Specificity	0.436	0.631	0.626	0.581
Absolute & Entropy (SF) Specificity	0.48	0.687	0.709	0.67
Absolute Data Accuracy	0.621	0.676	0.693	0.676
Entropy Data Accuracy	0.548	0.565	0.584	0.587
Absolute & Entropy Accuracy	0.640	0.679	0.699	0.668
Absolute & Entropy (SF) Accuracy	0.649	0.715	0.727	0.715

Following **Hypothesis 3**, which stated that fusing pupil and thermal features could increase success of stress detection, we further checked whether the fusion of pupil size and facial thermal changes were descriptive of stress. In Table 11, best classification results (accuracy is higher 83%) was obtained by ABRF, Decision Tree, Bagging and Random Forest methods with fusing pupil and thermal features. With respect to this result, we concluded that the **Hypothesis 3** was verified.

Following **Hypothesis 4**, which stated that using absolute signal and entropy based features could increase success of stress detection, we further checked whether absolute signal and entropy based features were descriptive of stress. In Table 11, best classification result was obtained by ABRF method (accuracy is higher 83%) with absolute signal and entropy based features. With respect to this result, we concluded that the **Hypothesis 4** was also verified.

Table 11. Pupil & Thermal Data (Fusion) Classification Results, SF: Selected Features

	Decision Tree	**Random Forest**	**AdaBoost**	**Bagging**
Absolute Data Sensitivity	0.728	0.711	0.75	0.756
Entropy Data Sensitivity	0.639	0.594	0.644	0.628
Absolute & Entropy Sensitivity	0.778	0.806	0.839	0.822
Absolute & Entropy (SF) Sensitivity	0.694	0.817	0.839	0.839
Absolute Data Specificity	0.609	0.715	0.721	0.676
Entropy Data Specificity	0.531	0.598	0.564	0.575
Absolute & Entropy Specificity	0.782	0.793	0.793	0.737
Absolute & Entropy (SF) Specificity	0.771	0.832	0.838	0.81
Absolute Data Accuracy	0.668	0.713	0.735	0.715
Entropy Data Accuracy	0.584	0.596	0.604	0.601
Absolute & Entropy Accuracy	**0.779**	0.799	0.816	0.779
Absolute & Entropy (SF) Accuracy	0.732	**0.824**	**0.838**	**0.824**

According to the sensitivity results graph (see Figure 22), when ABRF and Random Forest algorithms were used for thermal and pupillary data with absolute and entropy based features, sensitivity greater than 80% was achieved. Moreover, when Decision Tree algorithm was used for pupillary data with absolute and entropy based features, more than 80% sensitivity was also reached. Moreover, using absolute and entropy based features together has higher sensitivity than using these features separately. In comparison with pupillary and thermal data, pupillary data has higher sensitivity. Besides, fusing pupil and thermal data improves sensitivity 5 percent considering ABRF, Bagging and Random Forest.

Specificity results graph (see Figure 23) shows that, using ABRF, Bagging and Random Forest algorithms on thermal and pupillary data with absolute signal and entropy based features (after feature selection), more than 80% specificity was reached. Furthermore, when Decision Tree algorithm was used for thermal data with absolute signal based features, more than 80% success was also achieved. In comparison with pupillary and

thermal data, thermal data has higher specificity. Besides, fusing pupil and thermal data improves specificity 5-10 percent considering ABRF, Bagging and Random Forest.

According to the accuracy results graph (Figure 24), using ABRF, Bagging and Random Forest algorithms on thermal and pupillary data with absolute and entropy based features (after feature selection) more than 80% success was achieved. As a result, ABRF performed best accuracy among other methods. Using absolute signal and entropy based features together increases accuracy (5%) more than using these features separately. In comparison with pupillary and thermal data, pupillary data has higher accuracy. Besides, fusing pupillary and thermal data improves accuracy (10-15%) significantly.

Feature selection algorithm increases accuracy 5 percent, sensitivity 2-3 percent among ABRF, Bagging and Random Forest algorithms. Unfortunately, feature selection had bad effect on Decision Tree method considering sensitivity, specificity and accuracy results. This warrants further investigation.

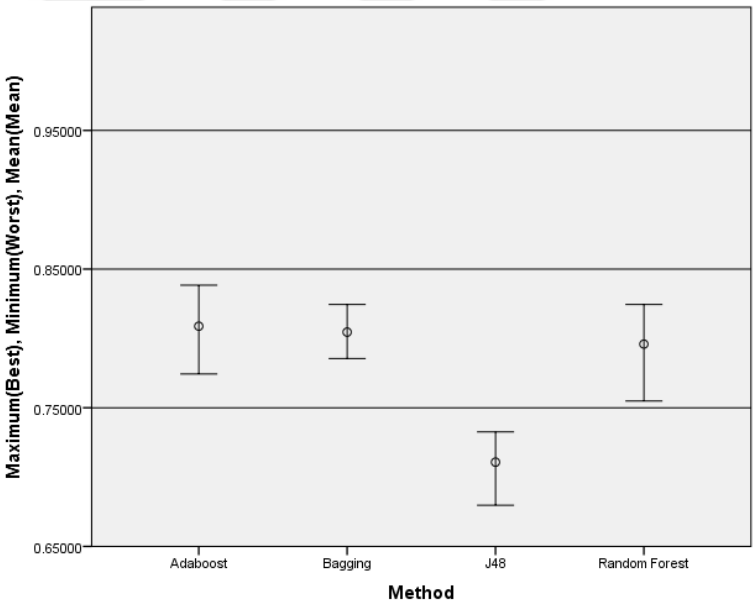


Figure 21. Classification Algorithm's Accuracy Results

In comparison with classification methods, Figure 21 shows best and worst classification results of ABRF, Bagging, Random Forest and Decision tree algorithms. ABRF and Bagging results are very close but best accuracy is reached with ABRF. Random Forest method has also high accuracy results but its variation is greater than other methods. In addition, Decision Tree algorithm has 10-15% worse accuracy than other algorithms.

Ensemble learning methods (ABRF, Bagging, Random Forest), combine predictions of multiple learning algorithms that is why they often lead to a better predictive performance than a single learner (Decision Tree). They are well-suited when small differences in the training data produce very different classifiers. Main drawbacks are computation time and lower interpretability.

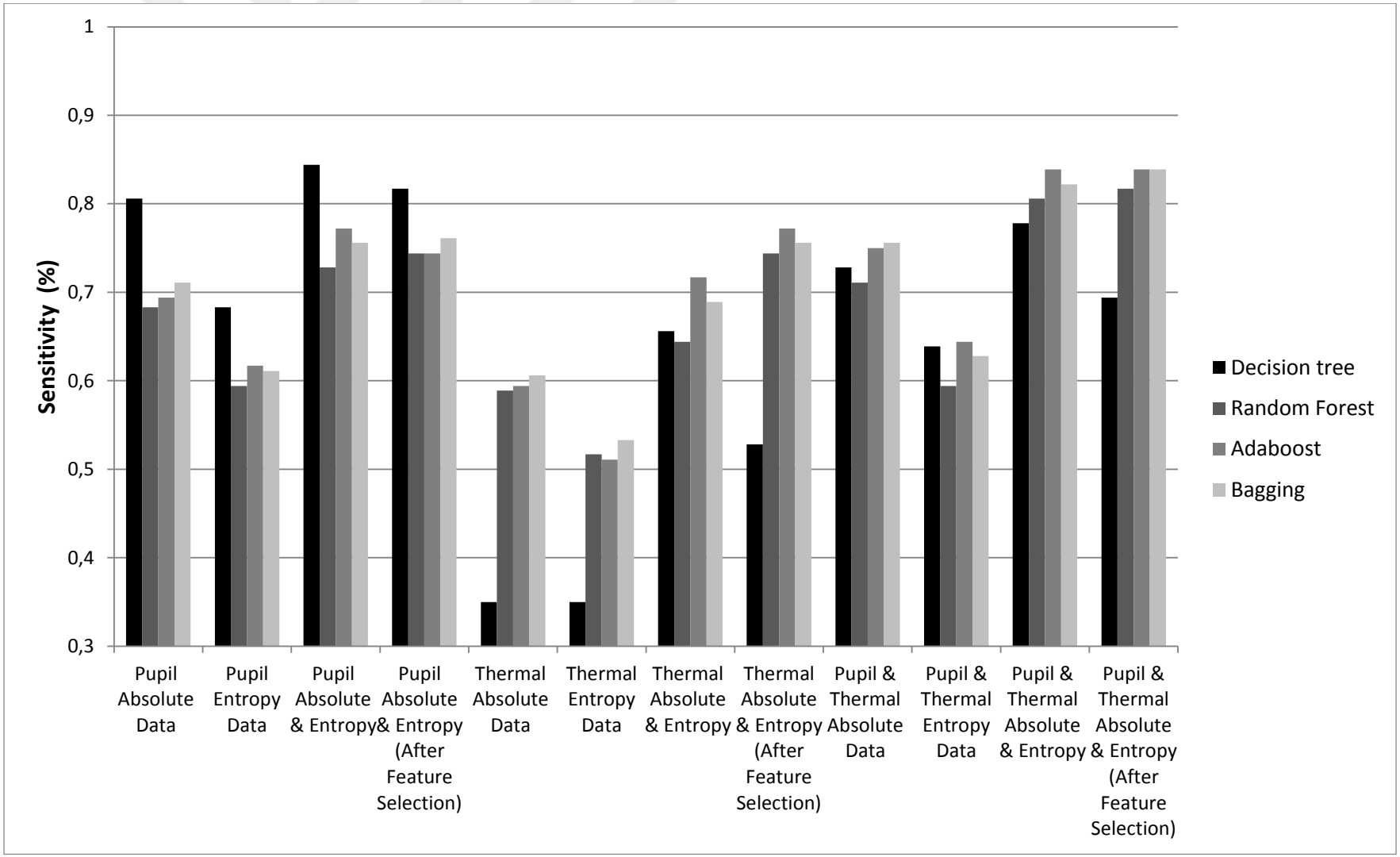


Figure 22. Sensitivity Results

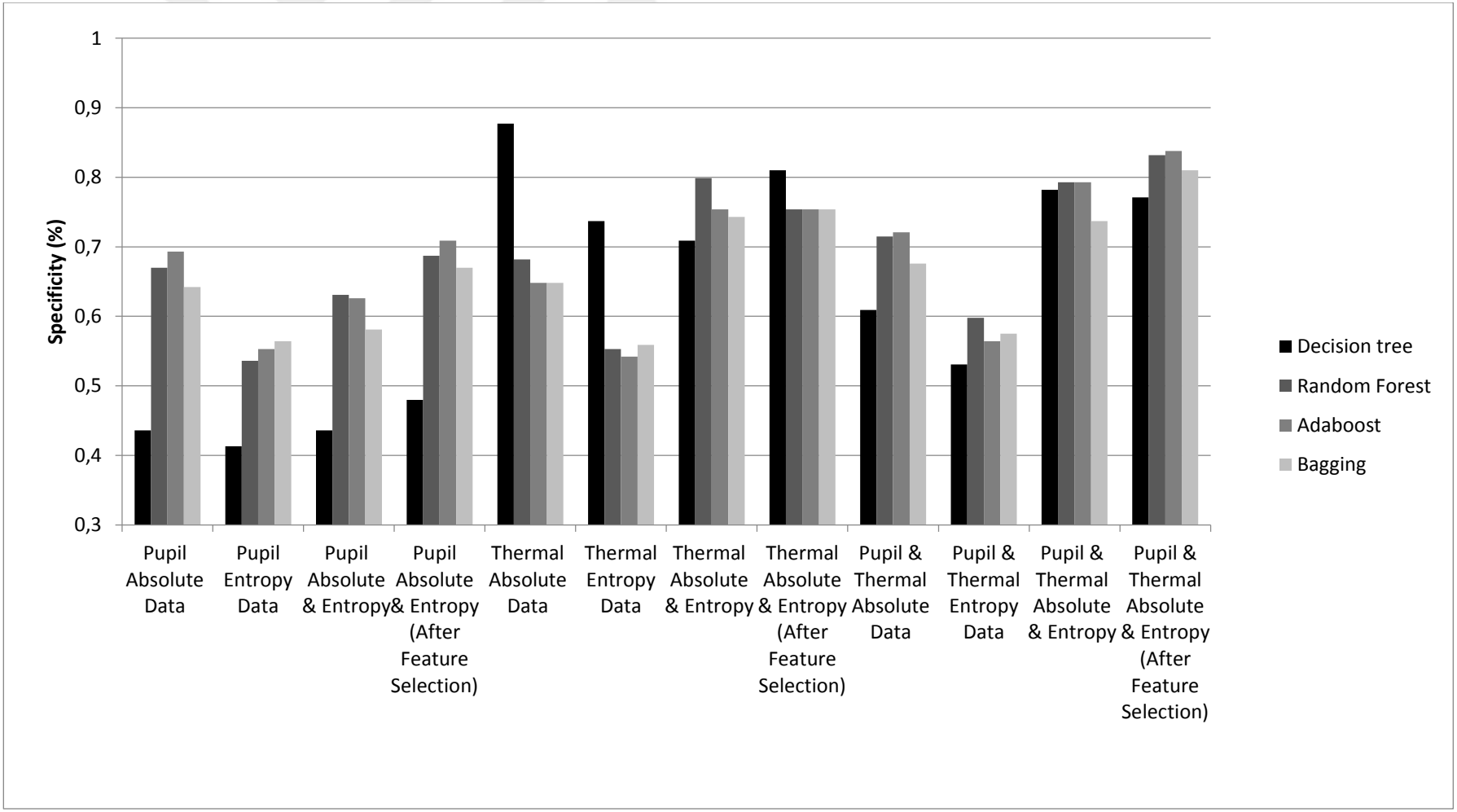


Figure 23. Specificity Results

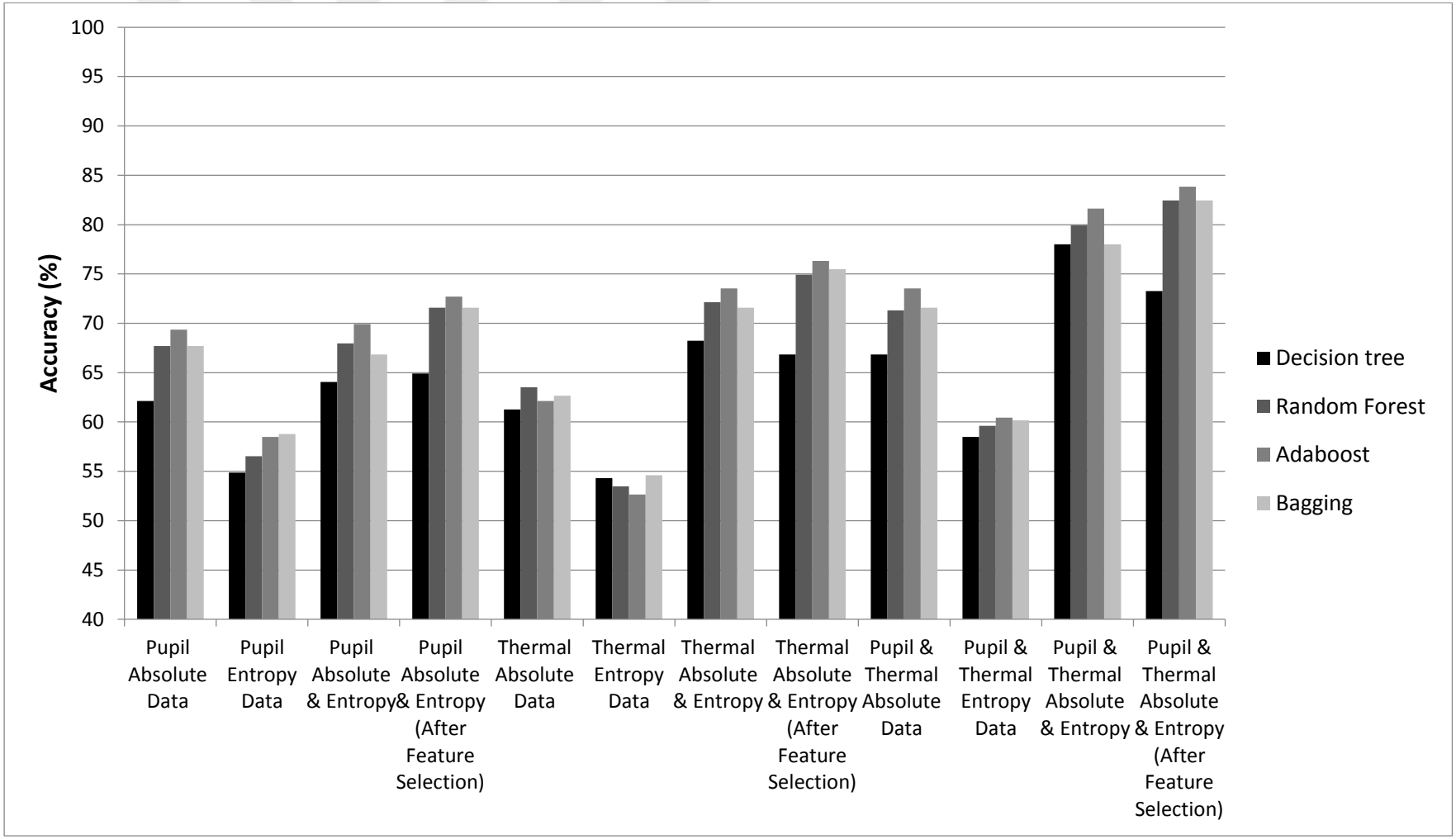


Figure 24. Accuracy Results

5.3. Receiver Operating Characteristic (ROC) Curve Results

Classifiers performances can be evaluated in terms of the area under the Receiver Operating Characteristic (ROC) curves (graphical plots of correct detection rate versus false alarm rate for a binary classifier system as its discrimination threshold is varied).

Figure 25 illustrates the predictive performance of the four classifiers, ABRF, Bagging, Random Forest and Decision Tree with pupil and thermal features. The results show that ABRF, Bagging and Random Forest method performs relatively well compared to Decision Tree method in terms of ROC curve.

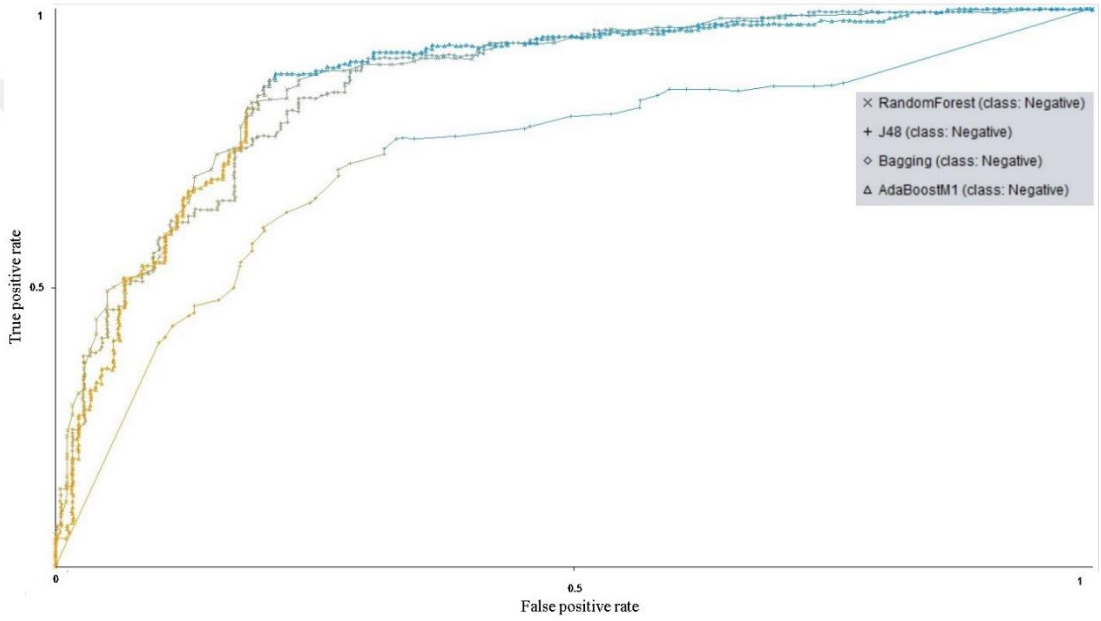


Figure 25. ROC of classification methods for all features

The AUC is an effective and combined measure of sensitivity and specificity. AUC scores of our study can be seen in Figure 26. The best score was obtained using the Pupil and Thermal Features (Weka feature selected) with Random Forest and ABRF classification methods. Fusing pupil and thermal features increased ROC AUC for our classification methods. And also feature selection method increased ROC AUC except Decision Tree algorithm. Perhaps, Weka's feature selection methods were not good/compatible with the way Decision Tree works.

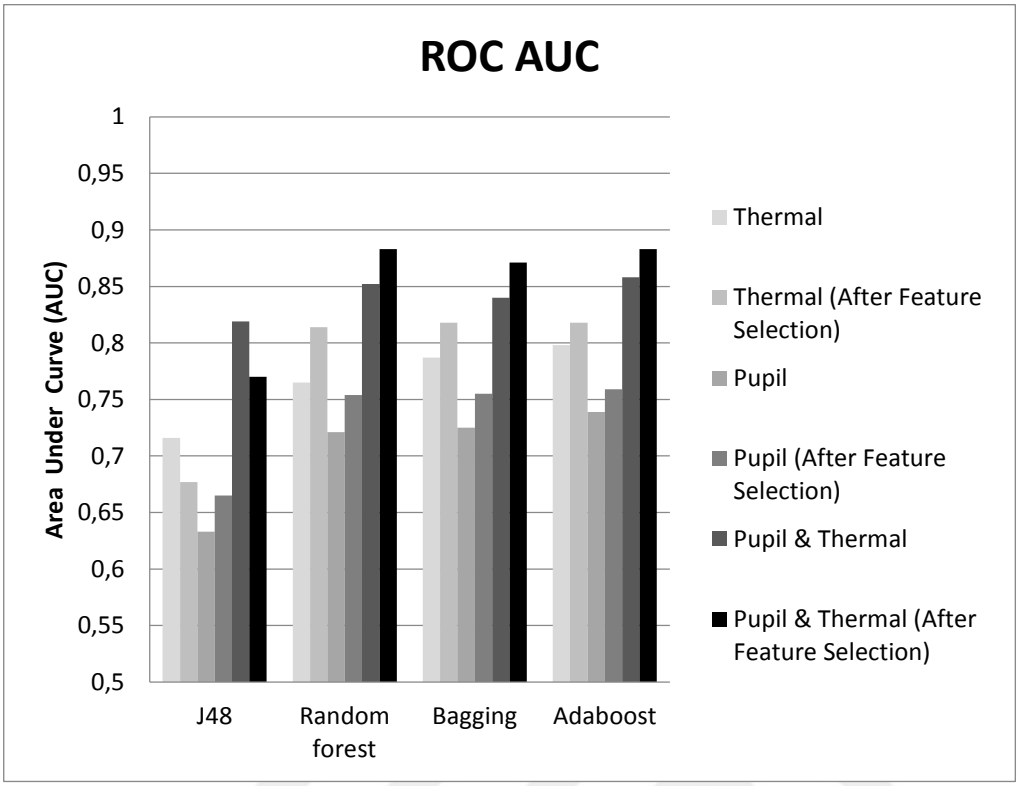


Figure 26. Area under curve scores

5.4. Effects of Parameters on Accuracy

Classification performance scores were reached after separate investigation of three parameters for pupil and thermal records in classification success: entropy window size ($W_{entropy}$), slope window size (W_{slope}), moving average filter window size (W_{moving}). Because of the different physiological windows of pupil dilation and facial temperature, different optimum parameters were determined and used for classifications. In Figure 27, accuracy results are shown when two parameters are fixed but one parameter is manipulated for pupil and thermal data. After careful inspection of Figure 27, we preferred the following parameters in our classifications: W_{slope}: 120 time points for the pupil and 200 time points for the thermal signal,$W_{entropy}$:200 time points for the pupil and 160 time points for the thermal signal, W_{moving}: 20 time points for the pupil and 15 time points for the thermal signal. In Figure 27, darker areas indicate higher accuracy and lighter areas indicate lower accuracy in classification results. It is important to note that when all variables are varied at the same time, optimum results may differ. Furthermore, the choice of W_{slope} and $W_{entropy}$ affected the accuracy only around 4%, but the choice of the W_{moving} was more important, because this changed the performance around 20%.

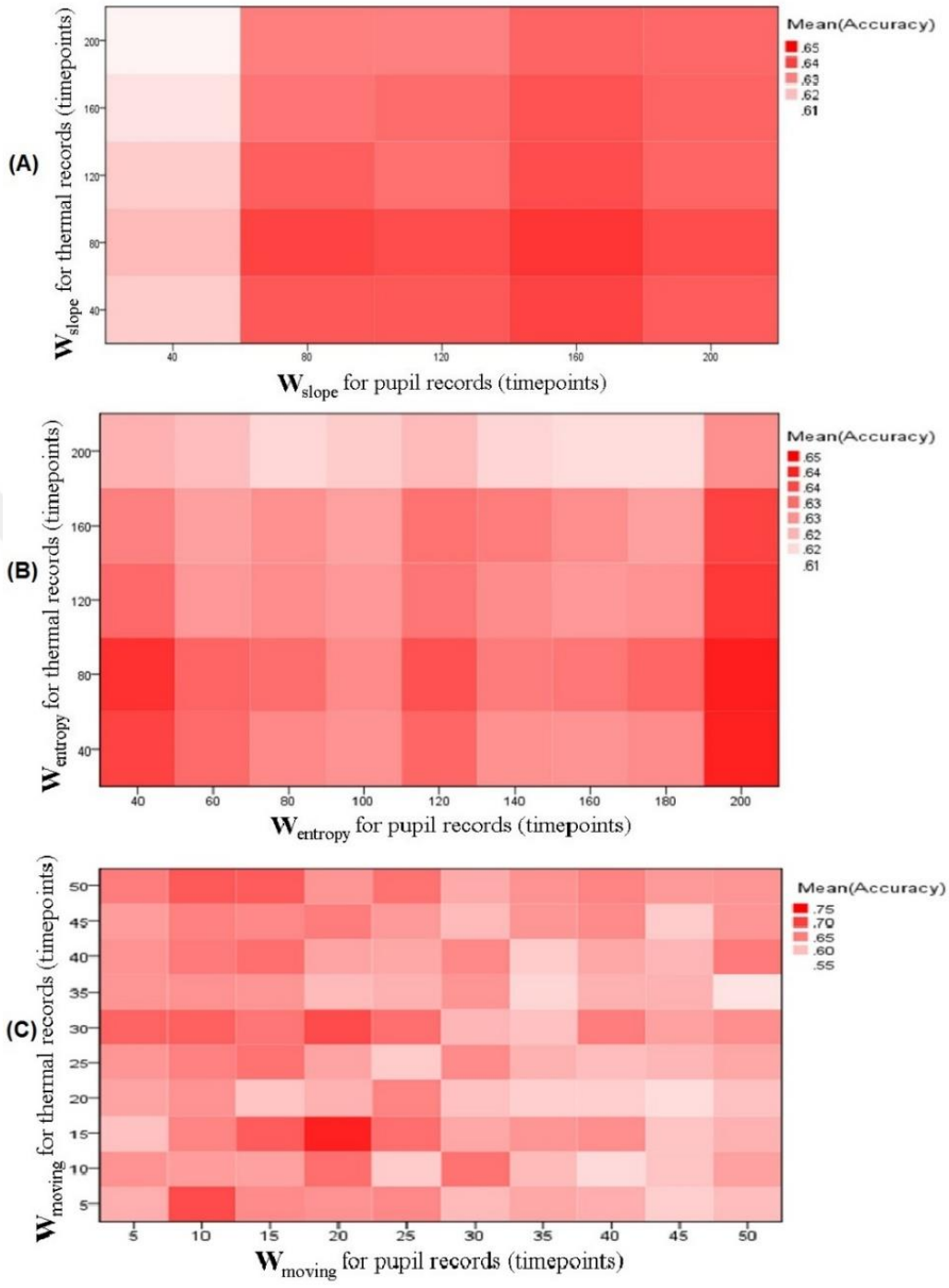


Figure 27. Effect of parameters on classification results (X and Y-axes show varying window size for the pupil and thermal records respectively) (A) slope window size (B) entropy window size (C) moving average filter window size

5.5. Experimental Results

A total of 11 subjects participated, but 2 of them were excluded because more than 30% of their records were missing due to eye-blinks. All participants had completed the experiment once. Acceptable values of PANAS (Watson, Clark, & Tellegen, 1988) scores are such that, for a positive mood the mean was 29.7 (SD: 7.9), for a negative mood the mean was 14.8 (SD: 5.4). Our average positive score was 29.85 and negative score was 15.14, which were in the acceptable ranges for all healthy subjects. Correlation between the right and left pupil diameter samples was $r = 0.876$, $p < 0.01$ therefore merging right and left eyes into a single record was justified.

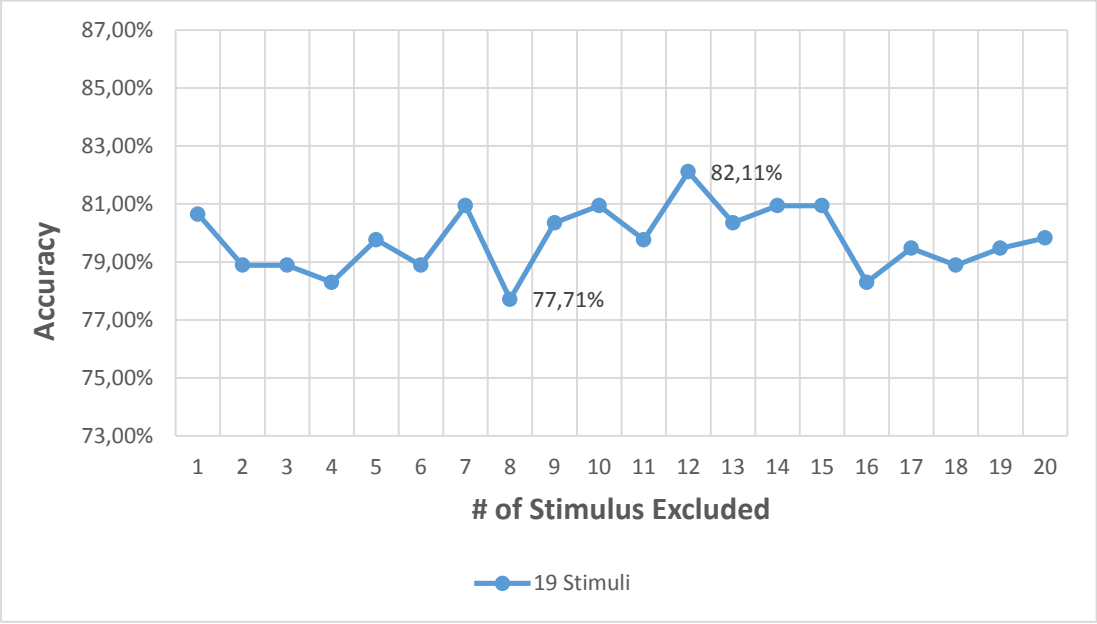


Figure 28. Stimulus Effects on Accuracy Results, number of stimulus (one at a time) was excluded. (In comparison when all (20) stimuli used, accuracy is % 83.8)

The classification results are investigated with respect to effects introduced by individual stimuli and participants' in order to analyze the homogeneity of findings. In Figure 28, for a particular method (weka feature selection and ABRF classification) the effect of stimulus on thermal and pupil data is demonstated by excluding one stimulus at a time. Accuracy results varied between 77.71% and 82.11%. In this figure, best result was obtained with all stimuli included, but twelfth stimulus excluded. Twelfth stimulus had the minimum effect on overall accuracy and eighth stimulus had the maximum effect. Every stimulus was useful for overall accuracy (83.85%) so we concluded that every stimulus was not necessary for our experiment. On the basis of each subject, the results are presented in figure 33, at the end of this chapter. Third participant had the worst stress detection accuracy. Third participant was the oldest subject in our group.

5.6. Similarity Results

All features in Table 5 were tested for validity by running cosine similarity method in SPSS Software (Statistical Package for the Social Sciences, Version 20; IBM statistic tool, NY, USA) as follows. In each stimulus, the 9 IAPS pictures' arousal/valence values were averaged in order to produce a single arousal/valence value. Then the entire pupil data was pulled together for each feature. The cosine similarity between this value and the average arousal/valence value were computed. This process was repeated for thermal data and thermal features. Some of the features had high similarity values (>0.6) as shown in Figure 29. For the pupil measurements, the features that had high similarity values (>0.6) to IAPS ratings were: Mean (Entropy), Median (Entropy), SD (Absolute), Slope2 (Absolute) and CurveCorrelation2 (Absolute). For the thermal measurements, the features that had high similarity values (>0.6) to IAPS ratings were: Min (Entropy) and Max (Entropy). As seen from Figure 29, among the ten best features selected by Weka, only three features had low cosine similarity values.

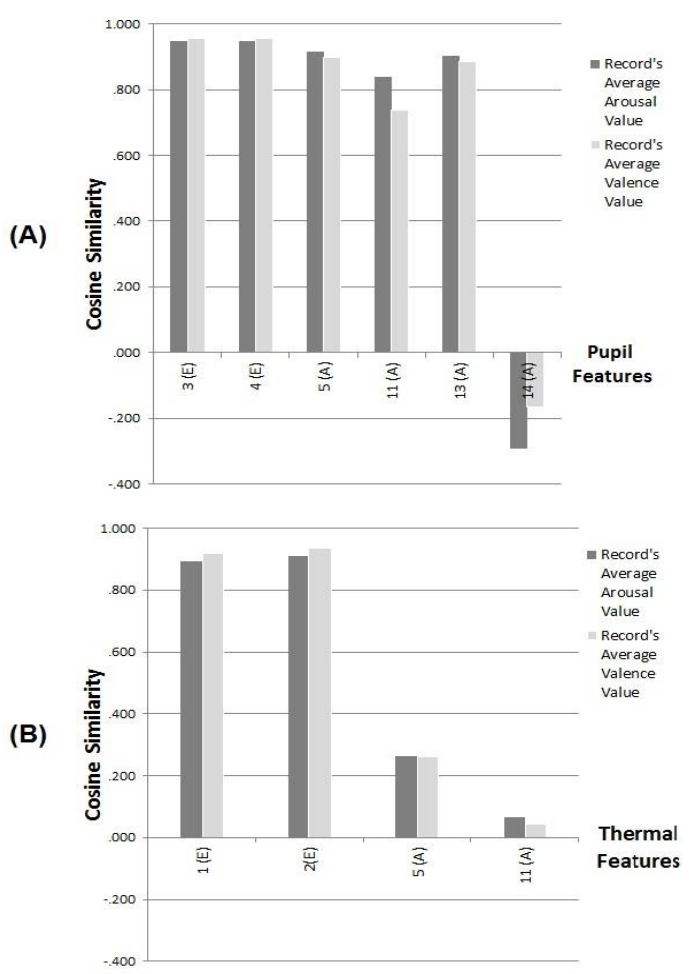


Figure 29. Verification of validity of selected pupil and thermal features with IAPS valence and arousal values (X-axis contains reference numbers of features from Table 5). Cosine Similarity between stimulus average arousal/valence values and selected pupil (A) and thermal (B) features

5.7. Analysis of significant features and classification results

Significant features (from Table 7) and selected features (from Table 5) are illustrated in Table 12. Features that are common (painted) were studied in detail and some of these common features were plotted (Figure 30, Figure 31 and Figure 32).

Table 12. Comparison of Selected Features and Significant features

ID. #	Feature Name	Pupil Features		Thermal Features	
		Selected (Weka)	Significant (SPSS)	Selected (Weka)	Significant (SPSS)
1	Min	-	(E)	(E)	-
2	Max	-	(E)	(E)	(A)
3	Mean	(E)	(E)	-	-
4	Median	(E)	(E)	-	-
5	SD	(A)	(A)	(A)	(A)
6	Kurt	-	(A)	-	-
7	Skew	-	-	-	-
8	KLD1	-	(A)	-	-
9	KLD2	-	(A)	-	-
10	Slope 1	-	-	-	-
11	Slope 2	(A)	(A)	(A)	-
12	CurveCorrelation1	-	(A)	-	-
13	CurveCorrelation2	(A)	(A)	-	-
14	CurveCorrelationDif	(A)	(A)	-	-

According to Figure 30, Figure 31 and Figure 32, “pu_ent_median”, “pu_ent_mean” and “pu_abs_slope2” features in Part I (green) clearly differed from Part II (blue). However, for the third and twelfth stimuli, discrimination of Part I and Part II was not clearly possible according to these three features.

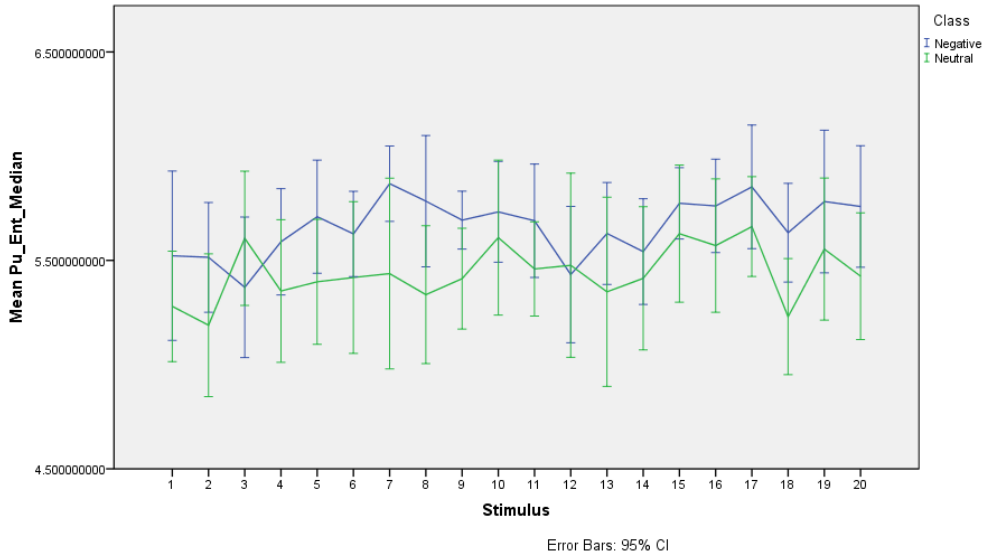


Figure 30. Part I and Part II feature (Pupil Median) difference for all participants

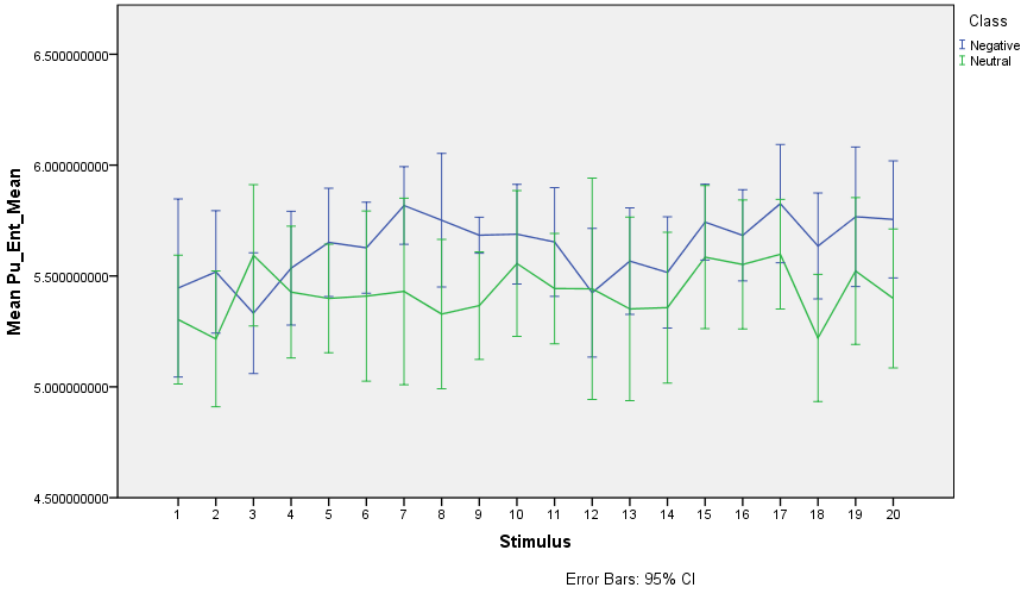


Figure 31. Part I and Part II feature (Pupil Mean) difference for all participants

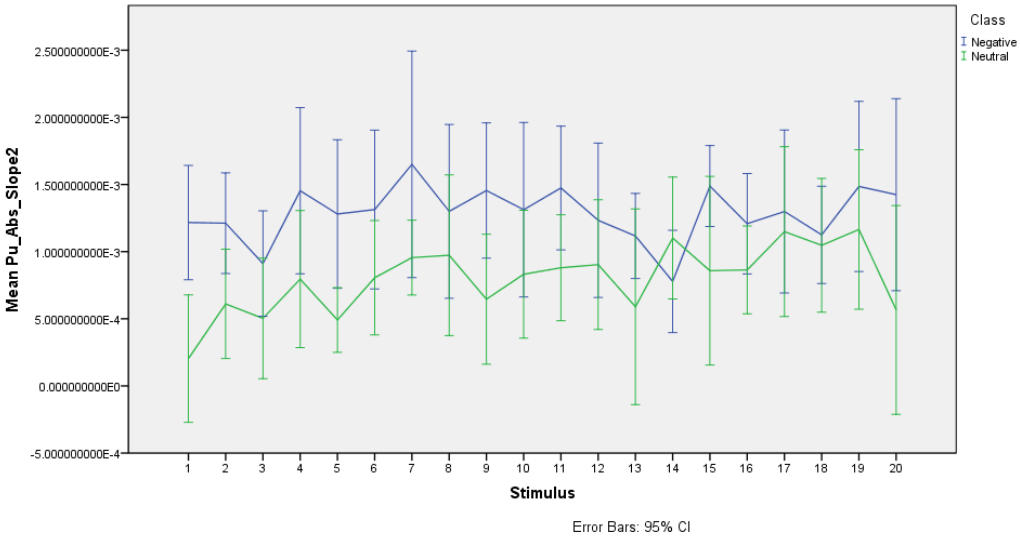


Figure 32. Part I and Part II feature (Pupil Slope2) difference for all participants

5.8. Debriefing Results

Each participant was interviewed and debriefed on the true purpose of the study. According to participants' Debriefing Form (Appendix E), second part (negative) of our experiment was found more stressful and disturbing than the first part. Moreover, viewing negative images resulted in loss of concentration and inability to find and count arrows correctly in time. In Figure 33, participant's data was analyzed with ABRF method. Except for subject 33, the classification of each participant was higher than 77%.

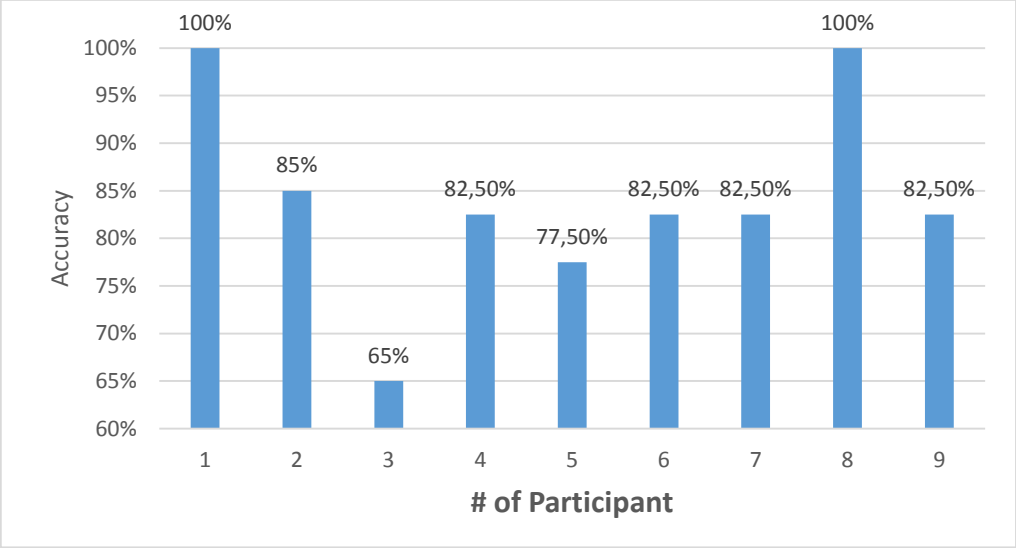


Figure 33. Participant's Effect on Accuracy Result.

CHAPTER 6

DISCUSSION

In the literature, the success rates of the stress detection through pupil dilation and facial thermal signals are contingent upon cumbersome offline data processing. When pupil dilation is concerned, according to Ren (Barreto, Gao, & Adjouadi, 2013) the best average accuracy of "relaxation" versus "stress" states of the computer user, through the monitoring and processing of the pupil signal is 83%. In their study, several pre-processing steps such as Kalman filtering, Wavelet denoising, Walsh transform were required which prevented real-time dynamic pupil analysis. On the other hand, our methods are applicable to real-time analysis with little effort.

Pedrotti et al. (2014) proposed a method for relating pupillary behavior to psychological stress. During a simulated driving task, pupil diameter and electrodermal activities were recorded. In that study, Neural Network classifier was able to reach 79.2% precision whereas our precision result was 83.9% considering pupil and thermal data.

According to Zhai et al. (2005), in order to monitor the stress state of computer users based on Blood Volume Pulse, Galvanic Skin Response and Pupil Diameter signals, 80% accuracy was reached with SVM method. Zhai & Barreto (2006) also implied that, non-invasive, non-intrusive real-time assessment of the affective state of a computer user could be achievable and by using four physiological signals: BVP, GSR, PD and ST, acceptable levels (up to 90.10%) of differentiation between "relaxed" and "stressed" states could be observed. Proposed system was unsuitable for real-time and remote sensing.

Researches show that, electrodermal activity signal has less discriminating power compared to pupil signals for stress classification (Pedrotti et al., 2014; Ren et al., 2013; Zhai & Barreto, 2006).

When facial thermal signal is concerned, classification accuracy of nearly 80% was achieved with thermal facial imaging while classifying baseline versus high arousal and valence levels (Nhan & Chau, 2010). In that study, five ROIs were tracked so classification required extra work. In another study (Khan et al., 2006), infrared

thermal sensing of positive and negative affective states was achieved with manual specification of facial thermal feature points (FTFP) resulting in 66.3-83.8% accuracy rates. FTFP required human intervention, so its use seemed to be limited.

In Yuen et al. (2009)'s study, to classify physical and emotional stress based on facial ST, elevation of temperatures in the facial region, particularly in the forehead, periorbital, eardrum and cheek regions were used. Remarkable increase of skin temperature in the prefrontal region was seen at the onset of emotional stress, while increased number of hot pixels in the periorbital region was monitored under physical stress. Our study was similarly based on thermal imaging in periorbital region. Quantitative assessment techniques were not implemented by Yuen et al. (2009) which made this study subjective and its results incomparable.

According to Cross et al. (2013)'s study, mental and physical stresses were remotely detected using ST and EO. Accurate heart and respiration rates could also be sensed from both thermal and radar signatures. Because of processing so many complex thermal facial tracking methods and real-time detection objective, linear classification method (LDA) was preferred. On the other hand, Cross et al. (2013) also claimed that classification of physical and psychological stressor was greater than 90%.

Giakoumis et al., (2013) proposed a subject-dependent emotion recognition system to detect psychological stress via GSR and ECG. In this work, video game competition was the stimulus and LDA classifier with sequential backward search (SBS) feature selection method was used and 95% accuracy was reached. Major disadvantages of this work were using subject dependent experiment and their system didn't work real-time and remotely. However, our system works remotely and near real-time.

Observer stress for an observer of a real-life environment is a new concept created by Sharma & Gedeon, (2014). Observer stress was detected based on EEG, GSR and ST. Genetic algorithm was used for feature selection to build a SVM classifier and observer stress was recognized with an accuracy of 98%. The outcome of this research was to predict human stress response to real-life environments. Proposed system was unsuitable for real-time, remote sensing and did not identify actual stress.

In comparison with these studies (see Table 13), our method has the advantage to classify data in near real time[6], once training is accomplished offline. Obviously, in a real time scenario, the pre-processing steps as well as the classification procedure introduces a lag. In certain situations, where the entire thermal record is necessary, this lag can be as high as 18 seconds, but generally, we estimate that classification on the fly can be achieved with a lag of 10-15 seconds. However, the eye tracker and

[6] Real-time processing may sometimes jeopardize the data capturing process. Therefore, thread management should be optimized accordingly.

thermal camera we used, are sensitive, sophisticated and a little bit expensive measurement devices that applications may have difficulty in providing.

In accordance to our hypotheses, we found that stress could solely be predicted using either pupil diameter or facial temperature over 70% accuracy. With respect to the **Hypothesis 1**, when we categorized our participants' results into two parts, namely neutral and stress, pupillary responses allowed us to classify emotion (stress). That is why we can easily say that pupillary response differs with respect to the emotional state the subject is currently in. Relation between IAPS pictures arousal, valence values and pupillary responses was demonstrated using cosine similarity method.

With respect to the **Hypothesis 2**, participants' stress was evaluated over 70% accuracy based on the variations in facial temperatures. At stressful times during our experiment, participants' facial temperatures raised in comparison to other times. Our results show us that stress can be evaluated by measuring thermal changes of facial area. As a result, facial thermal variations could also be an indicator of the stress.

Our analysis further showed that when applying fusion techniques on thermal and pupil features, stress classification accuracy was enhanced over 83.8%. The success rate of our method is on the high end in comparison to other methods in the literature. The accuracy rates we have obtained show the complementary potential of the of pupil and facial thermal recordings

This result supported **Hypothesis 3**, expressing that using pupil and thermal features together will increase success of stress detection. Moreover, when using entropy based features stress classification accuracy was 60% while using absolute signal features made accuracy 73.53%. Fusing entropy based and absolute signal features enhanced stress detection accuracy over 83.84%. This result supported **Hypothesis 4**, using features that capture the difference of the rise and fall profiles in thermal and pupil signals will increase success of stress detection.

Furthermore, all participants declared that in the second part of our experiment, they were more affected and disturbed from negative IAPS pictures which made the subject miscount the arrows in pictures. Therefore we can say that we have behavioural evidence regarding the success of our method, both in stress induction and stress detection.

Table 13. Review of emotion & stress detection studies.

Reference (Citation)	Stimulus	Measurements	Classification Methods	Reported Accuracy	Disadvantage
Our Study	Visual stimuli (IAPS) with varying arousal and valence content (Psychological Stress)	PD, ST (facial)	ABRF Bagging Random Forest Decision tree	% 83.8 % 82.8 % 81.0 % 74.2	Doesn't work real time.
(Ren et al., 2013)	Stroop Color Word Test (Mental Stress)	PD, GSR	Multilayer Perceptron and Naive Bayes classifier	% 84.21 accuracy	Doesn't work real time and remotely So many pre-processing steps such as Kalman filtering, Wavelet denoising, Walsh transform are needed.
(Zhai & Barreto, 2006)	Paced Stroop Test (Mental Stress)	PD, BVP, GSR, ST(finger)	SVM Decision tree Naive Bayesian network	% 90.10 % 88.02 % 78.65	Doesn't work real time and remotely
(Pedrotti et al., 2014)	Simple Driving Test with external stressful stimuli (Psychological Stress)	PD, EDA	wavelet transform and neural networks	% 79.2 (precision)	EDA doesn't correlate with self-reports Doesn't work real time and remotely
(Nhan & Chau, 2010)	Visual stimuli (IAPS) with varying arousal and valence content (Psychological Stress)	ST (facial)	Fisher LDA classifier with Genetic Algorithm	% 66.3- 83.8	human intervention needed (Manual specification of thermal feature points) many medications (analgesics, antidepressants, antihypertensives, antispasmatics, melotonin, and niasin), motion artefacts and loss of focus can affect thermal results

(Yuen et al., 2009)	Running exercise (physical stress) and Quiz (emotional stress)	ST (Facial)	None	None	No quantitative assessment. Subjective results.
(Giakoumis et al., 2013)	Video-game competition, arithmetic questions were also asked (Psychological Stress)	GSR, ECG	LDA-based classifier and sequential backward search (SBS) feature selection.	95%	Doesn't work real time and remotely Subject dependent
(Carl B. Cross et al., 2013)	Computerized version of the Stroop Color-Word Interference Test, (Mental Stress) Pedaling a recumbent exercise bicycle (Physical Stress)	ST and EO	LDA ANN and SVM	100% 90%	Sensitive to individual movements Doesn't measure emotional stress
(Sharma & Gedeon, 2014)	Interview experiment & Meditation experiment (Observer Stress for an observer of a real-life environment.)	EEG, GSR and ST	Individual-independent SVM Genetic algorithm for feature selection	98%	Doesn't work real time and remotely Doesn't identify actual stress. Identify only environmental stress

CHAPTER 7

CONCLUSION

In the scope of this thesis, we conducted an experimental study on detecting stress via measurement of the physiological responses, pupil dilation and facial thermal changes. In the experiment we designed, IAPS pictures that varied in pleasure and arousal axes were used as stimuli. Our results showed that negative contents reliably evoke greater pupillary response and facial temperature compared to neutral pictures.

The most important contribution of our study is increased stress detection performance by feature-based fusion of pupil dilation and facial thermal changes. The results of the research outlined show a promising relation between the stressful state of stimulus and the physiological signals monitored. An important characteristic of the stress detection method suggested here is that, it is generic for capturing physiological signals with variable rise and fall profiles. This advantage is attained through the use of entropy which is robust for all kinds of physiological sensors.

Overall, the best results were obtained by the fusion of both pupil and thermal data features. Furthermore, classification with ABRF method outperformed classification with other methods. By using ABRF with both absolute and entropy-based features, classifications' rates in the range of 80-84 percent accuracy were reached.

Considering that the measurement of pupil and periorbital temperature can be made unobtrusively, the method proposed herein for stress detection stands out with its feasibility and speed. The pre-processing, feature extraction and classification methods implemented in this thesis are fast, robust and can be applied to real-time emotion classification.

7.1. Future Work

In this study, the experiment takes place in near real time. Making all process in real time will be our future work. Besides, using deep learning technique can increase the accuracy of the stress detection and can be seen another next goal.

In our experiment, IAPS pictures with high arousal values were used after neutral ones. In real time scenarios, this is not possible. Because real time stimuli evolve

dynamically. Feature extraction and classification for dynamic stimuli would make the framework presented in this study more applicable.

An experimental limitation of the procedure mentioned in this thesis was that we used a template matching algorithm in order to track the head. Advanced tracking methods like particle filtering might help us to acquire more sensitive and robust data from thermal camera that can increase accuracy of classification.

A future direction is to involve the Emotional Quotient Test results of the participants in emotion detection. The utility of this test for increasing the performance is stress detection remains to be investigated.

Other classification methods should also be examined and their potential to achieve even a higher accuracy in the detection of stress must be evaluated. One of the most common methods of classifying data, SVM remains to be explored. Additionally, larger collections of experimental data need to be gathered, to allow for the development of a stronger classifier.

The experiment we have designed reveals a limited emotional palette compared to a dynamic multi-media experience with unpredictable shifts between neutral and negative emotions. Testing the promise of the chosen features for a dynamic emotional environment is necessary to identify the actual real-life performance in stress detection.

Overall we have presented a successful stress classification scheme in a lab environment. Our method is unobtrusive and fast. The promise of this method using dynamic stimuli along with other classification techniques is a new investigation area brought along with our findings.

REFERENCES

Andreassi, J. L. (2006). Pupillary response and behavior. *Psychophysiology: Human Behavior & Physiological Response, 5 ed.*, 350–371.

Baltaci, S., & Gokcay, D. (2012). Negative sentiment in scenarios elicit pupil dilation response. In *Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI '12* (p. 529). New York, New York, USA: ACM Press. https://doi.org/10.1145/2388676.2388787

Baltaci, S., & Gokcay, D. (2016). Stress Detection in Human–Computer Interaction: Fusion of Pupil Dilation and Facial Temperature Features. *International Journal of Human–Computer Interaction*, 1–11. https://doi.org/10.1080/10447318.2016.1220069

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In *Handbook of Psychopysiology* (pp. 276–92).

Berkovitz, B. (Producer), Kirsch, C. (Producer), Moxham, J. B. (Producer), Alusi, G. (Producer), & Cheesman, T. (Producer). (2013). 3D Head and Neck Anatomy with Special Senses and Basic Neuroanatomy. United Kingdom: Primal Pictures.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7962581

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Calvo, R. A., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective*

Computing, *1*(1), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory 2nd Edition* (2nd ed.). Wiley-Interscience. Retrieved from http://www3.interscience.wiley.com/cgi-bin/bookhome/110438582?CRETRY=1&SRETRY=0

Cross, C. B. (2013). An Investigation of Thermal Imaging to Detect Physiological Indicators of Stress in Humans.

Cross, C. B., Skipper, J. A., & Petkie, D. (2013). Thermal imaging to detect physiological indicators of stress in humans. In G. R. Stockton & F. P. Colbert (Eds.) (p. 87050I). International Society for Optics and Photonics. https://doi.org/10.1117/12.2018107

Czaja, S. J., & Sharit, J. (1993). Stress reactions to computer-interactive tasks as a function of task structure and individual differences. *International Journal of Human-Computer Interaction*, *5*(1), 1–22. https://doi.org/10.1080/10447319309526053

de Araujo, D. B., Tedeschi, W., Santos, A. C., Elias, J., Neves, U. P. C., & Baffa, O. (2003). Shannon entropy applied to the analysis of event-related fMRI time series. *NeuroImage*, *20*(1), 311–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14527591

De Santos Sierra, A., Sánchez Ávila, C., Guerra Casanova, J., & Bailador Del Pozo, G. (2011). A stress-detection system based on physiological signals and fuzzy logic. *IEEE Transactions on Industrial Electronics*, *58*(10), 4857–4865. https://doi.org/10.1109/TIE.2010.2103538

Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice* ((2 ed.).). Springer.

Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. *Refrigeration And Air Conditioning*. https://doi.org/10.1111/1467-9639.00050

Ekman, P. (1985). *Telling lies: clues to deceit in the marketplace, politics, and marriage*. Norton. Retrieved from https://books.google.ru/books?id=RfAMAQAAMAAJ

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252. https://doi.org/10.1007/BF00987191

Erdem, A., & Karaismailoglu, S. (2010). Neurophysiology of Emotions. In *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives: Psychological, Cognitive and Neuroscientific Perspectives* (p. 1). IGI Global.

Fujigaki, Y., & Mori, K. (1997). Longitudinal Study of Work Stress Among Information System Professionals. *International Journal of Human-Computer Interaction*, *9*(4), 369–381. https://doi.org/10.1207/s15327590ijhc0904_3

Fujimas, I. (1998). Pathophysiological expression and analysis of far infrared thermal images. *IEEE Engineering in Medicine and Biology Magazine*, *17*(4), 34–42. https://doi.org/10.1109/51.687961

Gane, L., Power, S., Kushki, A., & Chau, T. (2011). Thermal Imaging of the Periorbital Regions during the Presentation of an Auditory Startle Stimulus. *PLoS ONE*, *6*(11), e27268. https://doi.org/10.1371/journal.pone.0027268

Giakoumis, D., Tzovaras, D., & Hassapis, G. (2013). Subject-dependent biosignal features for increased accuracy in psychological stress detection. *International Journal of Human-Computer Studies*, *71*(4), 425–439. https://doi.org/10.1016/j.ijhcs.2012.10.016

Gökçay, D. (2011). Emotional axes: Psychology, psychophysiology and neuroanatomical correlates. In *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives* (pp. 56–73). IGI Global.

Gokcay, D., Baltaci, S., Karahan, C., & Turkay, K. D. (2011). Prediction of affective states through non-invasive thermal camera and EMG recordings. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6975 LNCS, pp. 319–321).

Gokcay, D., & Yildirim, G. (2011). Problems Associated with Computer-Mediated Communication Cognitive Psychology and Neuroscience Perspectives. *Affective Computing and Interaction*, 244–261. https://doi.org/10.4018/978-1-61692-892-6.ch011

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, *11*(1), 1–21. https://doi.org/10.1080/00401706.1969.10490657

Izard, C. E. (2009). Emotion theory and research: highlights, unanswered questions, and emerging issues. *Annual Review of Psychology*, *60*, 1–25. https://doi.org/10.1146/annurev.psych.60.110707.163539

Jaimes, A., & Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*, *108*(1), 116–134. https://doi.org/10.1016/j.cviu.2006.10.019

Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011). Physiological signals based human emotion Recognition: a review. In *2011 IEEE 7th International Colloquium on Signal Processing and its Applications* (pp. 410–415). IEEE.

https://doi.org/10.1109/CSPA.2011.5759912

Jing Zhai, J., Barreto, A. B., Craig Chin, C., & Chao Li, C. (2005). Realization of Stress Detection using Psychophysiological Signals for Improvement of Human-Computer Interaction. In *Proceedings. IEEE SoutheastCon, 2005.* (pp. 415–420). IEEE. https://doi.org/10.1109/SECON.2005.1423280

Jonghwa Kim, & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(12), 2067–2083. https://doi.org/10.1109/TPAMI.2008.26

Kensinger, E. A. (2004). Remembering Emotional Experiences: The Contribution of Valence and Arousal. *Reviews in the Neurosciences*, *15*(4), 241–252. https://doi.org/10.1515/REVNEURO.2004.15.4.241

Khan, M., Ward, R., & Ingleby, M. (2006). Infrared Thermal Sensing of Positive and Negative Affective States. In *2006 IEEE Conference on Robotics, Automation and Mechatronics* (pp. 1–6). IEEE. https://doi.org/10.1109/RAMECH.2006.252608

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, *42*(3), 419–427. https://doi.org/10.1007/BF02344719

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. Morgan Kaufmann Publishers Inc.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. &. (2008). International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual (Rep. No. A-8).

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8497555

Meinshausen, N. (2006). Quantile Regression Forests. *The Journal of Machine Learning Research*, *7*, 983–999.

Morency, L.-P., Mihalcea, R., & Doshi, P. (2011). Towards multimodal sentiment analysis. In *Proceedings of the 13th international conference on multimodal interfaces - ICMI '11* (p. 169). New York, New York, USA: ACM Press. https://doi.org/10.1145/2070481.2070509

Nhan, B. R., & Chau, T. (2010). Classifying Affective States Using Thermal Infrared Imaging of the Human Face. *IEEE Transactions on Biomedical Engineering*,

57(4), 979–987. https://doi.org/10.1109/TBME.2009.2035926

Partala, T., Jokiniemi, M., & Surakka, V. (2000). Pupillary responses to emotionally provocative stimuli. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00* (pp. 123–129). New York, New York, USA: ACM Press. https://doi.org/10.1145/355017.355042

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1), 185–198. https://doi.org/10.1016/S1071-5819(03)00017-X

Pavlidis, I. (2003). Continuous physiological monitoring. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)* (Vol. 2, pp. 1084–1087). IEEE. https://doi.org/10.1109/IEMBS.2003.1279434

Pavlidis, I., Dowdall, J., Sun, N., Puri, C., Fei, J., & Garbey, M. (2007). Interacting with human physiology. *Computer Vision and Image Understanding*, *108*(1), 150–170. https://doi.org/10.1016/j.cviu.2006.11.018

Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J.-R., Mérienne, F., Benedetto, S., & Baccino, T. (2014). Automatic Stress Classification With Pupil Diameter Analysis. *International Journal of Human-Computer Interaction*, *30*(3), 220–236. https://doi.org/10.1080/10447318.2013.848320

Qian, M., Aguilar, M., Zachery, K. N., Privitera, C., Klein, S., Carney, T., & Nolte, L. W. (2009). Decision-level fusion of EEG and pupil features for single-trial visual detection analysis. *IEEE Transactions on Bio-Medical Engineering*, *56*(7), 1929–37. https://doi.org/10.1109/TBME.2009.2016670

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, *4*(1), 77–90.

R.R. Seeley, T. D. S. and P. T. (2008). *Anatomy & Physiology*. New York, New York, USA: McGraw-Hill.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2011). *Psychology of Reading* (second edi). Psychology Press.

Ren, P., Barreto, A., Gao, Y., & Adjouadi, M. (2013). Affective Assessment by Digital Processing of the Pupil Diameter. *IEEE Transactions on Affective Computing*, *4*(1), 2–14. https://doi.org/10.1109/T-AFFC.2012.25

Ring, E. F. J. (1998). Progress in the measurement of human body temperature. *IEEE Engineering in Medicine and Biology Magazine*, *17*(4), 19–24. https://doi.org/10.1109/51.687959

Sharawi, M. S., Shibli, M., & Sharawi, M. I. (2008). Design and implementation of a human stress detection system: A biomechanics approach. In *Proceeding of the 5th International Symposium on Mechatronics and its Applications, ISMA 2008*. https://doi.org/10.1109/ISMA.2008.4648856

Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine*, *108*(3), 1287–1301. https://doi.org/10.1016/j.cmpb.2012.07.003

Sharma, N., & Gedeon, T. (2014). Modeling observer stress for typical real environments. *Expert Systems with Applications*, *41*(5), 2231–2238. https://doi.org/10.1016/j.eswa.2013.09.021

Shastri, D., Merla, A., Tsiamyrtzis, P., & Pavlidis, I. (2009). Imaging facial signs of neurophysiological responses. *IEEE Transactions on Bio-Medical Engineering*, *56*(2), 477–84. https://doi.org/10.1109/TBME.2008.2003265

Sun, F.-T., Kuo, C., Cheng, H.-T., Buthpitiya, S., Collins, P., & Griss, M. (2012). Activity-Aware Mental Stress Detection Using Physiological Sensors (pp. 211–230). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29336-8_12

Surakka, V., & Hietanen, J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *29*(1), 23–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9641245

Tobii Technology. (2010). Tobii Eye Tracking - An introduction to eye tracking and Tobii Eye Trackers. *Technology*, 12. Retrieved from http://www.tobii.com/Global/Analysis/Training/WhitePapers/Tobii_EyeTracking_Introduction_WhitePaper.pdf?epslanguage=en

Wang, J. (2011). Pupil dilation and eye tracking. *A Handbook of Process Tracing Methods for Decision …*, 1–33. Retrieved from http://books.google.com/books?hl=en&lr=&id=DBx5AgAAQBAJ&oi=fnd&pg=PA185&dq=Pupil+Dilation+and+Eye-tracking&ots=0tvxuLdqy8&sig=q65VvtsIwXbtOO-_CeZ1y4Si3XM

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3397865

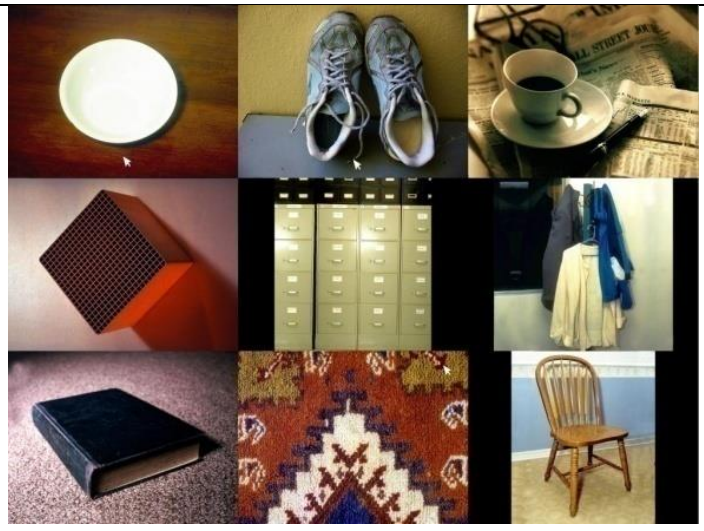
Witten, I. . H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. *Machine Learning*. Retrieved from http://www.amazon.com/Data-Mining-Practical-Techniques-Management/dp/0120884070/ref=sr_1_2?s=books&ie=UTF8&qid=1329477320&sr=1-2

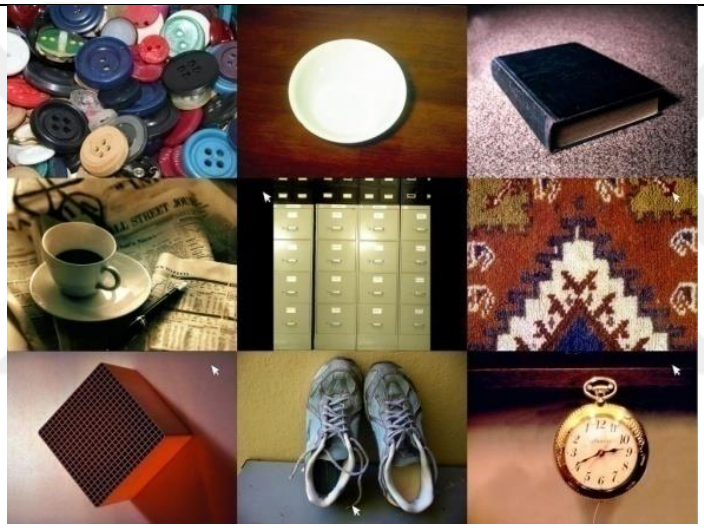
Yuen, P., Tsitiridis, a., Kam, F., Jackman, J., James, D., Richardson, M., … Lightman, S. (2009). Emotional & physical stress detection and classification using thermal imaging technique. *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*, *7486*, P13–P13. https://doi.org/10.1049/ic.2009.0241

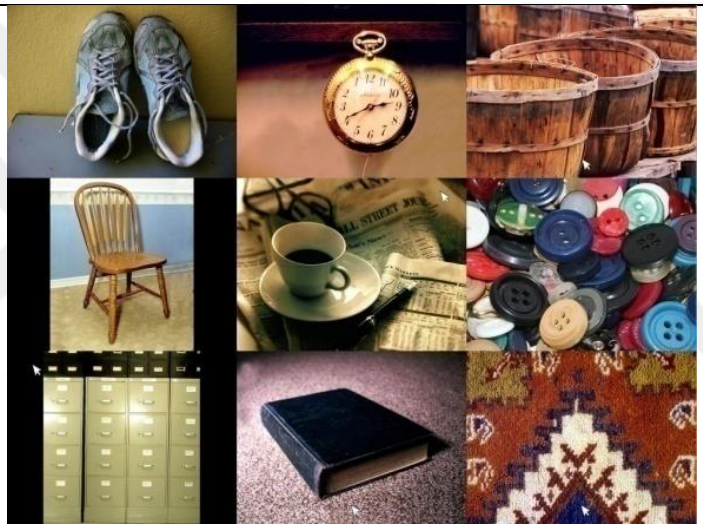
Yun, C., Shastri, D., Pavlidis, I., & Deng, Z. (2009). O' game, can you feel my frustration? In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (p. 2195). New York, New York, USA: ACM Press. https://doi.org/10.1145/1518701.1519036

Zhai, J., & Barreto, A. (2006). Stress detection in computer users based on digital signal processing of noninvasive physiological variables. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *1*, 1355–8. https://doi.org/10.1109/IEMBS.2006.259421

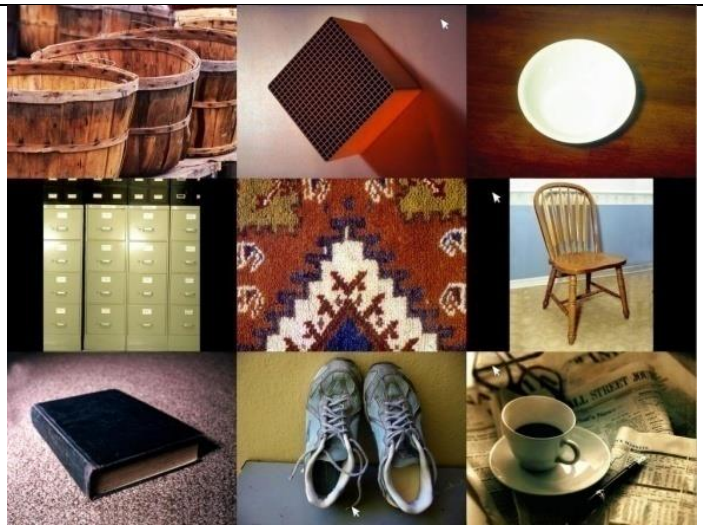
APPENDIX A : STIMULI FROM PART I

# of Stimulus	Grid of IAPS Images	IAPS #
1		7090 7235 7001 7179 7006 7190 7032 7057 7185
2		7006 7032 7057 7185 7224 7217 7090 7179 7235

3		7179 7217 7057 7001 7235 7190 7090 7006 7224
4		7001 7006 7090 7057 7224 7179 7185 7032 7190
5		7179 7090 7057 7185 7224 7041 7235 7217 7190

6		7224 7090 7032 7217 7041 7057 7179 7190 7001
7		7032 7190 7041 7235 7057 7001 7224 7090 7179
8		7224 7001 7217 7179 7090 7190 7006 7057 7235

9		7041 7090 7235 7190 7032 7185 7217 7179 7006
10		7041 7090 7185 7001 7032 7190 7006 7217 7179
11		7006 7217 7179 7032 7057 7190 7235 7224 7041

12		7041 7185 7006 7224 7179 7235 7090 7032 7057
13		7057 7041 7090 7001 7235 7224 7190 7217 7185
14		7235 7057 7006 7185 7217 7224 7090 7032 7001

15		7041 7001 7006 7032 7185 7217 7190 7057 7179
16		7032 7190 7217 7006 7235 7057 7001 7179 7090
17		7041 7006 7224 7179 7032 7090 7057 7217 7185

18		7057 7006 7217 7001 7032 7190 7224 7090 7041
19		7190 7185 7224 7235 7006 7001 7057 7090 7217
20		7057 7001 7185 7224 7090 7190 7006 7179 7041

APPENDIX B : STIMULI FROM PART II

# of Stimulus	Grid of IAPS Images	IAPS #
1		8230 2900 3030 3103 9332 9423 6550 3213 6940
2		9423 6940 9925 2900 9584 1120 3103 1300 3213

3		9332 1300 1120 6940 9925 9584 8230 2900 3213
4		3213 9925 3030 6550 8230 9332 9423 6940 3103
5		6550 9584 3103 3030 9925 9332 1300 2900 8230

6		3213 6940 3103 8230 1120 1300 6550 3030 9584
7		3213 6550 3103 9423 1300 9925 1120 2900 8230
8		9584 9423 3030 9332 8230 9925 6550 3213 6940

9		6940 9332 9584 3213 6550 3030 9925 8230 1300
10		3213 8230 1120 6550 9332 9584 9423 1300 3030
11		6550 9332 8230 9584 1120 3030 6940 2900 9925

12		3213 8230 6940 6550 3030 1300 9332 1120 9925
13		6940 9332 3030 3103 3213 8230 9423 2900 9925
14		1120 3030 3103 1300 9925 3213 6550 9584 2900

15		9925 3030 3213 2900 9423 9332 3103 6940 1120
16		3213 1120 9584 9925 1300 9423 3030 6550 2900
17		1120 3030 9925 9584 6940 3103 2900 8230 9332

18		3103 3213 1300 9423 6940 6550 9925 9584 3030
19		6550 6940 9423 3030 2900 9925 1300 9584 1120
20		9332 8230 3213 9925 9423 2900 1300 3030 6940

APPENDIX C : PANAS TEST

Bu ölçek farklı duyguları tanımlayan bir takım sözcükler içermektedir. Son iki hafta nasıl hissettiğinizi düşünüp her maddeyi okuyun. Uygun cevabı her maddenin yanında ayrılan yere (puanları X ekleyerek) işaretleyin.

		Çok az (hiç)	**Biraz**	**Ortalama**	**Oldukça**	**Çok fazla**
1	İlgili					
2	Sıkıntılı					
3	Heyecanlı					
4	Mutsuz					
5	Güçlü					
6	Suçlu					
7	Ürkmüş					
8	Düşmanca					
9	Hevesli					
10	Gururlu					
11	Asabi					
12	Uyanık					
13	Utanmış					
14	İlhamlı					
15	Sinirli					
16	Kararlı					
17	Dikkatli					
18	Tedirgin					
19	Aktif					
20	Korkmuş					

APPENDIX D: DEMOGRAPHIC INFORMATION FORM

Kişisel Bilgiler:
Adı Soyadı: Uygulama Tarihi:
… / … / …

Cinsiyeti: Kadın ( ) Erkek ( )

Doğum Tarihi: … / … / …

Yaşı: …

Medeni Hali: Evli ( ) Bekar ( ) Dul ( ) Boşanmış ( )

Mesleği: ……

El Tercihi: Sağ ( ) Sol ( )

Eğitim Durumu: İlkokul (0-5 yıl) ( )
Ortaokul (6-8 yıl) ( )
Lise (9-12 yıl) ( )
Üniversite (12+) ( )

Sağlık Durumuna İlişkin Bilgiler:
İşitme Bozukluğu: Var ( ) Yok ( )

Varsa düzeltilmiş mi? ……

Görme Bozukluğu var mı? Var ( ) Yok ( )

Varsa hangisi? Miyop ( ) Astigmat ( ) Hipermetrop ( )

Varsa düzeltilmiş mi? ……

Renk Körlüğü: Var ( ) Yok ( )

Fiziksel Özür: Var ( ) Yok ( )

Varsa türü: ……

Geçirdiği Önemli Rahatsızlıklar (Psikiyatrik, Nörolojik veya Psikolojik):
Halen Kullanmakta Olduğu İlaç: Var ( ) Yok ( )

Varsa ilacın/ilaçların adı: .....

Uzun Süre Kullanıp Bıraktığı İlaç: Var ( ) Yok ( )

Varsa ilacın/ilaçların adı: ……

Varsa kullanım süresi: ……

APPENDIX E: DEBRIEFING FORM

1. Deneyimizi nasıl buldunuz?

2. Kendi performansınızı nasıl degerlendiriyorsunuz?

3. Okları tutarlı olarak sayabildiğinizi düşünüyor musunuz?

4. Birinci kısım ile ikinci kısım arasında ne gibi bir fark hissettiniz?

5. Birinci kısım ile ikinci kısım arasında heyacanlanmanızda degisiklik oldu mu?

6. Resimlerin icerikleri hakkinda ne düşünüyorsunuz?

7. Eklemek istediğiniz başka şeyler var mı?

APPENDIX F: INFORMED CONSENT FORM

ARAŞTIRMAYA GÖNÜLLÜ KATILIM FORMU

Bu araştırma, Enformatik Enstitüsü Sağlık Bilişimi Bölümü öğretim üyelerinden Y. Doç. Dr. Didem Gökçay tarafından yürütülmekte olup Sağlık Bilişimi Doktora Programı öğrencisi Serdar Baltacı'nın doktora tez çalışmasının bir gereği olarak yapılmaktadır. Bu form sizi araştırma koşulları hakkında bilgilendirmek içindir.

Çalışmanın Amacı

Günlük hayatımızda olumluluk açısından farklı içeriklere sahip birçok görsel uyaranla karşılaşmaktayız. Bu uyaranların aynı zamanda heyecan verici olma, stres yaratma gibi özellikleri de bulunabilir. Bu araştırmanın amacı, tüm bu maruz kaldığımız uyaranların duygusal ve bilişsel süreçlerde davranışlarımız üzerindeki etkilerini incelemek ve vücut fizyolojimizde ne gibi değişimler yarattığını gözlemlemektir.

Bize Nasıl Yardımcı Olmanızı İsteyeceğiz?

Çalışma sırasında sizden 2 deneye katılmanız ve her birinde 20'şer resmi değerlendirmeniz istenmektedir. Değerlendirme sırasında görsel uyaranın üzerindeki okları saymanız ve '+' şeklindeki odaklanma işaretini gördüğünüz sırada ok sayısını sesli bir şekilde belirtmeniz beklenmektedir. Daha sonra size söylediğiniz ok sayısının doğru ya da yanlış olduğu belirtilecektir. Çalışmaya katılmayı kabul ettiğiniz takdirde, doldurmanız gereken anketler hakkında bilgilendirileceksiniz. Çalışma süresi toplam yarım saat olarak planlanmıştır.

Sizden Topladığımız Bilgileri Nasıl Kullanacağız?

Araştırmaya katılımınız tamamen gönüllülük temelinde olmalıdır. Çalışmada sizden kimlik veya çalıştığınız kurum/bölüm/birim belirleyici hiçbir bilgi istenmemektedir. Cevaplarınız ve sizden alınan veriler tamamıyla gizli tutulacak, sadece araştırmacılar tarafından değerlendirilecektir. Katılımcılardan elde edilecek bilgiler toplu halde değerlendirilecek ve bilimsel yayımlarda kullanılacaktır. Sağladığınız veriler gönüllü katılım formlarında toplanan kimlik bilgileri ile eşleştirilmeyecektir.

Katılımınızla ilgili bilmeniz gerekenler:

Bu çalışmada gözbebeği büyümesini ve hareketlerini takip edip kayıt altına almak için bir göz izleme cihazı kullanılmaktadır. Bilgisayar ekranının arkasında bir kamera fark edebilirsiniz, o kamera ile de yüz bölgenizin sıcaklığı saptanmaktadır. Bu cihazlar insan sağlığı ya da ruhsal durumu açısından en ufak bir risk teşkil etmemektedir.

Araştırmayla ilgili daha fazla bilgi almak isterseniz:

Bu çalışmaya katıldığınız için şimdiden teşekkür ederiz. Çalışma hakkında daha fazla bilgi almak için ODTÜ öğretim üyelerinden Y. Doç. Dr. Didem Gökçay (ODTÜ Enformatik Enstitüsü, A-216, (0 312) 210 3750, E-posta: dgokcay@metu.edu.tr) ile iletişim kurabilirsiniz.

Yukarıdaki bilgileri okudum ve bu çalışmaya tamamen gönüllü olarak katılıyorum.

(Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

İsim Soyad Tarih İmza

---/----/-----

APPENDIX G: APPROVED CONSENT FORM

(By METU Ethics Committee)

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

19 EKİM 2016

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
uygulamalietik@metu.edu.tr
www.uygam.metu.edu.tr

Sayı: 28620816 /413

Konu: Değerlendirme Sonucu

Gönderilen: Yrd.Doç.Dr. Didem GÖKÇAY
Sağlık Bilişimi.

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

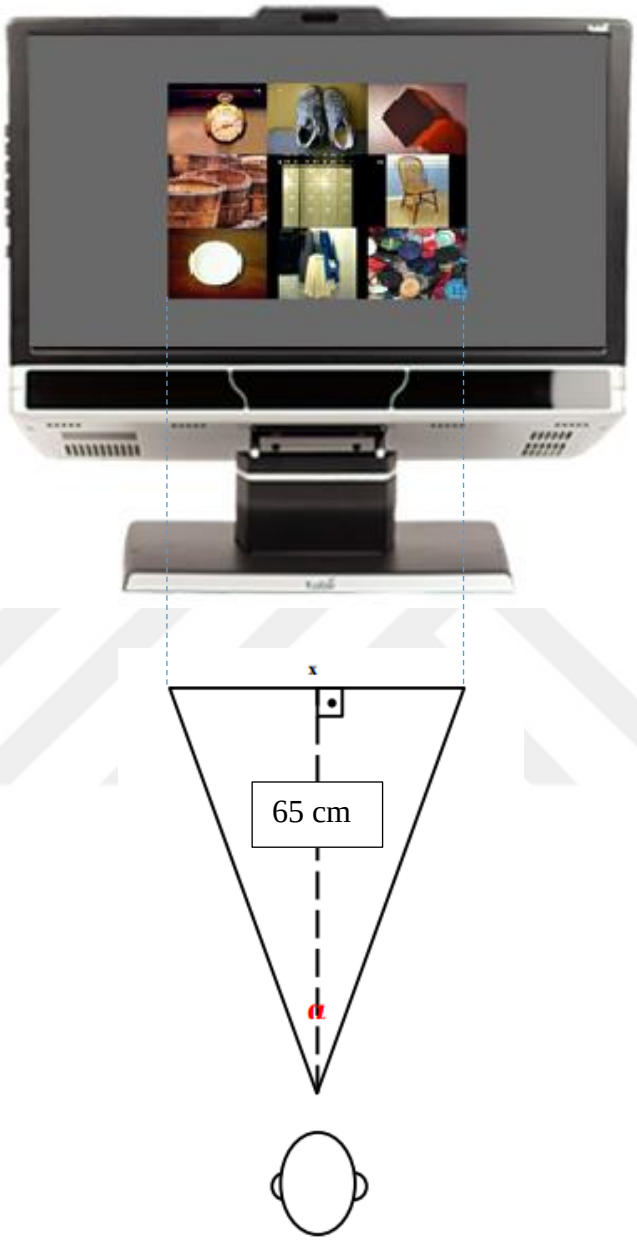
İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın : Yrd.Doç.Dr. Didem GÖKÇAY;

Danışmanlığını yaptığınız Serdar BALTACI'nın "İnsan-bilgisayar etkileşiminde stresin ve duygusal değişimlerin fizyolojik ölçümlerle incelenmesi" başlıklı araştırması İnsan Araştırmaları Kurulu tarafından uygun görülerek gerekli onay **2016-FEN-052** protokol numarası ve **19.10.2016-02.10.2017** tarihleri arasında geçerli olmak üzere verilmiştir

Bilgilerinize saygılarımızla sunarız.

APPENDIX H: VISUAL ANGLE COMPUTATION



tan(15°)= X/65

X=17.11 cm

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name : Baltaci, Serdar
Nationality : Turkish
Date and Place of Birth : 1 January 1980, İzmir
Marital Status : Married
Driving License : B
Phone : +90 533 493 99 81
Email : serdarbaltaci@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
M.Sc.	Ege University International Computer Institute, Information Technology	2004
B.Sc.	Ege University, Computer Engineering	2001
High School	İzmir Atatürk Anatoly High School	1997

WORK EXPERIENCE

Year	Place	Enrollment
2007- 2016	Prime Ministry	Software Architect and Developer
2005-2006	Turkish Armed Forces	Software Project Officer
2002-2005	Havelsan A.S.	Senior Software Engineer
2001-2002	Ege University	Researcher

PUBLICATIONS & PRESENTATIONS

• Baltaci, S., Gokcay, D, Model-free analysis of fMRI images based on Shannon entropy in a sliding window. Poster Presentation, (2010), Human Brain Mapping, Barcelona.

• Gokcay, D., Baltaci, S., Karahan, C., & Turkay, K. D. (2011). Prediction of affective states through non-invasive thermal camera and EMG recordings. In Lecture Notes in Computer Science (including subseries Lecture Notes in AI and Lecture Notes in Bioinformatics) (Vol. 6975 LNCS, pp. 319–321).

• Baltaci, S., & Gokcay, D. (2012). Negative sentiment in scenarios elicit pupil dilation response. In Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI ’12 (p. 529). New York, New York, USA: ACM Press. https://doi.org/10.1145/2388676.2388787

• Baltaci, S., & Gokcay, D. (2014). Role of pupil dilation and facial temperature features in stress detection. In 2014 22nd Signal Processing and Communications Applications Conference (SIU) (pp. 1259–1262). IEEE. https://doi.org/10.1109/SIU.2014.6830465

• Baltaci S, Gokcay, D. (2015). Evaluation of features derived from pupil dilation in a stress induction experiment. European Conference on Visual Perception,Abstracts.

• Baltaci, S., & Gokcay, D. (2016). Stress Detection in Human–Computer Interaction: Fusion of Pupil Dilation and Facial Temperature Features. International Journal of Human–Computer Interaction (Vol. 32, pp. 1–11) https://doi.org/10.1080/10447318.2016.1220069

FOREIGN LANGUAGES

Native Turkish, Advanced English

VITA

Serdar Baltaci was born in İzmir on January 1, 1980. He received his B.S. degree in Computer Engineering from Ege University in August 2001. He worked in government institutions and companies as a software developer position since then. He continued his education at the same university, in Information Technology field. His main areas of interest in academics are pattern recognition, affective computing and image processing.