UNIVERSITY OF TARTU
Institute of Computer Science
Computer Science Curriculum

Artem Bachynskyi

Emotional State Recognition Based on Physiological Signals

Master's Thesis (30 ECTS)

Supervisor: Ilya Kuzovkin, MSc
Supervisor: Raul Vicente Zafra, PhD

Tartu 2018

Emotional State Recognition Based on Physiological Signals

Keywords:
Emotion Recognition, Emotion Classification, Physiological Signals, EEG, ECG, Machine Learning

CERCS: P170

Abstract:

Emotional state recognition is a crucial task for achieving a new level of Human-Computer Interaction (HCI). Machine Learning applications penetrate more and more spheres of everyday life. Recent studies are showing promising results in analyzing physiological signals (EEG, ECG, GSR) using Machine Learning for accessing emotional state. Commonly, specific emotion is invoked by playing affective videos or sounds. However, there is no canonical way for emotional state interpretation. In this study, we classified affective physiological signals with labels obtained from two emotional state estimation approaches using machine learning algorithms and heuristic formulas. Comparison of the method has shown that the highest accuracy was achieved using Random Forest classifier on spectral features from the EEG records, a combination of features for the peripheral physiological signal also shown relatively high classification performance. However, heuristic formulas and novel approach for ECG signal classification using recurrent neural network ultimately failed. Data was taken from the MAHNOB-HCI dataset which is a multimodal database collected on 27 subjects by showing 20 emotional movie fragment's. We obtained an unexpected result, that description of emotional states using discrete Eckman's paradigm provides better classification results comparing to the contemporary dimensional model which represents emotions by matching them onto the Cartesian plane with valence and arousal axis. Our study shows the importance of label selection in emotion recognition task. Moreover, obtained dataset have to be suitable for Machine Learning algorithms. Acquired results may help to select proper physiological signals and emotional labels for further dataset creation and post-processing.

Emotsionaalse Seisundi Tuvastamine Füsioloogiliste Signaalide Baasil

Võtmesõnad:
Emotsioonide Tunnustamine, Emotsioonide Klassifikatsioon, Füsioloogilised Signaalid, EEG, EKG, Masin Õppimine

CERCS: P170

Lühikokkuvõte:

Emotsionaalsete seisundite tuvastamine on väga tähtis inimese ja arvuti vahelise suhtlemise (Human-Computer Interaction, HCI) jaoks. Tänapäeval leiavad masinõppe meetodid ühe enam rakendust paljudes inimtegevuse valdkondades. Viimased uuringud näitavad, et füsioloogiliste signaalide analüüs masinõppe meetoditega võiks võimaldada inimese emotsionaalse seisundi tuvastamist hea täpsusega. Vaadates emotsionaalse sisuga videosid, või kuulates helisid, tekib inimesel spetsifiline füsiloogiline vastus. Antud uuringus me kasutame masinõpet ja heuristilist lähenemist, et tuvastada emotsionaalseid seisundeid füsioloogiliste signaalide põhjal. Meetodite võrdlus näitas, et kõrgeim täpsus oli saavutatud juhumetsa (Random Forest) meetodiga rakendades seda EEG signaalile, mis oli teisendatud sagedusintervallideks. Ka kombineerides EEG-d teiste füsioloogiliste signaalidega oli tuvastamise täpsus suhteliselt kõrge. Samas heuristilised meetodid ja EEG signaali klassifitseerimise rekurrentse närvivõrkude abil ebaõnnestusid. Andmeallikaks oli MAHNOB-HCI mitmemodaalne andmestik, mis koosneb 27 isikult kogutud füsioloogilistest signaalidest, kus igaüks neist vaatas 20 emotsionaalset videolõiku. Ootamatu tulemusena saime teada, et klassikaline Eckman'i emotsionaalsete seisundite nimekiri oli parem emotsioonide kirjeldamiseks ja klassifitseerimiseks kui kaasaegne mudel, mis esitab emotsioone valentsuse ja ärrituse teljestikul. Meie töö näitab, et emotsiooni märgistamise meetod on väga tähtis hea klassifitseerimismudeli loomiseks, ning et kasutatav andmestik peab sobima masinõppe meetodite jaoks. Saadud tulemused võivad aidata valida õigeid füsioloogilisi signaale ja emotsioonide märkimise meetodeid uue andmestiku loomisel ja töötlemisel.

Contents

1 Introduction

Emotional state recognition is an estimation of humans organism state after a particular stimulus. According to physiological studies, different emotional states cause some changes in the sympathetic branch of the autonomous nervous system, which regulates the functioning of heart, respiration and body heat [New80]. Thus, all mentioned features can be measured using available medical techniques like electrocardiography, electromyography, electrodermal activity, and electroencephalography, because the emotional state also reflects on the brain activity.

Recognition of human emotions may provide more comfortable experience in the human-computer interaction (HCI). A multimedia industry may benefit from emotional recognition applications by adjusting their content according to consumers mood [SLPP12]. Affordable devices that can accurately measure heart rate, electrocardiogram or electroencephalogram are widely represented on the current market, so such HCI application becoming more and more realistic.

The goal of this thesis is to compare two emotional paradigms in terms of emotional state recognition based on physiological signals. In particular, in this work, we took labels from self-assessed emotional keywords and self-assessed valence / arousal and transformed both of them to low, medium and high categories in valence / arousal scales. Next, we classified affective emotional state records from the MAHNOB-HCI dataset using heuristic and machine learning methods.

In the background chapter, I described existing concepts of emotion categorization and physiological signals that recapitulate changes in emotional states. Also, emotional elicitation techniques and affective physiological databases were presented. In the methods chapter machine learning and data processing, approaches were explained. Result chapter combines presentation and interpretation of the reach performances in different experimental setups. In the conclusion chapter, we end up with surprising results, that more straightforward way for emotional state estimation using discrete emotional keywords in most cases provided better classification performance.

2 Background

In the first part of this section, we will give a general description of emotion phenomena and introduce discrete and dimensional (valence / arousal) emotional paradigms. Next, physiological signals that indicate emotional state changes will be presented. The third part of the chapter will provide information about existing techniques for emotion elicitation. Finally available physiological databases that were collected using mentioned approaches will be described in details.

2.1 Theories of Emotions

Throughout the ages, philosophers were thinking about the origins of emotions. Beginning with Aristotle, who established the four humors - components of the human body that are responsible for an emotional state. Further, Galen called them “sanguine”, “choleric”, “melancholic” and “phlegmatic” [Mer87]. With the arrival of the twentieth century, physiology as a part of medicine experienced rapid development. Based on research, the majority of scientists agree that three basic emotions (fear, anger, and happiness) are the most distinguishable and significantly differ from each other [WMBB13]. However, emotional expressions are much more compound, so more complex multidimensional models were developed to provide a comprehensive representation of human emotions [Rus80]. In this section, we describe two most popular models of human emotional state.

2.1.1 Discrete Emotion Model

The most well-known emotional paradigm belongs to one of the mos authoritative physiologists of 20th century Paul Ekman. During his research in Papua New Guinea, he remarked that some facial expressions of isolated tribe members also manifested in another culture [Ekm93], which means that some emotional expressions are common for all human beings. Based on his observations he proposed six basic emotions: anger, disgust, fear, happiness, sadness, and surprise [EF71]. Despite the fact, that more efficient paradigms was developed, described one still considered suitable for emotional state representation.

Figure 1. Six basic emotions by Ekman [EF71] (Image took from [Man]).

Further research by Izard [Iza77] [ATAS88] modified Ekman's model. Carrol Izard took into account physiological features and proposed next ten classes: anger, disgust, fear, sadness, contempt, shame, guilt, joy, surprise and interest.

The primary advantage of emotion categorization in a discrete manner is plainness for the majority of people that allows for their extensive use in surveys and questioning. Unfortunately, they are not suitable for human-computer interaction. Humans tend to name the same emotional state using different words. For example, words that describe emotions may vary depending on context, cultural background, and personality. That is why we need something more efficient to represent a human emotional state.

2.1.2 Valence / Arousal Emotion Model

Russel [Rus80] proposed the two-dimensional emotional model in 1980. The main idea is to divide emotions into two components: valence (or pleasure) and arousal (or activation). Negative values on a valence axis correspond to negative emotional state, and positive values correspond to a positive state, respectively. Arousal axis defines "power" of emotion or how it's active or passive depending it's positive or negative. Origin of mention two axes represents a neutral state.

Whissel proposed to divide the considered model into quadrants like in Cartesian plane and assign to them emotions from the discrete paradigm [Whi09] (see Figure 2). First quadrant (positive valence, positive arousal) specifies active and positive emotional states like excitement, surprise, happiness, etc. The second quadrant (negative valence, positive arousal) corresponds to high-activated negative emotion like anger, fear, distress. The third quadrant (negative valence, negative arousal) indicates negative calm states

such as disgust or sadness. The last, fourth quadrant (negative valence, positive arousal) describes positive states with low power; the best example will be relaxation, satisfaction, and calmness.

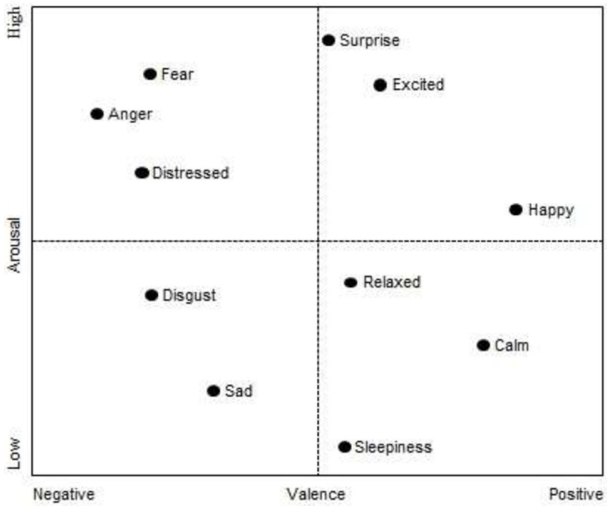


Figure 2. Whissel Wheel model [RMMI+14].

However, this model has a disadvantage, emotions that lie in the same quadrant might be indistinguishable in some cases. To tackle this problem, a technique called self-assessment manikin (SAM) was developed. It allows to accurately collect human response in an intuitive way that is readily interpretable on valence / arousal model. Initially, Semantic Differential Scale (SDS) proposed by Mehrabian [MR74] was used for measuring valence and arousal. SDSs data collection process was inconvenient and was not suitable for non-English speakers, because implied answering questions in English for evaluation of the experienced emotion [MR74]. SAM was developed by Lang in 1980 [Lan80]; it contains a set of pictures shown in Figure 3.

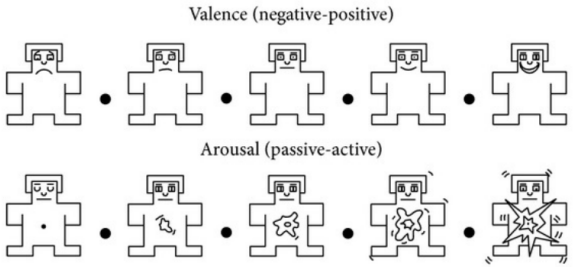


Figure 3. Self-assessment manikin [BF17].

On the SAM picture, high valence is encoded as a smiling face figure, and a gloomy character depicts low valence. For the arousal scale, a character with widely opened eyes

stands for positive value and figure with entirely closed eyes for negative. Thanks to its simplicity SAM can be applied to a range of emotion elicitation methods and allows to collect responses not only from non-English speakers but also from children and adults [LHCJL85], as well as various clinical populations [BL94].

2.2 Physiological Signals Associated with Emotion

When talking about traditional human-computer interaction, the emotional state of the user is usually not taken into account. However, usage of it might significantly increase the efficiency of intercommunication in such applications as e-learning, games and health care. The most native way of emotion manifestation by a human are facial expressions, voice tone, gestures, and posture, but all of the mentioned attributes can be controlled and faked. Thus, researchers of the affecting computing are focused on the physiological signals, because they are regulated by a sympathetic branch of the autonomous nervous system and cannot be elicited by any conscious or purposive control [New80]. The most used physiological signals for emotional state estimation are:

- Electrical activity of the brain (electroencephalography, magnetoencephalography [ASK+15])
- Cardiac function (electrocardiography and heart rate variability)
- Skin conductance (galvanic skin response)
- Muscle electrical activity (electromyography)
- Respiration
- Temperature

2.2.1 Electroencephalography

Electroencephalography (EEG) is a method for recording electrical brain activity. Electrical phenomena of the exposed cerebral hemispheres of mammals were found in 1975 by Richard Caton [Sch74]. In 1924 Hans Berger recorded the first human EEG. According to D. Millet, EEG was called "one of the most surprising, remarkable, and momentous developments in the history of clinical neurology" [Mil02]. Contemporary EEG is non-invasive in most cases — the electrodes are placed on the scalp surface. A common way to place them is called "10-20" system [Sil63]. This is an internationally recognized set of rules of placement and interpretation for EEG recordings. Each electrode placement area named by letter which encodes the lobe, or area of the brain (pre-frontal (Pf), frontal (F), temporal (T), parietal (P), occipital (O), and central (C)) and a number (to distinguish multiple electrodes on the same region (see Figure 4). "10-20" requires at least 19

electrodes, however, depending on task this number might be higher. It was called in such way because the distances between the contiguous electrodes can be 10% or 20% of the total left-right or front–back length of the brainpan.

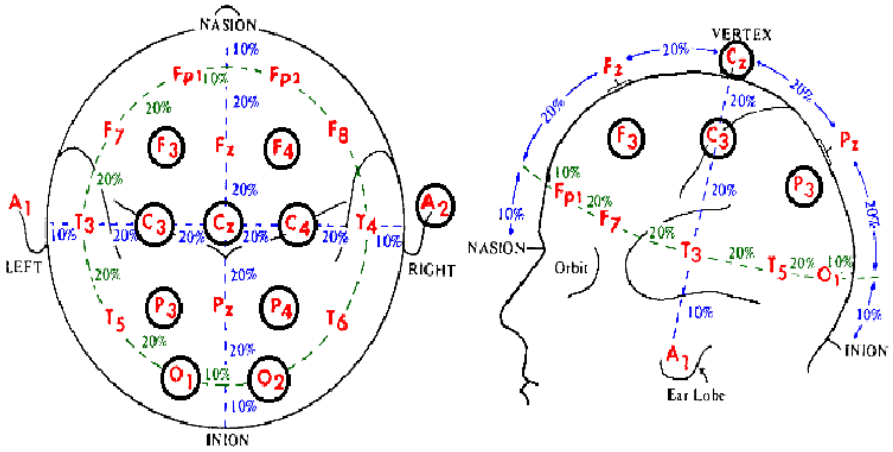


Figure 4. EEG electrodes placement according 10-20 system [AG15].

EEG is a way of monitoring the brain activity monitoring by measuring voltage changes appeared after activation of the nerve cells (neurons) [Tep02]. From the clinical perspective, EEG measures unprompted electrical activity during the time. The primary sources for clinical diagnosis are brain responses to a motor, sensory and cognitive events that are called Event-Related Potentials, and power spectrum analysis which transforms the signal from time to frequency domain. There are five standard frequency bands: delta (less than 4 Hz), theta (from 4 to 7 Hz), alpha (from 8 to 15 Hz), beta (from 16 to 31 Hz) and gamma (more than 32 Hz). Expression of each band is used for dysfunction detection. For example, epilepsy, coma, encephalopathies, brain death and sleep disorders could be recognized. Also, EEG is used for the initial diagnosis of brain tumors, strokes, and other brain disorders. Despite the popularity of EEG, such brain imaging techniques as magnetic resonance imaging (MRI) and computed tomography (CT) are more suitable for diagnostics task. EEG has poor spatial resolution compared to fMRI and CT. However, it outperforms them in temporal resolution. Thus, it is appropriate to use in the emotional state recognition tasks.

2.2.2 Electrocardiogram

Electrocardiography(ECG) is the approach of recording the electrical activity of the heart using electrodes attached to the skin. These electrodes detect the minimal electrical changes on the skin surface that appear from the heart muscle's contraction and relaxation during each heartbeat [NLdS05]. Conventional 12-lead ECG includes ten electrodes that

are placed on the patient's arms, legs and along the chest. They allow monitoring heart activity from different angles in vertical and horizontal electrical planes. Throughout each heartbeat, a healthy heart has an orderly progression of special events that represent mechanics of cardiac muscle activity. In contemporary cardiology, these events are called P, Q, R, S, T points and are shown in Figure 5.

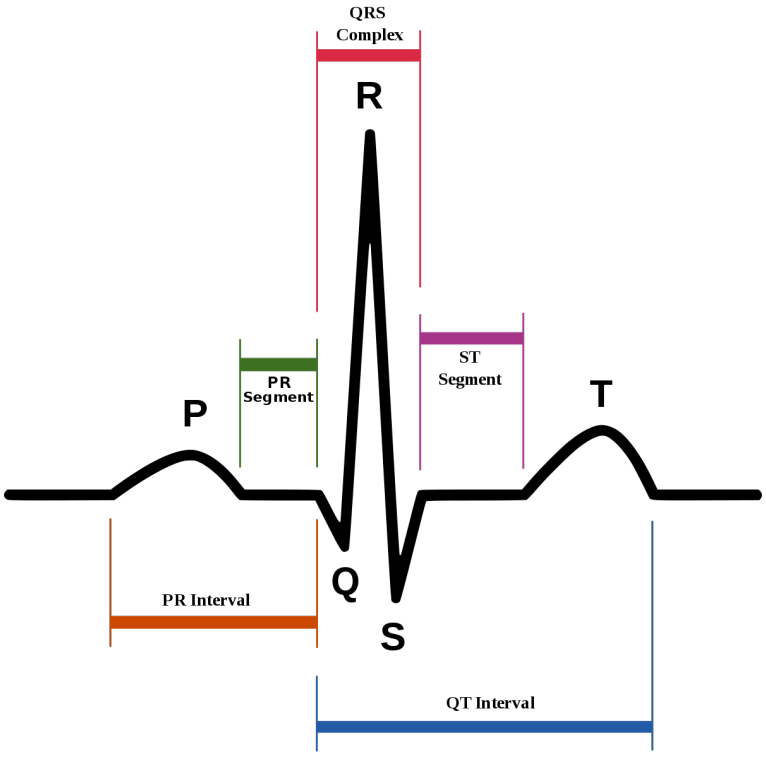


Figure 5. Visualization of PQRST events [img18].

For the experienced doctors, an ECG contains a significant amount of information about the state of the heart and cardiovascular system. Also, an ECG can be used to measure the heart rate variability, the size and position of the heart chambers, the presence of any damage to the heart's muscle cells or conduction system, the effects of cardiac drugs, and the function of implanted pacemakers. Data scientists attempted ECG problems for last two decades [Kon01], however, always faced with a lack of data. Luckily, portable ECG recorders widely arise on the market in recent years, which significantly increased the amount of data. Results were not long in coming, Stanford research group released a paper [RHH+17] where they claimed human (doctor) accuracy in various types of arrhythmia detection and reached such performance using data from 30,000 unique patients.

2.2.3 Heart Rate Variability

Heart rate variability (HRV) is a phenomenon of volatility in the time interval between the heartbeats. This physiological mark is measured by the length of consecutive inter-beat intervals (RR-intervals). Primary methods for tracking heartbeats are ECG, blood pressure, ballistocardiograms (monitoring of barely noticeable body fluctuations caused by heart activity), and the pulse wave signal extracted from a photoplethysmograph (PPG) [LZJ+08] (a technique based on an optical sensor for pulse measure, widely spread in modern wearable devices). However, ECG is considered being the most reliable, because it contains direct information about cardiac muscle contraction in its waveform. For HRV analysis performs in time and frequency domains. For time domain those features are the number of pairs of successive RR-intervals that differ by more than 50 ms, mean, standard deviation and range of RR-intervals. For frequency domain, the high-frequency band (HF) from 0.15 to 0.4 Hz, the low-frequency band (LF) from 0.04 to 0.15 Hz, and the very low-frequency band (VLF) from 0.0033 to 0.04 Hz are extracted using discrete Fourier transform [BAGC86] and allow to detect anomalies in the heart functions.

2.2.4 Galvanic Skin Response

Galvanic Skin Response (GSR) is a feature of human body continuously change its electrical characteristics. In different sources it is also called skin conductance, electrodermal response (EDR), psychogalvanic reflex (PGR), skin conductance response (SCR), sympathetic skin response (SSR) and skin conductance level (SCL) [wik18a]. Theoretically, GSR is based on hypothesis, that skin resistance depends on the state of sweat glands (tubular structures of the skin that produce sweat [wik18b]). Since sweating is out of control of the human, it is regulated by the autonomous nervous system (ANS). High arousal of the sympathetic branch of ANS increases the sweat gland activity, which accordingly increases skin conductance and another way around. This is the reason why GSR can be considered as an index of emotional and sympathetic responses. Complementary studies are saying about dependencies between electrodermal response and such emotional states like stress, drowsiness, and engagement [NWC13]. GSR signal is elementary to collect. Initially, two electrodes should be placed on the second and third finger as shown in Figure 6.

Then, low-voltage current is supplied to the placed electrodes, and the changes in conductivity are measured in units of micro-Siemens (μS). Traditionally, GSR data is obtained with the sampling rates from 1 to 10 Hz, but some new equipment can provide resolution up to 1 kHz.

2.2.5 Muscle Electrical Activity

The procedure for measurement of electrical muscle activity is called electromyography (EMG). Electric potential generated by the motor neurons is recorded by electrodes

Figure 6. Galvanic Skin Response recording procedure [Imo].

when the internal electrical source or brain command activates the cells. For the clinical purposes, the obtained signal is interpreted as a sound, graph or numerical values for further analysis by a specialist. The given technique can be used for the diagnostics of nerve and muscle dysfunction and problems with the nerve-to-muscle signal transmission.

2.2.6 Respiration

Respiration (breathing, ventilation) is physiological of phenomena of air transfer in and outside the lungs. During this process, blood saturates with oxygen from the gas obtained in the breath stage, in the exhalation stage, the body gets rid of carbon dioxide. Breathing, like other physiological factors, reflects on the organisms state changes, emotions are not an exception. Increased heart rate and more rapid braising in humans with high arousal state were shown [NTD97] [Boi98]. Also, animal and human studies show, that high-frequency respiration patterns were observed during fear and the anxiety state [Dav92]. Anxiety as a primary emotion is considered to be the most distinguishable based on the breathing pattern. In this case, increase in respiration frequency is not related to oxygen consumption, but is caused by parasympathetic nervous system [MH01] which is responsible for the emotional regulation. Respiration rate widely used in polysomnography (sleep studies) and polygraph tests, along with the other physiological signals mentioned in this chapter. There are several methods for measuring the breathing rate. Impedance pneumography involves measures of velocity and force of chest movements during respiration; capnography technique controls the concentration or pressure of carbon dioxide in the respiratory gases. They both are used in clinical practice. However, alternative wearable devices with the electrocardiogram, photoplethysmogram, and accelerometry sensors allow making the same monitoring with

relatively high accuracy.

2.2.7 Temperature

Temperature is quite advantageous and straightforward physiological signal. Value of the temperature may vary, depending on the place of the body, physical activity of the person and measuring equipment. Since changes in temperature correlate with emotional state adjustments, it is widely used in polygraphs and another type of emotional state monitoring.

2.3 Emotion Elicitation Techniques

Emotion elicitation is a crucial part of affective science. During the with years of studies in this field, researchers have concurred that no single sensory domain would be ultimate for emotional invocation. Thus, a wide range of different techniques was developed. Mostly, they include visual, auditory and even social or environmental effects. Some approaches combine various modalities, like audio + visual by showing affective videos. However, the hottest topic of recent years is Virtual Reality (VR) [OAHT17] application for the considered task. The virtual environment gives more mobility to the subject allows adjusting the stimulus to provide higher emotional arousal. In the next sections, all mentioned techniques will be described in more details, with examples of affective databases and their experimental setups.

2.3.1 Visual Stimuli

International Affective Picture System (IAPS) [LBC99] is the best-known database for affective state studies; it provides pictures to conduct experiments for different physiological research like emotions and attention. It was developed by the National Institute of Mental Health Center for Emotion and Attention at the University of Florida. This database was developed over several years, and in 2005 it contained 956 color photos varying from everyday items and situations like cats or furniture to very specific, with maimed humans or sexual scenes.

IAPS has one advantageous feature that distinguishes it from other datasets. All photos in it were evaluated in valence and arousal scales. During the collection of the database, subjects' assessments were collected in valence, arousal and dominance dimensions using the earlier described technique called SAM. An experimental setup was as follows: all images were divided into 16 batches 60 images in each, with 6 seconds for showing a picture and 15 seconds for rating them by the subject. Overall, 50 men and 50 women took part in this experiment. Also, the same tests were conducted on children but with extend time and more child-friendly explanations.

However, usage of IAPS in studies is too broad, making stimulus known to subjects prohibiting its use in subsequent trials. Thus, datasets analog to IAPS were collected. For instance, Geneva Affective Picture Database (GAPED) contained 730 validated pictures and aimed to increase the availability of visual and emotional stimuli [DGS11]. For elicitation of negative emotions, they used photos of spiders, snakes and violent scenes. The Nencki Affective Picture System (NAPS) incorporates even more images (1,356) with high resolution, which are categorized into five groups (landscapes, objects, faces, animals, and humans) [MŻJG14]. There is also an open-access dataset OASIS (Open Affective Standardized Image Set) [KLB17], it simulates IAPS [LBC99], but is available in public domain with no restrictions. All the databases mentioned above were validated in valence / arousal dimensions using the SAM technique.

2.3.2 Audio Stimuli

Audio stimuli are used much less often. Understandably, mining of audio data requires more effort comparing to images. However, dataset called International Affective Digitized Sounds (IADS) was created [SJ08], in comparison with previously collected databases that had an insufficient number of dimensions (separating only positive emotions from negative one), it was validated as well as IAPS in valence / arousal domain. Given database consist of 111 emotionally evocative sounds for the full range of emotions. Recent studies [NVG$^+$15] shown comparative results of elicitation for further emotional state detection based on physiological signals. The used HRV as a data for emotion recognition and obtained 84.72% on the valence dimension, and 84.26% on the arousal dimension. Other studies that solve the same problem but on the data collected via IAPS [LBC99] or similar dataset provided almost the same results, so audio emotion elicitation is suitable for physiological, affective data collection.

2.3.3 Videos Stimuli

Video stimuli are the most widely used way for emotion elicitation in laboratory studies [BDCC15]. Combination of audio and visual channels provide a more realistic emotional experience which is closer to real life. First trials in the creation of short video for emotion elicitation were made by Philippot [Phi93], also, Gross and Levenson [GL95]. However, selected clip segments were able to invoke only strong emotions. Therefore, during the next studies, researchers attempted to expand the range of emotions that possible to elicit by multimedia stimuli. In Table 1 we presented currently available databases.

Table 1. Annotated Video Databases for Emotion Elicitation [BDCC15].

Name	Size and Duration	Emotional labels
HUMAINE [DCCM+11]	50 clips from 5 seconds to 3 minutes long	Wide range of labels at a global level (emotion-related states, context labels, key events, emotion words, etc.) and frame-by-frame level (intensity, arousal, valence, dominance, predictability, etc.)
FilmStim [SNSP10]	70 film excerpts from 1 to 7 minutes long	24 classification criteria: subjective arousal, positive and negative affect, a positive and negative affect scores derived from the Differential Emotions Scale, six emotion discreteness scores and 15 mixed feelings scores
DEAP [KMS+12]	120 one-minute music videos	Ratings from an online self-assessment on arousal, valence and dominance and physiological recordings with face video for a subset of 40 music videos
MAHNOB-HCI [SLPP12]	20 film excerpts from 35 to 117 seconds long	Emotional keyword, arousal, valence, dominance and predictability combined with facial videos, EEG, audio, gaze and peripheral physiological recordings
EMDB [CLGÁG12]	52 non-auditory film clips of 40 seconds long	Global ratings for the induced arousal, valence, dominance dimensions
VIOLENT SCENES DATASET [DPGS12]	25 full-length movies	Annotations include the list of the movie segments containing physical violence according to two different definitions and also include 10 high-level concepts for the visual and audio modalities (presence of blood, fights, gunshots, screams, etc.)
LIRIS-ACCEDE [BDCC15]	9,800 excerpts from 8 to 12 seconds long	Rankings for arousal and valence dimensions

A dataset HUMAINE collected by Douglas-Cowie [DCCM+11] contains several subsets of natural and affected responses. Because of the broad range of labels, it's not applicable for machine learning task, though well represents basic principles of affective computing.

The FilmStim database [SNSP10] collected by Shaefer contains 70 video clips validated by physiologists for experimental use. There are ten videos for every emotion (neutral, tenderness, amusement, sadness, fear, disgust, anger). 24 items of assessment were collected from 364 subjects, which makes this dataset validated on the highest number of participants. However, desired the emotional state that expressed by physiological activity leads only up to 10 seconds [Kap11], so in given dataset, a real response is vanished in time because of video length from 1 to 7 minutes.

Koelstra built a dataset called DEAP [KMS+12], it consist of 120 music video fragments, each minute long and was validated in valence / arousal scale on 14 subjects. Furthermore, physiological signals like EEG, GSR, blood volume pressure, temperature, and respiration. Unfortunately, some of the used videos are under the copyright, another one, collected from YouTube are no longer available. It suggests that such databases have to be based on the open-source content.

Another affective database presented in the Table 1 called MAHNOB-HCI created by Soleymani [SLPP12]. This dataset is multimodal, besides 20 clips extracted from movies, it also contains records of various physiological signals and eye gaze data for 27 participants. As all previous databases, it validated in valence / arousal dimensions.

Emotional movie database (EMDB) [CLGÁG12] is another set of videos for emotions elicitation. The critical feature of this database that all 54 clips provided without sound. It was done to enhance the capacity of future experiments.

Demarty released one specific affective database in 2012 [DPGS12]. It called violent scene Dataset, consequently, containing movies with cruel fragments. Because of lopsided thematics of content in given database, only highly aroused negative emotions can be studied.

Massachusetts Institute of Technology (MIT) also have their affective database [JBC14]. Using GIF files (short looped videos without sound) from social networks they obtained 2.5M user annotations for 17 discrete emotions. Such broad control group makes this dataset very promising for further research.

In conclusion, the database called LIRIS-ACCEDE, which overcomes the limitations of all predecessors, was released in 2015 by Baveye [BDCC15]. 9,800 videos were collected under creative commons licenses, that means no issues with copyrights for public sharing. All videos in a considered database are from 8 to 12 seconds long, which makes it convenient for physiological signal measurements during emotion elicitation. Also, all records are accompanied by valence / arousal values collected via self-assessment.

2.3.4 Other Types of Stimuli

According to definition of emotion given by Klaus R. Scherer [Sch05]:"An episode of interrelated, synchronized changes in the states of all or most of the five organismic the evaluation of an external or internal stimulus event as relevant to major concerns of the organism", stimulus for emotion elicitation may be literally any event from the environment. Thus, Healy and Picard from MIT used the real-world driving task to draw drivers into the stressful state [HP05]. Each of 24 subjects drove at least 50 minutes, while physiological signals like ECG, GSR, and respiration were recorded. Further studies based on collected data [HP05] showed a correlation between drivers stress level and changes in skin conductivity and heart rate. This research demonstrates how different stimuli can be used for emotion elicitation.

Another alternative way for emotional invocation is using dyadic interaction described by Roberts, Tsai and Coan [RTC07]. The main idea of given technique is to engage interlocutors into such conversation, which maximizes the elicitation of emotion. For instance, contrary emotions were detected during discussions in a controversial field, whereas, positive emotions were obtained in conversations about enjoyable topics. A mentioned approach is not suitable for physiological data collection, because it's hard to estimate exact time of real emotional invocation.

Virtual and augmented reality (VR & AR) is new but very promising tools for affective computing. Recent study [OAHT17] shows that mentioned technologies can efficiently elicit emotionally and add flexibility to the experimental setup. Adjustability of the virtual environment allows generation of individual stimulus that would provide higher aroused emotional expression.

2.4 Datasets of Human Affective States

Currently, there is no benchmark for emotional state recognition algorithms. However, several databases presented in Table 2 are using in research. All considered datasets are multimodal, that means multiple streams of data (mostly physiological) were record during each session. DEAP is a well-known dataset [KMS+12] of EEG signals form emotional state recognition; it also has records of peripheral physiological signals like GSR and respiration. However, it's more frequently appears in papers dedicated to research emotional brain activity. Another dataset called DECAF [ASK+15] is a Magnetoencephalography-based Multimodal Database for Decoding Affective Physiological Responses. It had the wider range of signals like ECG, EMG and infra-red video as additional modality and was collected using the same short videos as in the DEAP. ASCERTAIN [SWA+17] and AMIGOS [CASP17] are relatively new and unexplored datasets with both visual and physiological (EEG, ECG, GSR) data streams. Regrettably, brain activity data of each of them were collected using devices from the public market with dry electrodes, which might provide a signal of lower quality. But still, they are

very interesting for further research regarding peripheral philological signal analysis.

Current poor number of affected state datasets with physiological signals explained by difficulties and costliness of data collection and processing. However, with the arrival of cheap and accurate wearable sensors, the number of such databases might severely increase.

Table 2. Datasets of physiological signals for emotional state recognition [CASP17].

Dataset	**No. Part.**	**No. of Stimuli**	**Purpose**	**Modalities**	**Annotations**
DEAP [KMS+12]	32	40	Implicit affective tagging from EEG and peripheral physiological signals	EEG, GSR, Respiration Amplitude, Skin Temperature, Blood Volume, EMG and Electrooculogram. Visual for 22	Self-assessment of arousal, valence, liking, dominance and familiarity. participants
DECAF [ASK+15]	30	40	Affect recognition	MEG, Near-infra-red facial video, horizontal Electrooculogram, ECG and trapezius-Electromyogram	Self-assessment of valence, arousal and dominance. Continuous annotation of valence and arousal of the stimuli.
ASCERTAIN [SWA+17]	58	36	Personality and Affect	EEG, ECG, GSR and Visual	Big-Five personality traits, self-assessment of valence and arousal.
AMIGOS [CASP17]	40	36 & 16	Affect, personality, mood and social context recognition	Audio, Visual, Depth, EEG, GSR and ECG	Big-Five personality traits and PANAS. Self-assessment of valence, arousal, dominance, liking, familiarity and basic emotions. External annotation of valence and arousal.
MAHNOB-HCI [SLPP12]	27	20	Emotion recognition and implicit tagging	Visual, Audio, Eye Gaze, ECG, GSR, Respiration Amplitude, Skin temperature, EEG	Self-assessment of valence, dominance, predictability and emotional keywords. Agreement/disagreement with tags.

3 Methods

In this chapter, we describe the data and the methods used for the classification emotional state. First, a detailed specification of the multimodal database MAHNOB-HCI will be provided. Then, we introduce approaches for physiological signal preprocessing which involve filtration, annotation, feature extraction, and normalization. In the next section, heuristical methods for valence / arousal estimation will be described. In the last part, we will introduce several machine learning algorithms for data classification.

3.1 The Data

As mentioned in sections 2.4 and 2.3.3, MAHNOB-HCI [SLPP12] is a multimodal dataset which was collected for two purposes. First, to build a collection of emotional responses to videos, second, for implicit image tagging. In this work, we are considering only the first part that is intended for emotional recognition. 20 clips that were used for emotion elicitation were selected from 155 movie scenes. They were shown to more than 50 participants and rated in valence / arousal / dominance scales. Videos with the highest number of tags (i.e., video with the highest number of fear tags were selected to invoke fear state) [SLPP12] appeared in the final compilation. Some of them are from such popular movies like “Hannibal”, “The Pianist”, “Kill Bill”, “Gangs of New York”, etc. The duration of them was between 34.9 and 117 s long (M = 81.4 s; SD = 22.5 s). However, corresponding physiological records are padded by 30-second signals collected on the neutral video stimuli before and after the main video record. Originally, given dataset should include 540 sessions (27 participants and 20 effective videos). However, some sessions were corrupted, so 527 relevant records were obtained. Under the multimodal, we mean seven different sources of data that synchronously represent a state of the subject. Those modalities include: video of participants face recorded from several perspectives, eye gaze data, EEG, ECG, GSR, respiration, and temperature. All physiological signals (except EEG) were recorded with sampling rate 1025 Hz, afterwards, during post-processing, they were resampled to 256 Hz for computational efficiency. ECG signal contains data from 3 leads. However, only one, with the purest signal, was used – it was enough to annotate PQRST-complex and extract HRV features. EEG data were collected from 32 AgCl electrodes placed according to the international 10-20 system, which provided a comprehensive representation of the brain activity. Besides physiological response, self-assessed labels were collected. After each experimental session, a participant was interviewed using SAM technique described in section 2.1.2 in valence, arousal, dominance, predictability and nine discrete emotional labels (neutral, anxiety, amusement, sadness, joy (happiness), disgust, anger, surprise, and fear) [SLPP12].

Due to a high quality of gathered data, MAHNOB-HCI is widely used in emotional state recognition research. Soleymani et al. [SLPP12] provided a baseline for emotional

state recognition obtained on given dataset. By merging valence and arousal labels in three classes (low, medium and high). He achieved 46.2% / 45.5% on peripheral physiological data, 52.4% / 57.0% on EEG signal, 63.5% / 68.8% on Eye gaze and 67.7% / 76.1% on a fusion of previous two modalities. Further studies [WL17] in emotional state recognition based on peripheral physiological signals showed accuracies around 50% for separate patterns, for fusion they reached almost 55% and 56.83% in valence / arousal classes accordingly. Also, Ferdinando et al. [FSA17] in his studies hasn't achieved a significant improvement for valence classification, but for arousal, accuracy increased from 58.7% to 69.6%. Such increase resulted from the application of supervised dimensionality reduction techniques. Xu and Plataniotis [XP15] improved accuracies for EEG signal classification: they achieved 64.74% and 62.75% on three-class valence / arousal task by applying ANOVA [Dem60] feature selection mechanism. Studies, as mentioned earlier, are showing, that emotions are distinguishable via machine learning algorithms trained on physiological data, but unfortunately, accuracies are still not satisfying for industrial use.

3.2 Label extraction

MAHNOB-HCI provides several self-assessment indicators. In this work we were focused on the next three: self-assessed discrete emotion from the group of nine emotions mentioned in the previous section, self-assessed valence and arousal values (discrete values from 1 to 9) obtained using SAM described in the section 2.1.2. In both cases, classification into nine groups is an overwhelming task, taking into account the number of training samples and general complexity of emotional representation in the physiological signal. Thus, it was decided to reduce the number of classes for both valence and arousal into tree groups ("low", "medium" and "high"). In case of self-assessed valence and arousal, a technique is quite straightforward and is described in Table 3.

Table 3. Label mapping for self-assessed valence and aroual.

Affective state	**Self-assessed Valence and Arousal**
Low	1, 2, 3
Medium	4, 5, 6
High	7, 8, 9

For self-assessed emotional keywords, the procedure is a bit more complicated. As far as this approach refers to a discrete emotional paradigm, it has a mapping to the dimensional model provided by [FSRE07] (see Table 4).

Table 4. Label mapping for self-assessed emotional keywords.

Affective State	Mapping	
	To valence	To arousal
Low	fear, anger, disgust, sadness, anxiety	sadness, disgust, neutral
Medium	surprise, neutral	joy (happiness), amusement
High	joy (happiness), amusement	surprise, fear, anger, anxiety

Thus, two datasets with same class names were obtained for further comparison in machine learning classification.

3.3 Valence and Arousal Estimation from EEG Signal

Lots of studies claim that asymmetry of frontal hemispheres brain activity represents valence and arousal of emotions [Dav93]. This section will introduce several methods for valence / arousal classification based on EEG signal recorded during music listening. To estimate this activity, power spectrum bands (alpha and beta) from five EEG channels are calculated. These five channels are Fz, AF3, F3, AF4, F4 according to 10-20 system [Sil63] described in the Section 2.2.1.

3.3.1 Method Proposed by Kirke

This method was introduced by Kirke [KM11] in 2011. He proposed formula 1 and formula 2 for valence and arousal estimation. $left\_alpha$ and $left\_beta$ were calculated as a mean of spectral power recorded from channels AF3, F3 in 7-15 Hz and 16-31 Hz bands, while $right\_alpha$ and $left\_beta$ were calculated in the same way, but on the data from channels AF4, F4 channels.

$$valence = \log(left\_alpha) - \log(right\_alpha) \tag{1}$$

$$arousal = -(\log(right\_alpha) + \log(left\_alpha)) \tag{2}$$

Given method was tested on the data collected while participants listened to a sequence of ambient and hard rock music divided by silent fragments. To simulate different scales of valence / arousal, pitch and tempo of fragments were adjusted. There were three participants, and each was affected to achieve five valence states and five arousal states. Final classification accuracies if five classes were 80% and 70% for valence and arousal correspondingly.

3.3.2 Method Proposed by Vamvakousis

Another method for valence estimation using AF3, AF4, F3 and F4 channels was proposed by Vamvakousis [RV12] in 2012. A new formula shown in equation 3, calculates the difference of beta to alpha ratios from the left and right hemispheres.

$$valence = left\_beta/left\_alpha - right\_beta/right\_alpha \tag{3}$$

For estimation of arousal a brand new formula was introduced. It takes log_2 from the ratio of $front\_alpha$ to $front\_beta$, the formula shown in equation 4.

$$arousal = log_2(front\_beta/front\_alpha) \tag{4}$$

Given algorithm was tested on six subjects (3 females and 3 males), during the experiment where 12 sound stimuli (5-second records from the IADS database described in the section 2.3.2) were presented with 10 seconds rest before each stimulus. Tackling most common value obtained from 1-second fragments of corresponding EEG record, only 50% instances were correctly classified. Because of the unsatisfactory result, machine learning algorithm was applied on top of obtained values to improve performance.

3.3.3 First Method Proposed by Ramirez

A first method proposed by Ramirez in 2012 [VR12] had equation 5 and 6 for estimating valence and arousal respectively.

$$valence = alpha\_F4/beta\_F4 - alpha\_F3/beta\_F3 \tag{5}$$

$$arousal = alpha\_all/beta\_all \tag{6}$$

As described in the equations above, valence formula calculates the difference of alpha to beta ratios from the F4 and F3 channels. For arousal calculation, the ratio between spectral power means ($alpha\_all$ and $beta\_all$) taken from all channels (AF3, F3, AF4, F4). Unfortunately, an author not provided any information related to performance evaluation.

3.3.4 Second Method Proposed by Ramirez

A second method proposed by Ramirez in 2015 [RPLGV15] undergone some minor changes comparing to previous one. A formula for valence calculation, that is presented in equation 7, was slightly simplified. Now it is just a difference between alpha extracted from the F4 channel and beta from the F3 channel. For the arousal formula, that is described in equation 8, the only change is that numerator and denominator were swapped, so it is $beta\_all$ divided by $alpha\_all$.

$$arousal = alpha\_F4 - beta\_F3 \tag{7}$$

$$arousal = beta\_all/alpha\_all \tag{8}$$

This approach was tested on ten subjects. Ten sessions per subject were recorded, each 15 minutes long. During the session, selected music fragments were played with one second of silence in between. According to the experimental protocol, increased loudness and tempo make arousal and valence to increase correspondingly. The evaluation was performed regarding changes of the average valence / arousal values during the first and the last sessions. However, an accuracy of given method was not provided.

3.4 Preprocessing and Feature Extraction

Preprocessing plays a crucial role in the physiological data analysis pipeline. Measured signals may have different types of artifacts caused by sudden muscle movements, equipment, external inference and some other factors. These conditions impede further data processing. Fortunately, there are lots of techniques to eliminate most of the mentioned issues. The methods we used will be described further in this section. Feature extraction is also essential step before data classification. Correctly calculated and selected features will significantly affect the final score. In this section, we will describe the feature extraction approaches for EEG, ECG and other peripheral physiological signals we used in this work. The last topic of this section is data normalization. Besides the information about the emotional state, physiological signals contain lots of different insights considering subjects health condition. We have to take this information into account in order to eliminate physical properties and work only with emotional changes.

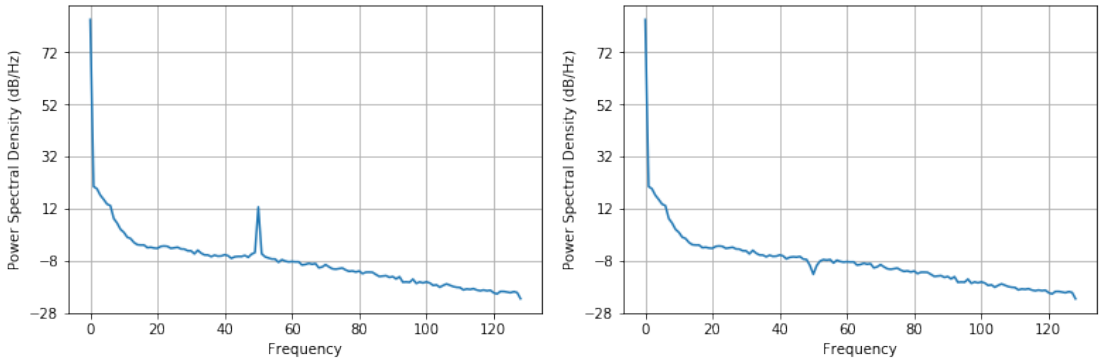
3.4.1 EEG Signal

During the EEG collection phase, internal factors could affect a signal, creating undesirable artifacts, that might impact classification results. In this section, the pipeline for electroencephalographic data processing will be described. In short, we can divide it into four stages:

1. Extraction of EEG signal from MAHNOB-HCI database
2. Remove baseline noise
3. Remove powerline noise
4. Calculate features using Fast Fourier Transform (FFT)

Data extraction was made using *bob* Python library. It has a function that provides the whole EEG record from MAHNOB-HCI databese by specifying a file and required channel name. For the second stage, a method called *detrend* from Python library scipy.signal was used. Baseline noise or trend is an artifact with much less frequency comparing to the normal range of signals fluctuation. In the EEG, such noise appears because of poor contact of electrodes, variations in temperature or bias in the equipment. To eliminate this problem, least-square fit [RN13] subtracts it from the initial data. Thirdly, powerline noise should be removed. It appears because of the electricity network, in frequencies 50 Hz and 60 HZ for Europe and the US correspondingly. These artifacts have a significant impact on the frequency domain, so aggravate further spectrum analysis of EEG. To avoid this issue, a technique called Notch-filter was used. It removes certain frequency component, by solving some difference equation, that has desired frequency as a parameter, and shows a relation between output and input signals. Results of the Notch-filter work are shown in Figure 7.

Figure 7. Results of the Notch-filter application.



(a) before filtration

(b) after filtration

The final step is to extract frequency domain representation of the input signal. We want to know, which amount of each frequency from some range is present in given sequence. To do that, an algorithm called fast Fourier transform (FFT) was used. According to this approach, decomposition of discrete Fourier transform (DFT) matrix is taken to computational efficiency [VL92], it reduces the complexity of computing from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$ where n is the length of the inoput signal. For its part, DFT calculates by a formula shown in equation 9, where x is an initial signal and X in an output of DFT).

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{2\pi i}{N} kn} = \sum_{n=0}^{N-1} x_n \cdot [\cos(2\pi kn/N) - i \cdot \sin(2\pi kn/N)] \tag{9}$$

FFT is a function that produces complex values, the absolute value of them shows the amount of that frequency present in the signal, imaginary part of those values represents the phase shift of the primary sinusoid in that frequency. To calculate Power Spectrum Density (PSD) from the obtained data, formula described in the equation 10 should be applied. In given equation, X_n is a spectrogram obtained after the FFT, and ν is a sampling rate of the initial signal.

$$PSD = \frac{|X_n|^2}{2\pi\nu} \tag{10}$$

PSD represents a distribution of power spread between frequency components that constitute a signal [SM+05]. To prepare EEG data for further machine learning classification, PSD is calculated on the short fragments sampled from all EEG channels by sliding window (commonly 1 second long with 25% overlapping). Afterward, PSDs obtained from the same channel shall be averaged. Thus, the 2d matrix where channels correspond to rows and frequencies correspond to columns will be flattened into a feature vector which would represent given record.

3.4.2 ECG Signal

ECG signal processing is quite specific because it requires detection of particular points (P, Q, R, S and T shown in Figure 5). The whole pipeline has the following structure:

1. Extract ECG signal from MAHNOB-HCI database
2. Remove baseline noise
3. Annotate R-peaks
4. Calculate distance between adjacent beats
5. Annotate other significant points (P, Q, S, T)
6. Calculate statistical features based on obtained points
7. Normalise obtained feature sets based on personal norms

MAHNOB-HCI dataset [SLPP12] contains ECG recorded in three leads, however, only single lead with the lowest amount of noise was used. Extraction of cardio records proceeds the same way as for EEG data described in section 3.4.1. Baseline drift is an artifact in ECG signal that hampers further processing and interpretation. It appears due to nature of skin conductance and some hardware features. To eliminate this issue, finite impulse response (FIR) filter was used. For each point of the input signal it calculates a "filtered value" using formula 11, where y contains filtered values, x is an initial signal,

b_i are mainly chose coefficients for removing noise from specific range, and N is the parameter called "filter order", that is also customizable.

$$y[n] = b_0 x[n] + b_1 x[n-1] + \cdots + b_N x[n-N] = \sum_{i=0}^{N} b_i \cdot x[n-i] \tag{11}$$

The third stage is R-peak detection. Implementation of the algorithm proposed by Hamilton [Ham05], which is a modification of well known Pan Tompkins [PT85] was used. It was chosen because of simplicity and computational efficiency. A considered algorithm has several main stages:

1. tacking the approximation of derivative of the signal,
2. tacking absolute value of the signal,
3. averaging obtained values over an 80 ms window,
4. detecting peaks in the result of the previous stage
5. applying the following rules to refine R-peak location:
   (a) ignore all peaks in the range of 400 ms from the highest peak,
   (b) if a peaks occurs, check if the raw signal has both positive and negative slopes; if not, this peak considered as baseline shift or noise,
   (c) if the peak is higher than the predefined detection threshold, mark it as R-peak, otherwise mark it as noise
   (d) if no peaks were detected in 1.5 lengths of the previous R-R interval, peak two times higher than threshold and at least 360 ms for the last R point might be detected as a new R-peak.

Having R-peaks, further feature extraction and ECG signal annotation can be done. Calculation of distances between adjacent beats can be done in several ways: first, calculate differences of indices of nearby R-peaks and work with this data, second, divide obtained distances by sampling rate, to get values in seconds. Q and S point correspond to minimum values in specific small range left and right from the R-peak. Obtaining P and T waves requires additional steps. Firstly, fragments with specific length should be taken towards left from Q point and toward the right from S point. Secondly, approximations of first and second derivatives should be taken to find the maximum value of considered section to find inflection point. Point obtained from the left segment is P and point obtained from the left segment is T.

Having annotation of whole ECG signal, following statistical features were calculated:

- mean, standard deviation, median, minimum, maximum, range (max-min) of distances between P, Q, R, S, T points of adjacent beats and the same features for the length of PQ, QS and ST segments;
- mean, standard deviation, median, minimum, maximum from ECG signal amplitude;
- pNN50 is a feature which shows the average number of times when interbeat interval surpass average interval by 50 milliseconds, studies are saying, that this measure helps to track parasympathetic activity [ENT84];
- mean, standard deviation, median, minimum, maximum, range of the HRV-distribution.

Afterwards, the obtained features will be combined in a single vector for further processing.

3.4.3 GSR, Respiration and Temperature Data

Galvanic skin response, respiration, and temperature are other physiological signals that carry valuable information for emotional recognition. Their processing is simpler comparing to ECG or EEG, but still requires several steps that will be described further. First of all, as mentioned in section 3.1, 30 seconds in the beginning and at the end of each record imply neutral state. Thus, a baseline for all videos was calculated on those first segment and subtracted from the reaming part. Afterward, for noise reduction purposes, Butterworth filter with cutoff frequencies 0.3 Hz and 1 Hz was applied to GSR and respiration data correspondingly. Next step was to extract the following features from considered signals. For GSR:

- mean, standard deviation, median, minimum, maximum of the signal,
- mean, standard deviation, median, minimum, maximum, minRatio, maxRatio of approximations of first and second derivatives of the signal,
- minRatio and maxRatio of GSR signal (the number of maximums and minimums divided by the length of the signal).

For the respiration data, following features were calculated:

- mean, standard deviation, median, minimum, maximum, range of the signal,
- mean, standard deviation, median, minimum, maximum, range maxRatio of the approximations of the first and second derivatives,

- mean, standard deviation, median, minimum, maximum, maxRatio of the pulse of the signal,
- mean, standard deviation, minimum, maximum, range, maxRatio of approximations of first and second derivatives of the signal,
- mean of the spectral power from [0, 0.1], [0.1, 0.2], [0.2, 0.3] and [0.3,0.4] Hz frequency bands,
- mean, standard deviation, minimum, maximum, range, maxRatio of signals amplitude, approximations of first and second derivatives of the amplitude of the signal.

maxRatio is a ratio of the number of extrema to the whole signal length, respiration pulse is a set of maximums of the respiration signals, mean of the pulse during the minute is called respiration rate. Temperature features contained mean, standard deviation, minimum, maximum and range normalized by Z-score (subtracted mean and divided by standard deviation). The mentioned features were described in papers by Wiem [WL17] and Wagner [WKA05].

As well as for ECG, GSR, respiration and temperature features can be gathered into a vector with no regards to different modalities. In the scientific literature it is called feature-level fusion [RG05].

3.4.4 Data Normalization

Usually, in machine learning and statistics, normalization is a procedure of rescaling the data using some mathematical transformations. To align a distribution of data to a normal distribution, mean μ is subtracted from all data. Then, this value is divided by a standard deviation σ of all data points as shown in equation 12.

$$X_{norm} = \frac{X - \mu}{\sigma} \tag{12}$$

However, this approach does not quite fit user data, because, as were mentioned earlier, features extracted from the physiological signals contain information about both emotional and physical state. Therefore, to eliminate these individual properties, the idem formula as in the equation 12 was used, but each user's features were normalized by mean and standard deviation calculated within the samples of the same user.

3.5 Data Classification

This section describes machine learning algorithms for solving classification task that were employed in this work. Each method performs in a particular way, having various

complexity, learning and prediction procedures, etc. Because of the mentioned reasons, their performance may vary. To obtain the best possible result, a set of approaches have to be considered.

3.5.1 Support Vector Machine

Support vector machine is a supervised algorithm that fits for solving both classification and regression tasks. The goal of SVM is to set up a hyperplane which will separate all training samples in two classes. This hyperplane if described by equation 13.

$$f(x) = w^t x + b \tag{13}$$

Where w^t is a two-dimensional vector according to the number of classes and b is bias. For multidimensional case, special functions called kernels (14) applied to input x as shown in equation 15.

$$k(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) \tag{14}$$

$$f(x) = w^t \varphi(x) + b \tag{15}$$

They project the data to higher-dimensional space, making it linearly separable, this technique also called “kernel trick”. The most used kernels are shown in Table 5.

Table 5. Common SVM kernels.

Name	**Formula**
Polynomial	$k(x_i, x_j) = (x_i \cdot x_j + 1)^d$
Gaussian	$k(x_i, x_j) = \exp\left(-\frcd{\|\|x_i - x_j\|\|^2}{2\sigma^2}\right)$
RBF	$k(x_i, x_j) = \exp\left(-\gamma \|\|x_i - x_j\|\|^2\right)$
Hyperbolic tangent	$k(x_i, x_j) = \tanh\left(x_i \cdot x_j - \delta\right)$

The most fitting hyperplane will be the one, that makes maximum margin between classes. The decision rule is based on the equation of the obtained hyperplane and forms in the next way: $sign$ function is applied to the mentioned equation with substituted values of the point to be classified. An obtained value will be -1 or 1 which would say about belonging to one or another class. For multiclass classification, an approach called one-vs-all is used, it means, that multiple hyperplanes are built to separate each class from all others, joint together.

3.5.2 Random Forest

Random Forest also called as random decision forest is a method that is based on decision trees and widely used for classification and regression. It was introduced by Ho in 1995 [Ho95], but algorithm had some shortcomings. In 2001 Breiman [Bre01] improved that and published a modified version that is still used. To describe the mechanics of Random Forest, firstly, it should be clarified how a decision tree works. A decision tree is a binary tree, the primary purpose of this algorithm is to find such variable-value pair from the training data, which will provide optimal (by some criteria) split. Recursively repeating described procedure for newly appeared subsets, the algorithm proceeds until next maximum depth was reached or subset can be no longer divided (such subset is called leaf), so optimal tree is found. To estimate a new split, the algorithm uses one of following metrics:

- information gain
- Gini impurity
- mean square

Selection of the metrics depends on the task. Single decision tree tends to overfit training data, to overcome this issue, a technique called ensemble is used. Ensemble method makes predictions by combining results of the set of classifiers, in case of Random Forest, it unites decision trees which are “weak” learners into one “strong” learner, as Figure 8 illustrates. Depending on the task, different methods for predictions aggregation can be used. For instance, in classification tasks, it takes mode or most frequent class, and for regression, it averages predictions.

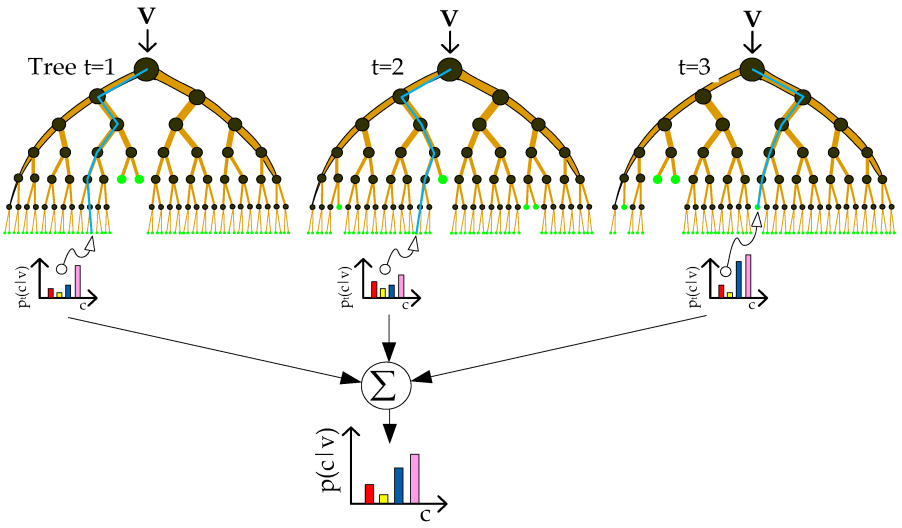


Figure 8. A Random forest algorithm classification [SLSH17].

As described earlier, a considered algorithm has word forest in its name because of a set of trees that it uses for predictions. Whereas, word random stands for a technique called Bagging or Bootstrap. It is used to train random forest; the main idea is to sample random subset of training data and features and learn particular decision tree based on it. Despite the fact that RF is a reliable classifier, it still has some weaknesses, for instance, in solving regression tasks, because of inability to predict beyond the range of training data. Also, it may significantly overfit noisy datasets.

3.5.3 Multilayer Perceptron

A multilayer perceptron (MLP) is a supervised learning algorithm that belongs to the class of feed-forward neural networks. MLP comprises simple hierarchically connected elements called artificial neurons. These elements are roughly simplified models of biological neurons [Ros61]. A function of a single artificial neuron – to produce a single value based on an input vector, as depicted in the Figure 9.

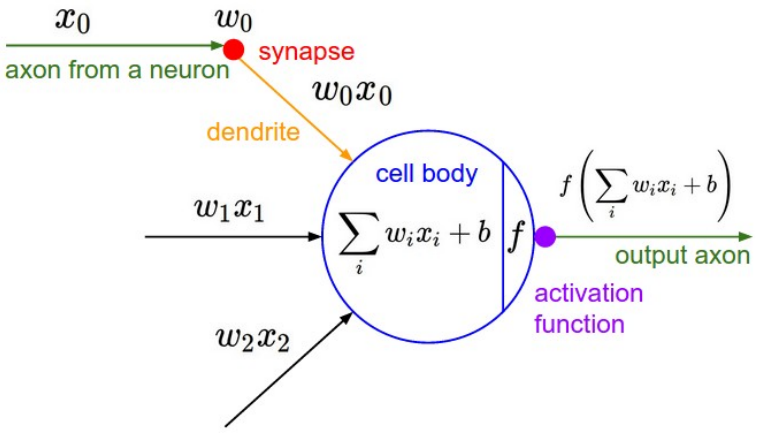


Figure 9. An artificial neuron model [cs2].

To obtain this value, a special function called activation function applies to the dot-product of the input signal with the neuron's weight vector plus bias term. Weights and bias are parameters of a neuron that adjust their values during the training phase to provide best possible classification results. There are several activation functions that used in practice. Sigmoid is initial one, its formula 16 and plot 10 are presented below. As seen, it produces outputs in a range [0, 1].

$$f(x) = \frac{1}{1 + e^{-x}} \tag{16}$$

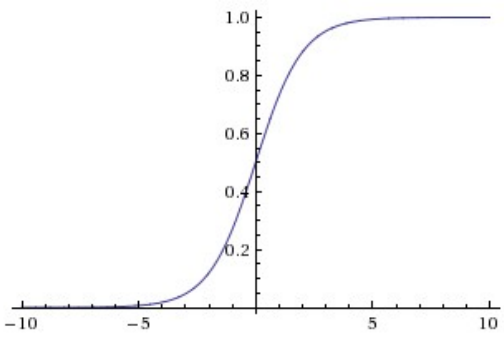


Figure 10. Sigmoid function plot [cs2].

Another activation function is called hyperbolical tangent (tanh), its domain is in [-1, 1]. However, they provide very small gradient is the input value if far from the origin (very small or very big). This phenomenon sometimes called “gradient vanishing”. To solve this problem rectified linear unit (ReLu) was used 17. It was applied by Glorot et al. [GBB11] for deep neural networks training and showed promising results.

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \tag{17}$$

MLP can be presented as an acyclic graph with neurons as vertices and connections between the neurons as edges, so outputs of a neuron on the given layer will be inputs of neurons on the next layer. Important to note, that neurons on the same level have no connections within the layer. Graphical representation of three-layer MLP is shown in Figure 11.

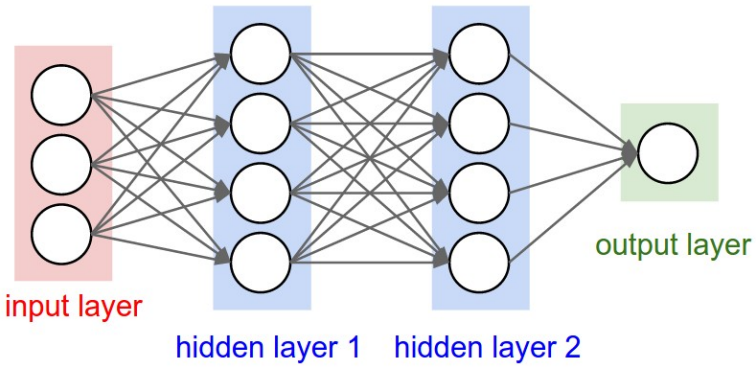


Figure 11. A three-layer neural network [cs2].

Mostly, the last layer of MLPs do not have activation, instead, depending on the task, a particular function for classification or regression is used. For mentioned learning

procedure, MLP uses learning technique called backpropagation, it applies iterative adjustment of weights until achievement of global minima approximation of task-specific function that is also called loss function. According to universal approximation theorem, a single-layer neural network with the finite number of nodes can approximate any continuous function [Hor91].

3.5.4 ANOVA Feature Selection

Analysis of variance (ANOVA) is a statistical technique for comparing means in three and more groups of numerical values (to perform it on two groups, Student t-test can be applied). To perform ANOVA, following hypothesis should be tested: a null-hypothesis says that means of n classes are equal (equation 18), while an alternative hypothesis assumes that at least one group with the statistically significant different mean (equation 19).

$$H_0 : \mu_1 = \mu_2 = ... = \mu_n \tag{18}$$

$$H_1 : \text{At least one of the } \mu \text{ is different.} \tag{19}$$

For hypothesis test, F-value has to be estimated by calculation of the ratio between variances: the first one is an inter-group variance and the second one is a variance within a group. Afterward, obtained F-value and corresponding degrees of freedom are used to calculate a p-value. If the mentioned p-value is less then α (mostly 0.05), the alternative hypothesis will be accepted. Concerning feature selection, the described algorithm applies to all features in the dataset by dividing each into groups by classes (labels) [BGC14]. Features that are not significantly different regarding labels will be thrown from the dataset. So, given method allows getting rid of features with low variability for the corresponding set of labels.

3.5.5 Recurrent Neural Networks

Recurrent neural networks (RNNs) are a specific type of neural networks intended to process sequential data. They are capable of both classification and regression tasks. This approach was invented by Hopfield [Hop82] in 1982. Further, in 1993 Schmidhuber [Sch92] improves this concept by introducing Long-short tern memory (LSTM) which is still considered as state of the art technique for sequence processing.

A simplified explanation of RNNs can be formulated in the following way. Such network has some “memory” about computations on the previous steps and uses it to process subsequent elements. Schematic representation of RNN is shown in Figure 12.

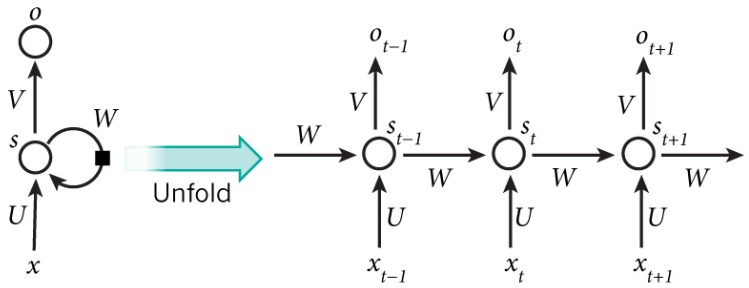


Figure 12. Folded and unfolded graphical representation of RNN [LBH15].

Unfolding the RNN means depicting recurrent neurons for each element of the sequence. According to the image above, x_t is the input for time point t (x_t can be a single number as well as vector); s_t is the hidden state for time point t In general, it is the mentioned "memory" of the network and calculates based on the current input x_t and previous hidden state s_{t-1} by formula $s_t = f(Ux_t + Ws_{t+1})$. f is an activation function. Mostly, it's tanh or ReLu described in section 3.5.3. o_t is output at the time t; it can be forwarded to another recurrent layer or function selected for the exact task. In contradiction to other types of neural networks, RNN uses the same parameters (U, W, V) for all steps, which means that the same procedure repeats with different inputs. This significantly reduces the number of parameters to learn but brings some disadvantages. s_t can not "remember" information captured from earlier steps. LSTMs that were mentioned above can mitigate this problem. LSTMs have no difference in architecture comparing to RNNs. However, they have a different structure of the hidden state which is shown in Figure 13.

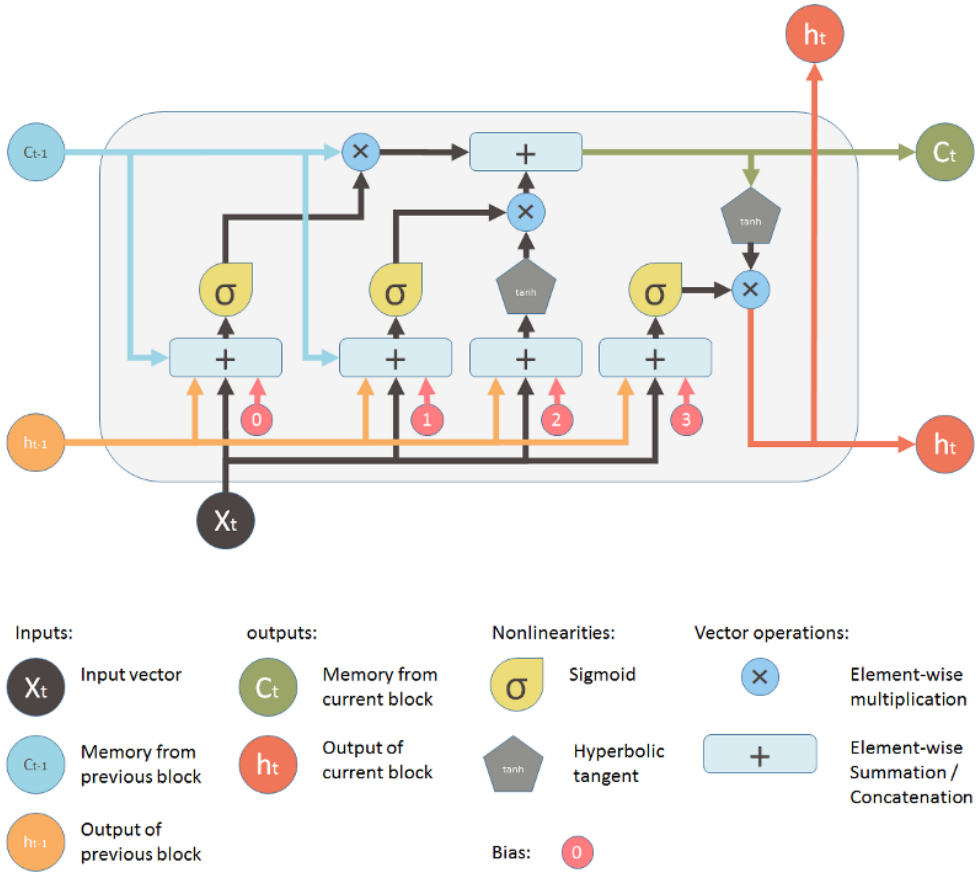


Figure 13. A hidden state structure of the LSTM block [Yan16].

In contrast to RNN, LSTM block outputs not only h_t values, but C_t values that provide long-term memory. One more difference, which also shown on Figure 13, is that LSTM output h_t goes to entire model output like in RNNs and to the next block simultaneously, this mechanism provides short memory to considered model. This type or recurrent units are very effective in the processing of long-term sequences.

4 Results chapter

In this chapter we will present comparison of results of physiological signal classification using the data sets that were described in Chapter 3.2. First, formula-based methods for valence arousal estimation from EEG signal, presented in the Chapter 3.3, were tested. As expected, they showed not promising results, so we instantly proceed to the machine learning approach. Next, random forest, SVM, multilayer perceptron and LSTM (See Chapter 3.5) classifiers were used. To evaluate classification performance we used a technique called Leave-one-out cross-validation. This technique implies testing the

model on the records of the single subject, whereas training was performed on all records from remaining subjects. Thus, having 26 users (initially it was 27, but one user had corrupted ECG and GSR signals, so we entirely removed him from the experimental setup), each training was held on 25 subjects (approximately 490 videos, because several users missed certain records) and tested on the one subject. The process is repeated 26 times and the average classification score is the estimate of the performance of a method.

4.1 Formula-based Methods Evaluation

Methods which estimate valence and arousal values from the EEG signal (See Chapter 3.3) require the calculation of average PSD in specific bands from five channels. In our experiments, for each record, we computed valence and arousal values from short fragments of EEG (from 800 to 1200 ms by sliding window on the signal), using mentioned formulas. Before converting given numbers to the low, medium and high classes, they were scaled according to the formula in the equation 20, where X_{usr} is a set of all obtained values from an exact subject.

$$x_{norm} = \frac{x - min(X_{usr})}{max(X_{usr}) - min(X_{usr})} \tag{20}$$

Thus, we took into account, that each user experienced minimal and maximal valence and arousal in different videos. After the scaling, we calculated the histograms for each record by three groups, to estimate density in low, medium and high notations. Example of the obtained histograms is shown in the Figure 14.

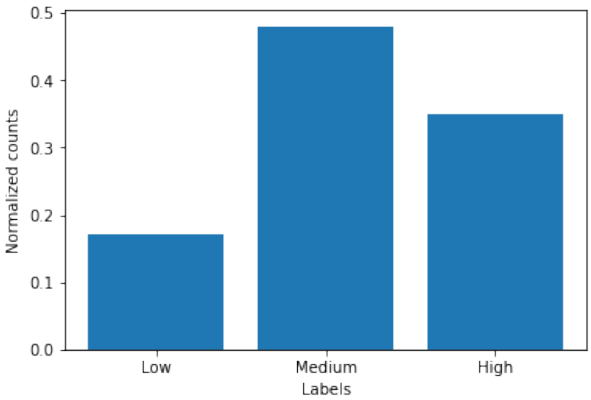


Figure 14. The histogram of the density of classes.

Taking the most represented class from the histogram as a whole record label is quite intuitive. However, we noticed, that the “medium” is dominant for most records. To consider this fact in the classification, we made our final predictions by dividing

elementwise a histogram of the record with an averaged histogram of the given user and picking a class that refers to a maximum of tree obtained values. Results of described classification using labels extracted from the self-assessed valence / arousal values and from the self-assessed discrete emotional keywords are shown in the Table 6 and Table 7 correspondingly.

Table 6. Formula-based classification results using labels from the self-assessed valence / arousal values. Chance level is 0.33.

Method	**Window Size (seconds)**	**Accuracies**	
		Valence	**Arousal**
By Kirke	800	33.07	31.55
	1000	30.22	33.46
	1200	26.99	32.12
By Vamvakousis	800	30.60	33.46
	1000	31.74	33.26
	1200	29.65	32.69
First by Ramirez	800	32.50	32.69
	1000	29.27	34.60
	1200	26.80	33.07
Second by Ramirez	800	34.03	34.79
	1000	33.07	34.60
	1200	35.17	35.93

Table 7. Formula-based classification results using labels from the self-assessed discrete emotional keywords.

Method	Window Size (seconds)	Accuracies	
		Valence	Arousal
By Kirke	800	33.07	37.07
	1000	29.65	35.74
	1200	29.65	35.17
By Vamvakousis	800	31.93	34.98
	1000	32.50	35.55
	1200	32.12	34.79
First by Ramirez	800	32.12	34.03
	1000	26.80	31.55
	1200	26.23	342.20
Second by Ramirez	800	33.65	35.17
	1000	31.93	37.26
	1200	33.26	34.79

According to the Table 7, we concluded that described approach does not work because the accuracies are not different from random chance. There might be several reasons for such results. First, mentioned formulas were initially designed to analyze EEG recorded during audio stimuli, whereas our dataset was collected using video stimuli which might cause different response of the brain. Another reason is that for usage of given formula, we need an exact timestamp of the emotional elicitation, to calculate a valence / arousal value directly after the stimuli. Unfortunately, for MAHNOB-HCI we don't have such opportunity.

4.2 Machine Learning Methods Evaluation

4.2.1 EEG Classification

Using EEG processing technique described in the Chapter 3.4.1, we obtained 32×69 features for 512 records for all subjects. ANOVA feature selection technique was used after the feature extraction stage. Results of for both valence and arousal classification using there algorithms and two label sets (extracted from the self-assessed valence / arousal values (LS1) and from the self-assessed discrete emotional keywords (LS2)) are shown in the Table 8 .

Table 8. EEG signal classification results using Leave-one-out cross-validation.

Classifier	Accuracy			
	Valence		Arousal	
	LS1	LS2	LS1	LS2
Random forest	40.72	**49.31**	38.52	**52.19**
SVM	40.38	43.17	40.26	44.86
MLP	39.03	40.29	39.97	44.31

We reached the highest accuracy for valence and arousal (49.32% and 52.19%) on the second label set (LS2). More detailed results of top classifiers are shown on the figure 15.

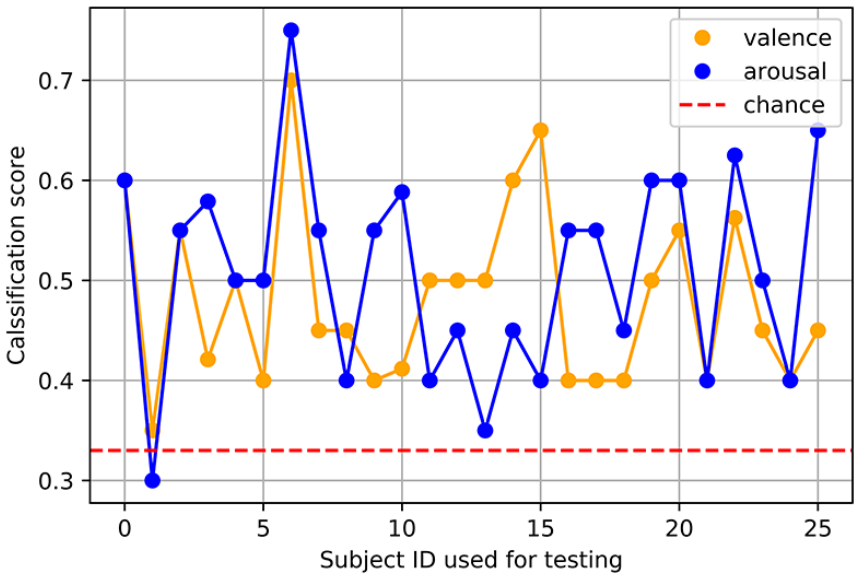


Figure 15. Top EEG classification rates using Leave-one-out cross-validation.

Based on presented plot we concluded, that specific subjects have more distinguishable or imperceptible brain response to affective videos, whereas the majority of participants valence and arousal can be predicted with around 50% accuracy.

4.2.2 ECG Classification

ECG was classified by 75 features described in the Chapter 3.4.2 and 26 subjects (20 samples for each user). Feature selection was made by ANOVA procedure. As for

EEG, we used two label sets and three classifiers. Leave-one-out subject cross-validation results are shown in the Table 9.

Table 9. ECG signal classification results using Leave-one-out cross-validation.

Classifier	**Accuracy**			
	Valence		**Arousal**	
	LS1	LS2	LS1	LS2
Random forest	37.19	39.34	37.35	40.40
SVM	**40.38**	38.73	40.26	**43.90**
MLP	38.70	38.37	38.66	**43.90**

We reached maximum valence classification score 43.9% on the first label using SVM algorithm with radial basis function kernel. Also, we obtained equally top scores (43.90) for arousal on the second label set.

4.2.3 Peripheral Signals Classification

Under peripheral signals, we mean GSR, respiration and temperature data. 89 features were extracted from the mentioned modalities and classified as a single vector. Feature selection technique, classification algorithms, and label sets were taken from the corresponding EEG and ECG experiments. Obtained results are presented in the Table 10.

Table 10. Peripheral signals classification results.

Classifier	**Accuracy**			
	Valence		**Arousal**	
	LS1	LS2	LS1	LS2
Random forest	40.75	**43.32**	41.68	**50.36**
SVM	40.38	39.33	40.26	43.90
MLP	39.75	39.35	40.46	44.09

In the given experiment, we obtained 43.32% accuracy for valence classification and 50.36% for arousal classification, both using random forest and second random set.

After three experiments, we decided to find out, which modality provides highest arousal accuracy for each user. Comparison results are shown on the Figure 16.

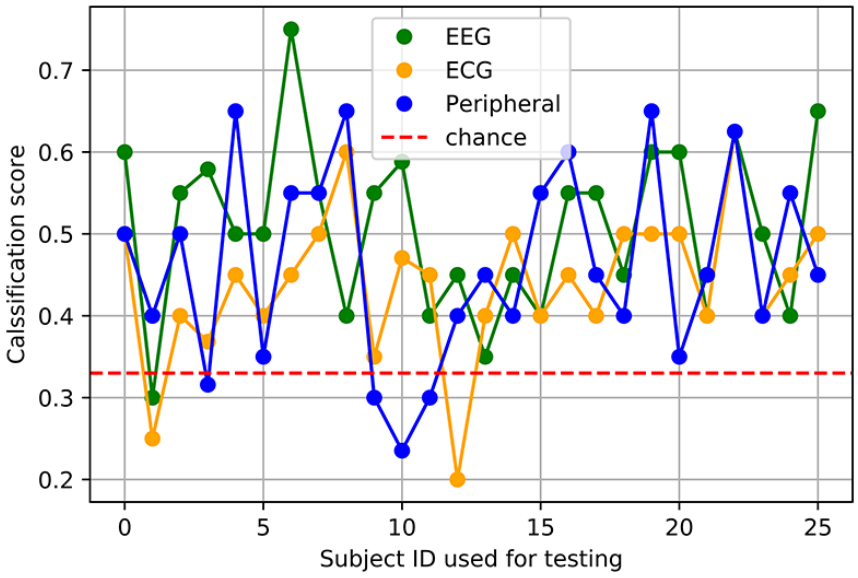


Figure 16. Valence classification results by tree modalities.

As expected, ECG and Peripheral results are visually more related. It can be explained because they are both regulated by the sympathetic nervous system, while EEG measures cortex activity.

4.2.4 ECG+Peripheral Singnals Classification

Based on results obtained in previous experiments, we decided to combine ECG, GSR, respiration and temperature features and classify them by described earlier pipeline. Acquired results are shown in the Table 11.

Table 11. ECG + Peripheral signals classification results.

Classifier	**Accuracy**			
	Valence		**Arousal**	
	LS1	LS2	LS1	LS2
Random forest	40.17	**41.22**	39.93	**46.25**
SVM	40.38	39.33	40.26	43.90
MLP	40.57	38.94	40.26	43.90

As we can see, the combination of modalities worsened our results. The accuracy of the peripheral signals arousal classification was 50.36%, since, union with ECG provided

46.25%. Despite that accuracy increased comparing to a single ECG classification. However, we have to say that the fusion of cardio with peripheral features does not provide any improvements. It is the same situation for the valence classification.

4.2.5 ECG Classification Using RNN

After receiving unsatisfactory results in ECG features classification, we proposed an alternative method based on recurrent neural networks. RNN's imply sequence. Thus, vectors of 50 RR-intervals were extracted from each ECG records. We applied the same normalization approach as were used in earlier experiments. However, mean and standard deviation were calculated using all intervals of given subject, and each interval was normalized by obtained values. Afterward, we trained RNN and LSTM models on two known feature sets for valence and arousal. The evaluation was held by the idem Leave-one-out subject cross-validation. Results are shown in the Table 12.

Table 12. ECG + Peripheral signals classification results.

Classifier	**Accuracy**			
	Valence		**Arousal**	
	LS1	LS2	LS1	LS2
RNN	34.13	**36.96**	30.55	33.81
LSTM	31.98	32.17	37.61	**38.25**

We concluded, that given approach not properly worked for valence estimation. However, for arousal prediction, LSTM model provided 38.25% accuracy, which is even lower than results from the first experiment, but higher than random choice. It must be said, that for some users they are higher or equal. (See Figure 17). Thus, we don't have to reject this method. It can be improved by adding more training samples or the increase of the input sequence length.

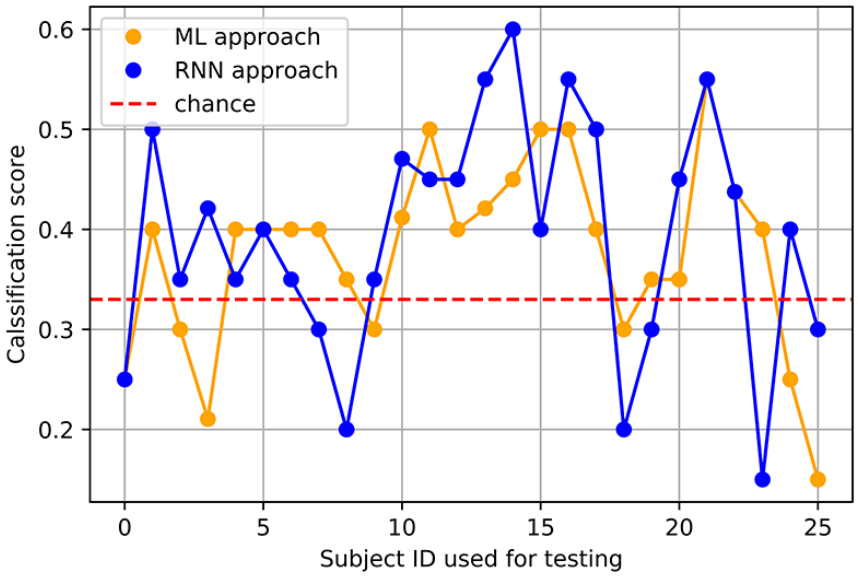


Figure 17. Leave-one-out cross-validation results of two ways of ECG classification.

5 Conclusion

The main goal of this thesis was to compare the methods for emotional state recognition based on physiological signals and several ways for categorization of emotions. To solve emotional recognition task we used heuristic and machine learning classification techniques. For emotion categorization, we considered Eckman's paradigm of discrete emotional states and dimensional model which represents emotions by matching them onto the Cartesian plane with valence and arousal axis.

Then, we introduced contemporary methods for emotional elicitation and subjects feedback collection. Two mentioned emotional paradigms have corresponding self-assessment gathering technique. Thus, we decided to compare how they affect on emotional state recognition. For these purposes, we needed a dataset, so an overview of existing human affective states databases was provided. Afterward, MAHNOB-HCI database was selected because of its quality and a wide range of modalities. We attempted to classify physiological signals from the dataset in low, medium and high valence and arousal, which, in turn, were extracted from the self-assessed discrete emotional keywords and from the valence / arousal scale provided by SAM.

Performing classification experiments, we obtained best highest leave-one-out accuracy scores for EEG and peripheral signals classification using machine learning. Usage of the label set extracted from the discrete emotional keyword provided better results in most experiments. This fact can be accounted in further datasets collections and processing. Also, we proposed a novel approach for heart rate data classification using RNNs. However, it was not successful because of the small size of the dataset.

For the future work, we are planning to test considered methods on the datasets with the higher number of subjects and trials for each person. Another thing to try is to collect and analyze a dataset with the precisely known moment of emotional invocation. In our opinion, it can be reached by using short, same length affective videos or affective images.

6 Acknowledgments

I gratefully acknowledge Ilya Kuzovkin and Raul Vicente Zafra for their guiding and support while writing this thesis. Also, I would like to thank Mohammad Soleymani from the University of Geneva for collecting MAHNOB-HCI database. Last but not least, I have to thank Wiem Mimoun Ben Henia from the Tunis El Manar University who provided insight and expertise that greatly helped the research.

References

[AG15] Priyanka Abhang and Dr.Bharti Gawali. Correlation of eeg images and speech signals for emotion analysis. 10:1–13, 01 2015.

[ASK+15] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe. Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing*, 6(3):209–222, July 2015.

[ATAS88] Allen, Chris T., Karen A., and Susan S. On assessing the emotionality of advertising via izard's differential emotions scale, Jan 1988.

[BAGC86] R. D. Berger, S. Akselrod, D. Gordon, and R. J. Cohen. An efficient algorithm for spectral analysis of heart rate variability. *IEEE Transactions on Biomedical Engineering*, BME-33(9):900–904, Sept 1986.

[BDCC15] Y. Baveye, E. Dellandréa, C. Chamaret, and L. Chen. Liris-accede: A video database for affective content analysis. *IEEE Transactions on Affective Computing*, 6(1):43–55, Jan 2015.

[BF17] Teah-Marie Bynion and Matthew Feldner. Self-assessment manikin, 01 2017.

[BGC14] Mahdi Bejani, Davood Gharavian, and Nasrollah Moghaddam Charkari. Audiovisual emotion recognition using anova feature selection method and multi-classifier neural networks. *Neural Computing and Applications*, 24(2):399–412, Feb 2014.

[BL94] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59, 1994.

[Boi98] Frans A Boiten. The effects of emotional behaviour on components of the respiratory cycle. *Biological Psychology*, 49(1):29 – 51, 1998.

[Bre01] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[CASP17] Juan Abdon Miranda Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. Amigos: A dataset for mood, personality and affect research on individuals and groups. *CoRR*, abs/1702.02510, 2017.

[CLGÁG12] Sandra Carvalho, Jorge Leite, Santiago Galdo-Álvarez, and Óscar F. Gonçalves. The emotional movie database (emdb): A self-report and

psychophysiological study. *Applied Psychophysiology and Biofeedback*, 37(4):279–294, Dec 2012.

[cs2]

[Dav92] M Davis. The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience*, 15(1):353–375, 1992. PMID: 1575447.

[Dav93] Richard J Davidson. The neuropsychology of emotion and affective style. 1993.

[DCCM+11] Ellen Douglas-Cowie, Cate Cox, Jean-Claude Martin, Laurence Devillers, Roddy Cowie, Ian Sneddon, Margaret McRorie, Catherine Pelachaud, Christopher Peters, Orla Lowry, Anton Batliner, and Florian Hoenig. The humaine database, 10 2011.

[Dem60] A. P. Dempster. Henry scheffé, the analysis of variance. *Technometrics*, 2(4):517–517, 1960.

[DGS11] Elise S. Dan-Glauser and Klaus R. Scherer. The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, 43(2):468, Mar 2011.

[DPGS12] Claire-Hélène Demarty, Cédric Penet, Guillaume Gravier, and Mohammad Soleymani. A benchmarking campaign for the multimodal detection of violent scenes in movies. In Andrea Fusiello, Vittorio Murino, and Rita Cucchiara, editors, *Computer Vision – ECCV 2012. Workshops and Demonstrations*, pages 416–425, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[EF71] Paul Ekman and Wallace V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129, 1971.

[Ekm93] Paul Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, 1993.

[ENT84] DAVID J Ewing, JM Neilson, and PAUL Travis. New method for assessing cardiac parasympathetic activity using 24 hour electrocardiograms. *Heart*, 52(4):396–402, 1984.

[FSA17] Hany Ferdinando, Tapio Seppänen, and Esko Alasaarela. Enhancing emotion recognition from ecg signals using supervised dimensionality reduction, 01 2017.

[FSRE07] Johnny RJ Fontaine, Klaus R Scherer, Etienne B Roesch, and Phoebe C Ellsworth. The world of emotions is not two-dimensional. *Psychological science*, 18(12):1050–1057, 2007.

[GBB11] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

[GL95] James J. Gross and Robert W. Levenson. Emotion elicitation using films. *Cognition and Emotion*, 9(1):87–108, 1995.

[Ham05] P Hamilton. Open source ecg analysis software documentation. 2002. *EP Limited*, 2005.

[Ho95] Tin Kam Ho. Random decision forests. In *Document analysis and recognition, 1995., proceedings of the third international conference on*, volume 1, pages 278–282. IEEE, 1995.

[Hop82] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[Hor91] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[HP05] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, June 2005.

[img18] Qrs complex, May 2018.

[Imo] Imotion. Gsr recording procedure. [Online; accessed May 1, 2018].

[Iza77] Carroll E. Izard. *Human emotions*. Plenum Press, New York, 1977.

[JBC14] Brendan Jou, Subhabrata Bhattacharya, and Shih-Fu Chang. Predicting viewer perceived emotions in animated gifs. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 213–216, New York, NY, USA, 2014. ACM.

[Kap11] Arvid Kappas. Emotion and regulation are one! *Emotion Review*, 3(1):17–25, 2011.

[KLB17] Benedek Kurdi, Shayn Lozano, and Mahzarin R. Banaji. Introducing the open affective standardized image set (oasis). *Behavior Research Methods*, 49(2):457–470, Apr 2017.

[KM11] Alexis Kirke and Eduardo R Miranda. Combining eeg frontal asymmetry studies with affective algorithmic composition and expressive performance models. In *ICMC*, 2011.

[KMS+12] Sander Koelstra, C. Mühl, Mohammad Soleymani, Jung Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Antinus Nijholt, and Ioannis Patras. Deap: A database for emotion analysis using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2012. eemcs-eprint-21368.

[Kon01] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89 – 109, 2001.

[Lan80] P. J. Lang. Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson, and T. H. Williams, editors, *Technology in Mental Health Care Delivery Systems*, pages 119–137. Ablex, Norwood, NJ, 1980.

[LBC99] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (iaps): Technical manual and affective ratings. Technical report, University of Florida, 1999.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[LHCJL85] Robert L. Hodes, Edwin Cook, and Peter J. Lang. Individual differences in autonomic response: Conditioned association or conditioned fear? 22:545–60, 10 1985.

[LZJ+08] Sheng Lu, He Zhao, Kihwan Ju, Kunson Shin, Myoungho Lee, Kirk Shelley, and Ki H. Chon. Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information? *Journal of Clinical Monitoring and Computing*, 22(1):23–29, Jan 2008.

[Man] Management Mania. Six basic emotions. [Online; accessed May 1, 2018].

[Mer87] Peter F. Merenda. Toward a four-factor theory of temperament and/or personality. *Journal of Personality Assessment*, 51(3):367–374, 1987. PMID: 16372840.

[MH01] Yuri Masaoka and Ikuo Homma. The effect of anticipatory anxiety on breathing and metabolism in humans. *Respiration Physiology*, 128(2):171 – 177, 2001.

[Mil02] David Millet. The origins of eeg. in 7th annual meeting of the international society for the history of the neurosciences (ishn), 2002.

[MR74] Albert Mehrabian and James Russell. An approach to environment psychology. 01 1974.

[MŻJG14] Artur Marchewka, Łukasz Żurawski, Katarzyna Jednoróg, and Anna Grabowska. The nencki affective picture system (naps): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, 46(2):596–610, Jun 2014.

[New80] P. P. Newman. *Physiology of Emotion*, pages 449–461. Springer Netherlands, Dordrecht, 1980.

[NLdS05] 1920 Niedermeyer, Ernst and 1935 Lopes da Silva, F. H. *Electroencephalography : basic principles, clinical applications, and related fields*. Philadelphia ; London : Lippincott Williams Wilkins, 5th ed edition, 2005. Includes bibliographical references and index.

[NTD97] Ivan Nyklíček, Julian Thayer, and Lorenz Doornen. Cardiorespiratory differentation of musically-induced emotions. 11:304–321, 01 1997.

[NVG+15] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. P. Scilingo. Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, 6(4):385–394, Oct 2015.

[NWC13] Nargess Nourbakhsh, Yang Wang, and Fang Chen. Gsr and blink features for cognitive load classification. In Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2013*, pages 159–166, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[OAHT17] P-H. Orefice, M. Ammi, M. Hafez, and A. Tapus. Design of an emotion elicitation tool using vr for human-avatar interaction studies. In Jonas Beskow, Christopher Peters, Ginevra Castellano, Carol O'Sullivan, Iolanda Leite, and Stefan Kopp, editors, *Intelligent Virtual Agents*, pages 335–338, Cham, 2017. Springer International Publishing.

[Phi93] Pierre Philippot. Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and Emotion*, 7(2):171–193, 1993. PMID: 27102736.

[PT85] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 1985.

[RG05] Arun A Ross and Rohin Govindarajan. Feature level fusion of hand and face biometrics. In *Biometric Technology for Human Identification II*, volume 5779, pages 196–205. International Society for Optics and Photonics, 2005.

[RHH+17] Pranav Rajpurkar, Awni Y. Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y. Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. *CoRR*, abs/1707.01836, 2017.

[RMMI+14] YUVARAJ RAJAMANICKAM, Murugappan M, Norlinah Mohamed Ibrahim, Ye Htut, Kenneth Sundaraj, Khairiyah Mohamad, Ramaswamy Palaniappan, Edgar Mesquita, and Marimuthu Satiyan. On the analysis of eeg power, frequency and asymmetry in parkinson's disease during emotion processing. 10:12, 04 2014.

[RN13] A Guruva Reddy and Srilatha Narava. Artifact removal from eeg signals. *International Journal of Computer Applications*, 77(13), 2013.

[Ros61] Frank Rosenblatt. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, CORNELL AERONAUTICAL LAB INC BUFFALO NY, 1961.

[RPLGV15] Rafael Ramirez, Manel Palencia-Lefler, Sergio Giraldo, and Zacharias Vamvakousis. Musical neurofeedback for treating depression in elderly people. *Frontiers in neuroscience*, 9:354, 2015.

[RTC07] N.A. Roberts, J.L. Tsai, and James Coan. Emotion elicitation using dyadic interaction tasks. pages 106–123, 01 2007.

[Rus80] James Russell. A circumplex model of affect. 39:1161–1178, 12 1980.

[RV12] Rafael Ramirez and Zacharias Vamvakousis. Detecting emotion from eeg signals using the emotive epoc device. In Fabio Massimo Zanzotto, Shusaku Tsumoto, Niels Taatgen, and Yiyu Yao, editors, *Brain Informatics*, pages 175–184, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[Sch74] Bruce S. Schoenberg. *Richard Caton and the electrical activity of the brain*. [publisher not identified], [Place of publication not identified], 1974.

[Sch92] Jürgen Schmidhuber. Learning complex, extended sequences using the principle of history compression. *Neural Computation*, 4(2):234–242, 1992.

[Sch05] Klaus R. Scherer. What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729, 2005.

[Sil63] Daniel Silverman. The rationale and history of the 10-20 system of the international federation. *American Journal of EEG Technology*, 3(1):17–22, 1963.

[SJ08] Ryan A Stevenson and Thomas W James. Affective auditory stimuli: Characterization of the international affective digitized sounds (iads) by discrete emotional categories. *Behavior Research Methods*, 40(1):315–321, 2008.

[SLPP12] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, Jan 2012.

[SLSH17] Xiaofeng Sun, Xiangguo Lin, Shuhan Shen, and Zhanyi Hu. High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field. *ISPRS International Journal of Geo-Information*, 6(8), 2017.

[SM+05] Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 1. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[SNSP10] Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippot. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, 24(7):1153–1172, 2010.

[SWA+17] R. Subramanian, J. Wache, M. Abadi, R. Vieriu, S. Winkler, and N. Sebe. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing*, pages 1–1, 2017.

[Tep02] Michal Teplan. Fundamental of eeg measurement. 2, 01 2002.

[VL92] Charles Van Loan. *Computational frameworks for the fast Fourier transform*, volume 10. Siam, 1992.

[VR12] Zacharias Vamvakousis and Rafael Ramirez. A brain-gaze controlled musical interface. In *Berlin BCI Workshop 2012-Advances in Neurotechnology*. Citeseer, 2012.

[Whi09] Cynthia Whissell. Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. *Psychological Reports*, 105(2):509–521, 2009. PMID: 19928612.

[wik18a] Electrodermal activity, Apr 2018.

[wik18b] Sweat gland, May 2018.

[WKA05] Johannes Wagner, Jonghwa Kim, and Elisabeth André. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 940–943. IEEE, 2005.

[WL17] M. B. H. Wiem and Z. Lachiri. Emotion sensing from physiological signals using three defined areas in arousal-valence model. In *2017 International Conference on Control, Automation and Diagnosis (ICCAD)*, pages 219–223, Jan 2017.

[WMBB13] Christine D. Wilson-Mendenhall, Lisa Feldman Barrett, and Lawrence W. Barsalou. Neural evidence that human emotions share core affective properties. *Psychological Science*, 24(6):947–956, 2013. PMID: 23603916.

[XP15] H. Xu and K. N. Plataniotis. Subject independent affective states classification using eeg signals. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1312–1316, Dec 2015.

[Yan16] Shi Yan. Understanding lstm and its diagrams – ml review – medium, Mar 2016.

Appendix

I. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Artem Bachynskyi**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

   1.1 reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

   1.2 make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

   of my thesis

   Emotional State Recognition Based on Physiological Signals

   supervised by Ilya Kuzovkin and Raul Vicente Zafra

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 21.05.2018