Institute of Software Technology

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Bachelorarbeit Nr. 666

From Physiological Signals to Emotions - An Integrative Literature Review

Evelin Grande

Course of Study: Informatik

Examiner: Dr. Daniel Graziotin

Supervisor: Lisa-Marie Michels

Commenced: February 1, 2022

Completed: August 1, 2022

Abstract

With the increase of daily human-computer-interactions and humans being highly emotional beings, came the research field of affective computing. More precisely, the need to incorporate emotions and emotion recognition into computer systems.
Humans often express their emotions to other humans through body language, namely facial expressions, gestures and posture. Those can be deliberately influenced and to avoid this, physiological signals can be used for emotion recognition. They are regulated by the autonomous nervous system and influenced by emotions. This includes the electrical activity of the brain, heart and skin, as well as the skin temperature and respiration patterns, among others.
Building systems that realize autonomous emotion recognition is a regularly studied topic and keeping an overview is difficult.
An emotion recognition system usually consists of data acquisition, data processing and the assignment of emotion classes to the measured signals. Therefore, the foundation of emotion recognition systems includes the creation of data sets, the methods to extract and select features from the collected physiological signals and the classification algorithms.
This thesis offers an explanation of the commonly used physiological signals, data sets and data processing methods, as well as an overview of some frequently used classifiers. However, the main objective is to provide an examination of topics that are currently being worked on. For this purpose, this thesis analyzes and reviews twenty-five studies published in the years 2018 to 2022 to report important findings or limitations and to highlight topics that are worth exploring further.

Kurzfassung

Mit den vermehrten täglichen Interaktionen zwischen Mensch und Computer und dem Umstand, dass der Mensch ein sehr emotionales Wesen ist, entstand das Forschungsgebiet Affective Computing. Genauer gesagt, die Notwendigkeit, Emotionen und Emotionserkennung in Computersysteme einzubeziehen.
Menschen drücken ihre Emotionen gegenüber Anderen oft durch Körpersprache, wie Mimik, Gestik und Körperhaltung, aus. Diese können absichtlich beeinflusst werden und um dies zu vermeiden können physiologische Signale zur Emotionserkennung verwendet werden. Diese werden durch das autonome Nervensystem reguliert und von Emotionen beeinflusst. Dazu gehören unter anderem die elektrische Aktivität des Gehirns, des Herzens und der Haut, sowie die Hauttemperatur und der Rhythmus der Atmung.
Die Erstellung von Systemen zur autonomen Emotionserkennung ist ein regelmäßig untersuchtes Thema und es ist schwer den Überblick zu behalten.
Ein Emotionserkennungssystem besteht in der Regel aus Datenerfassung, Signalverarbeitung und der Zuordnung von Emotionsklassen zu den gemessenen Signalen. Zu den Grundlagen von Emotionserkennungssystemen gehören daher die Erstellung von Datensätzen, die Methoden zur Extraktion und Auswahl von Merkmalen aus den erfassten physiologischen Signalen und die Klassifizierungssalgorithmen.
Diese Thesis bietet eine Erklärung der üblicherweise verwendeten physiologischen Signale, Datensätze und Datenverarbeitungsmethoden, sowie einen Überblick über einige häufig verwendete Klassifizierer.
Das Hauptziel ist jedoch eine Auseinandersetzung mit Themen, an denen derzeit gearbeitet wird. Zu diesem Zweck werden in dieser Thesis fünfundzwanzig Studien, die in den Jahren 2018 bis 2022 veröffentlicht wurden, analysiert und überprüft, um wichtige Erkenntnisse oder Einschränkungen aufzuzeigen und Themen hervorzuheben, die es verdienen, weiter erforscht zu werden.

Contents

List of Figures

List of Tables

List of Abbreviations

AMIGOS *A dataset for Multimodal research of affect, personality traits, and mood in Individuals and GrOupS*. 33

ANN Artificial Neural Network. 46

ANS autonomous nervous system. 28

BLR Boosted Logistic Regression. 46

BVP blood volume pulse. 27

CASE *Continuously Annotated Signals of Emotion*. 48

CNN Convolutional Neural Network. 37

CS Cuckoo Search. 46

CWT Continuous Wavelet Transform. 34

DBN Deep Belief Network. 47

DEAP *Database for Emotion Analysis Using Physiological Signals*. 32

DGCNN Dynmical Graph Convolutional Neural Network. 47

DT Decision Tree. 35

DTD Dynamic Threshold Difference. 48

DWT Discrete Wavelet Transform. 34

ECG electrocardiography. 22

EDA electrodermal activity. 25

EDMIL Deep Multiple Instance Leaning based Emotion Recognition Algorithm. 48

EEG electroencephalography. 22

EMG electromyography. 25

EmoDSN Emotion Recognition based on Deep Siamese Network. 48

EOG electrooculography. 29

FCN Fully Connected Network. 48

FFT Fast Fourier Transform. 34

FrFT Fractal Fast Fourier Transform. 34

GARAFED Graphical Assessment for RealLife Application-Focused Emotional Datasets. 53

GLM-NET Lasso and Elastic-Net Generalized Linear Models. 46

GSR galvanic skin response. 25

GWO Grey Wolf Optimizer. 46

HR heart rate. 27

HRV heart rate variability. 25

IADS International Affective Digitized Sounds. 30

IAPS International Affective Picture System. 30

KNN K-Nearest Neighbor. 36

LDA Linear Discriminant Analysis. 35

LOSOCV Leave-One-Subject-Out cross-validation. 38

LR Linear Regression. 35

LSTM Long Short-Term Memory. 35

MEG magnetoencephalography. 29

MKL Multiple Kernel Learning. 48

MLPNN Multi-Layer Perceptronneural Network. 46

MN Multinomial Regression. 46

MPED *Multi-Modal Physiological Emotion Database*. 48

mRMR Minimum Redundancy Maximum Relevance. 46

NB Naive Bayes. 37

NN Neural Network. 48

ODA online data adaptation. 46

OVPD *Odor–Video Physiological Signal Database*. 33

PCA Principal Component Analysis. 35

PPG photoplethysmography. 27

PSD Power Spectral Density. 47

PSO Particle Swarm Optimization. 46

RDFKM Regularized Deep Fusion Kernel Machine. 48

RF Random Forest. 35

RSP respiration rate. 25

SAM Self Assessment Manikin. 9

SC skin conductance. 28

SEED *SJTU Emotion EEG Dataset*. 32

SFFS Sequence Forward Floating Selection. 48

SKT skin temperature. 25

SRU Simple Recurrent Units for Ensemble Learning. 47

STFT Short-Time Fast Fourier Transform. 34

STRNN Spatail-Temporal Recurrent Neural Network. 47

SVM Support Vector Machine. 36

TBAG Bagging Trees. 46

UDA unsupervised domain adaptation. 46

XGBTREE Extreme Gradient Boost. 46

1 Introduction

Our daily life revolves more and more around online interactions, from social media to remotely working from home. This includes interactions with computer systems, ranging from online shopping websites to digital learning platforms and monitoring systems in healthcare environments.
At first sight, emotions and emotional intelligence are often seen as a completely independent field from such computer systems. Nevertheless, emotions interact with thought processes. Not necessarily in obvious ways, but emotions are important for intelligent functioning, according to Picard [1]. Additionally, human-computer-interaction follows the same principles as human-human-interactions and those rely heavily on the aspect of emotions [2]. Picard also stated that a computer could potentially appear more intelligent by recognizing and appropriately adapting to the user's emotion response [1]. Therefore, it is deemed beneficial to include the aspect of emotions in adaptive computer systems that directly interact with humans.
The field of research that was born from this is called *Affective Computing* [3].

It is crucial to give an adaptive computer system the ability to recognize emotions. This skill forms the basis to a system that influences its behaviour appropriate to a user's emotion.
This leads to the need for automated emotion recognition, which is a subject of interest in many different fields of research. It involves and draws from various research topics. Psychology for the background knowledge about emotions and medicine about how a human body reacts to emotional stimuli. Also included are computer sciences for the automation of the recognition process and mechatronics for the development of measurement devices and sensors.

A large part of human communication takes place through body language, namely facial expressions, gestures and posture. Therefore, people often influence these outward expressions to convey a certain image of themselves in front of others. Be it to be perceived as confident and self-assured to get that next promotion or not showing sadness when watching a movie with friends in order to not be considered weak by cultural standards.
However, when measuring emotions and emotional responses to stimuli, true and honest reactions are needed. Since it is not unlikely that people may withhold their initial facial

expression as a reaction to a new or changed situation, it is more reliable to use other types of emotional responses that are harder to manipulate.
The Autonomous Nervous System regulates various body parameters and is influenced by emotions [4]. Such parameters are the respiration rate of a person, the electrodermal activity, along with the skin temperature and electrocardiograms. Additionally, through electroencephalography brainwaves can also be utilized for the purposes of emotion recognition. All of these modalities differ in their applicability and are useful for different affective states.

Systems for physiological signal based emotion recognition are comprised of multiple steps. Firstly, the physiological signals and the corresponding emotions are measured. In a commonly used laboratory approach, the emotions are deliberately elicited, using specific stimuli. In many studies, this step is already covered by the usage of publicly available data sets. Those measured signals need to be processed before they can be used for the actual recognition. This consists of feature extraction and feature selection. Lastly, those selected features are then fed into a classifier. Such classifiers are mostly machine learning algorithms that are trained to correctly assign an emotion class to the given features.

Computers are still far away from truly perfect emotion recognition, but progress is being made every day, presented in many published studies and research to come.

1.1 Motivation

The field of automated emotion recognition is a vast one, with many different specializations and topics and it has existed for quite a long time now. With many studies being published every month, it is hard to keep up with all the new discoveries.
This thesis aims to provide a comprehensive overview of the tools usually utilized for emotion recognition, as well as a review of recent studies. Since it would be quite impossible to cover all research, this thesis only takes a look at the works in the specific time frame of the last five years. The focus is on emotion recognition systems based on physiological signals, instead of facial expressions or speech recognition.
It would be possible to specialize even further, for example, to limit the research to one modality only, or solely engage in a discussion about sensors or machine learning in respect to emotion recognition. However, many thesis already exist for this purpose. Consequently, this work tries to provide a more broad overview of the topics currently worked on.

1.2 Objectives

Since the main objective of this thesis is to present an overview of the important research topics in the last five years, a sample of research studies in this time frame is collected and their approach to emotion recognition systems is analyzed.
One goal is to explain the procedure of emotion recognition and highlight the typically used models and methods. These include data sets, emotion models and an examination of the various physiological signals the systems are based on. Methods for feature extraction and selection are covered, as well as different machine learning algorithms, which are tasked with classifying the physiological data in order to recognize emotions automatically. Furthermore, the thesis discusses the commonness of the different approaches and emphasize good performing instances.

1.3 Structure

First of all, Chapter 2 provides a summary of related work, i.e. various reviews concerning different topics in the field of emotion recognition.
Chapter 3 takes a look at the foundations of emotion recognition systems, following the thematic structure of the implementation of an emotion recognition system. The groundwork is laid with an explanation of emotion models. This is followed by the many various types of physiological signals and how they can be measured. Afterwards, methods to elicit emotions and the different ways to evaluate what the test subjects felt, are covered. Additionally, commonly used data sets are presented. An insight into feature extraction and selection methods is provided, as well as a selection of often used classification algorithms. Lastly, two used validation methods are going to be covered.
After the foundations are provided, Chapter 4 explains the methodology used to search, select and analyze the papers reviewed in this work. The resulting set of papers is presented in Chapter 5.
Chapter 6 discusses the findings of those papers, as well as similarities, agreements and differences. This Chapter is split into two parts, the first is following the steps of the experiment and data collection procedure. The second part covers the findings regarding the methods in the recognition procedure of the emotion recognition systems.
Lastly, Chapter 7 provides the conclusion and outlook for future work.

2 Related Work

Many reviews that examine studies about emotion recognition systems investigate not only physiological signals as modalities for emotion recognition, but rely heavily on studies working on emotion recognition for facial expressions, speech recognition or body language.
Mohammed and Hassan surveyed and analyzed studies regarding all of those different modalities in the context of human-robot-interaction and social robotics [5]. Another study addressed facial and spoken expressions alongside with physiological signals in their work and put the focus on studies concerning patients in a clinical environment [6]. Egger et al. [4] and Chunawale and Bedekar [7] both provided short overviews of all these modalities as well, alongside a comparison in advantages and limitations.
The subject of e-learning environments is added in a survey by Imani and Montazer [8]. They discuss different emotion recognition methods, including all previously listed modalities and which methods are suitable for e-learning environments.
Meanwhile, other reviews concentrate on solely one type of modality. For example, Ganapathy only looked at emotion recognition using speech [9], as did Koolagudi and Rao [10]. The latter also put an emphasis on exploring various emotional speech databases.
Regarding facial emotion recognition, Ko presents an overview of conventional and deep-learning based approaches [11]. More focus on machine algorithms for this purpose was provided by Koghai et al. [12]. Additionally, De and Saha also examined studies concerning emotion recognition systems based on facial expression and compared a handful of algorithms and their flaws and benefits [13].
The study of Landowska et al.[14] covered a more specialized field. They conducted a systematic literature review concerning automatic emotion recognition specifically for children with autism. Studies working with not only physiological signals, but also facial expressions and speech were included in this review.

Multiple reviews provide surveys about the foundations of emotion recognition systems that are solely based on physiological signals. One review that mainly focused on this was done by Wioleta [15]. Her paper provides a quick overview of signals and sensors, emotion models and emotion elicitation. She also showcases a few studies and highlights the amount of participants in each study, the used signal types and devices, stimuli and recognized emotions, as well as classification methods and their recognition rate.

Ali et al.[16] start their review with emotion theories and a description of the various physiological signal types. This is followed by an outline of the implementation steps, namely emotion elicitation, signal processing and classification. Furthermore, they present a list of eight related studies and emphasize the importance of generalized systems.
As previously mentioned, various reviews cover only one specific physiological signal and analyze studies working purely with this modality. Hasnul et. al documents the important foundations for emotion recognition systems based on electrocardiography (ECG) data and examine these in regards to healthcare systems [17]. A review about electroencephalography (EEG) based emotion recognition and the effect of music on human emotions was conducted by Hamada et al. [18].

A topic with more coverage is wearable devices. Here, many reviews concern themselves with the development of new sensors and measurement devices. For example, Saganowski et al. analyze and thoroughly critique several studies involving wearable devices for signal measuring [19]. They point out that none of the reviewed studies were comprehensive in every research stage. Nonetheless, they deem wearable devices to be a very promising research direction.
Wijasena et al. also reviewed studies that utilized wearable measurement devices [20]. They argued that such devices are an auspicious way to implement emotion recognition based on physiological signals in daily life.
A more specialized study was presented by Tzafilkou et al. [21]. They conducted a systematic literature review and collected studies dealing with the usage of sensors in mobile devices, especially smartphones, in regards to learning related emotions.
While not purely concerned with specifically wearable sensors, Dzedzickis et al. [22] provide a thorough overview of different sensors for physiological signals and their application.

Meanwhile, Calvo and D'Mello examined the topic of emotion recognition from a more theoretical perspective [23]. They describe different emotion theories and their effect on the research questions, methods, results and interpretations. They highlight problematic assumptions and limitations because of this.

A meta-analysis of the field of automated emotion recognition was conducted by Marrero-Fernández et al. in 2014 [24]. They reveal important topics and leading themes, as well as investigate in which countries those research is primarily based. They also note that many reviews about the topic of automated emotion recognition did not follow the standard of systematic literature reviews. This seems to be a persisting issue, as many reviews do not specify their methods thoroughly.

Nevertheless, integrative literature reviews about emotion recognition from physiological signals seem to be very rare. When searching for literature reviews, the majority of results are unspecified or systematic literature reviews.

3 Foundations

This chapter provides an initial overview of the foundations of emotion recognition systems and aims to explain the needed and used concepts and methods.
Firstly, emotions are explained, as well as two different types of emotion models that are commonly utilized by researchers for their measurements. This is followed by an examination of the various physiological signals and their properties and purposes. The elicitation material needed in order to deliberately evoke emotions and the related physical reactions is discussed together with methods to evaluate the measured emotions. Additionally, two commonly used data sets are described. Processing of the acquired data is explained in the sections about feature extraction and selection. Lastly, a few often used classification methods are discussed.

3.1 Emotions

Emotions themselves are a very subjective thing to most people and laymen may not put much thought into trying to define exactly what they are apart from some ”feeling”. But in order to conduct productive research on the topic, scientists had to define emotions. There are different definitions out there and not all scientists agree on one specific definition. The definition this thesis uses is from Keltner and Gross: emotions are ”episodic, relatively short-term, biologically-based patterns of perception, experience, physiology, action, and communication that occur in response to specific physical and social challenges and opportunities” [25].

Emotions must be distinctly and clearly identifiable in order to ensure unambiguous and accurate emotion recognition. Otherwise, emotion recognition systems would be prone to error and their use would not provide the desired and intended benefits.
Therefore, an understanding of the typically used emotion models is needed, which enable us, researchers and test subjects likewise, to distinguish and label the experienced emotions, whether they are elicited in a experiment setting or evoked in day-to-day situations. It is differentiated between a discrete and a continuous emotion model, both of which are often utilized in research and experiments. In this, both models differ in their usefulness and applicability, depending on which methods are used. This is covered

in Section 3.2, when the various physiological signals and what aspect of emotional responses they express are discussed.

3.1.1 Discrete Emotion Model

A discrete emotion model describes emotions as labels distinct from each other. Definitions for basic emotions are still debated under scientists and change with the progress of research.
Paul Ekman describes emotions as families of related states, rather than single affective or psychological states that are shared across cultures and are universally recognized [26], [27]. Ekman states there exists evidence for seven emotions, namely anger, fear, surprise, disgust, contempt, sadness and happiness. Although he expects evidence to be found in the future for more emotions to comply with his proposed characteristics for an emotion to be basic [26].
Not always are all of these basic emotions used for emotion recognition experiments, many utilize only a distinction between negative, positive and neutral emotions, like the commonly used data set SEED does [28].

3.1.2 Continuous Emotion Model

In contrast to the discrete approach, there are also some continuous models that record emotions in a two or three dimensional space. One such two-dimensional model is the Circumplex Model of Affect by James A. Russell, presented in 1980 [29]. In this model, emotions are arranged in the horizontal dimension from displeasure to pleasure and in the vertical dimension from sleep to arousal.
More commonly, the terms positive or negative valence are used instead of high or low pleasure.
Emotions are then distributed according to how they fit into this scheme, e.g. happiness is classified as positive valence and (moderately) high arousal and sadness as low valence and low arousal as seen in Figure 3.1.

Variations of Russell's model are often used in research and many physiological signals relate to either one of those measurements, sometimes even both. Therefore, the data sets DEAP and MAHNOB use this two dimensional emotion model to attribute the measured signals to the emotions perceived by their subjects.

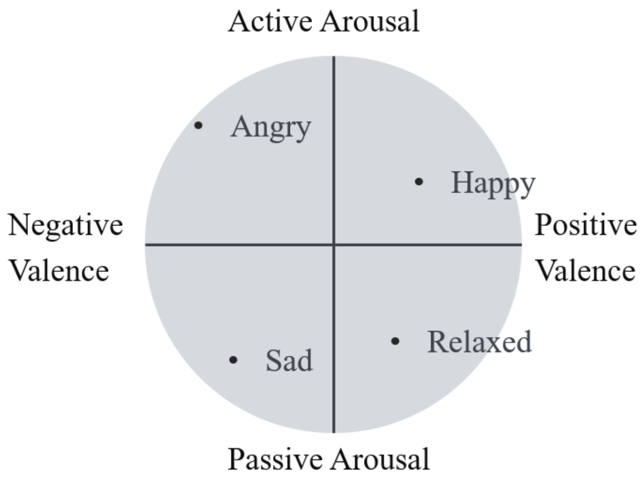

Figure 3.1: Two dimension system based on Russell's Circumplex Model of Affect [29].

3.2 Physiological Signals

Humans display emotions in many different ways. Especially when communicating with other people, emotions are a big part of the interaction and are expressed mainly through body language. Gestures, facial expressions and language - both tone and manner of expression - are the most important ways in which this emotional information is communicated to the other person. Accordingly, these external types of expression are also subject to social and cultural rules and expectations. This quickly leads to people influencing or manipulating their emotional expressions in order to avoid attracting negative attention in their environment.
At the same time, emotions do not only play a role in these communicative aspects, but the physiology of a person is also influenced by emotions. For example, respiration rate and heart rate increase when one is afraid, or slow down when relaxed.
These physiological signals are not as easily deliberately influenced and are therefore much more reliable for emotion recognition.
Such emotionally relevant bio-parameters are electroencephalography (EEG), electrocardiography (ECG)/heart rate variability (HRV), electromyography electromyography (EMG), galvanic skin response (GSR)/electrodermal activity (EDA), skin temperature (SKT) and respiration rate (RSP).
The applicability and usage varies depending on the type of signal.

3.2.1 Electroencephalography EEG

Electroencephalography is a method of measuring the electrical activity in the human brain. It consists of electrodes that are placed around the head according to a specific system. There are different resolutions, with correspondingly more or fewer electrodes. The international standard for many years was the 10/20 system with 21 electrodes [30]. This system has been extended to the 10/10 system with 64 channels arranged as shown in Figure 3.2 [31]. This is used in the SEED and DEAP data sets.

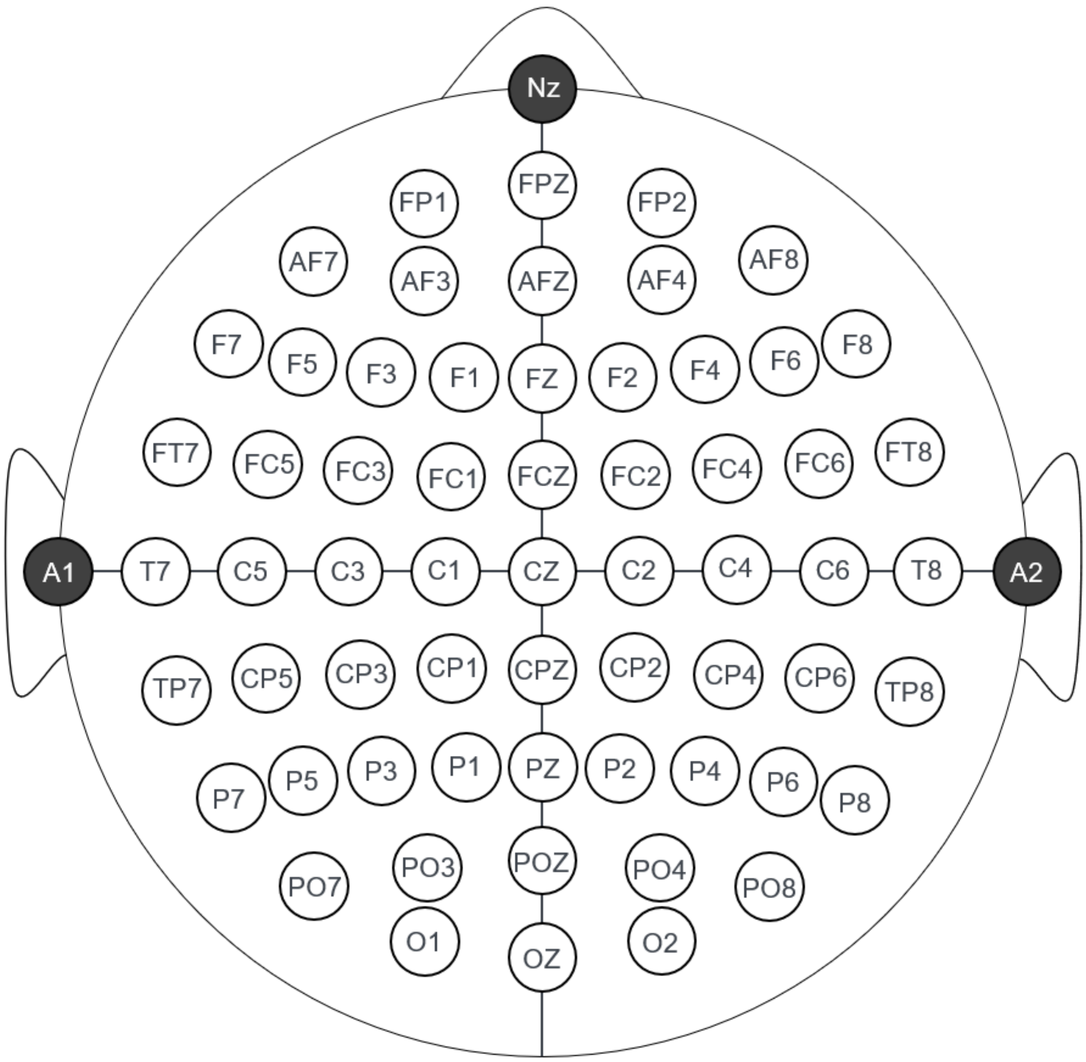

Figure 3.2: Placement of Electrodes in the 10/10 Standard based on [31].

The EEG signal itself is the fluctuation of voltage between a pair of electrodes over time [32]. The measurements are noise-sensitive and best used in a clinical environment.

The signals are divided into five different frequency bands: delta, theta, alpha, beta and gamma.
Delta waves are in the frequency range 0.5 - 4Hz and are usually associated with sleep. Theta waves range from 4-8 Hz and are mostly observed during relaxation and are associated with working memory load. Alpha waves are the basic rhythm of an awake but resting person with closed eyes. They range from 8 to 13 Hz. The frequency range 13 - 30 Hz covers the beta waves. They are observed when the brain is aroused and engaged in an activity. All frequencies above 30 Hz contain the gamma waves and are associated with high-level functions and high concentration. They are completely absent during sleep.
According to Zheng and Lu [28] 12 of these channels are the most effective for emotion recognition purposes. Those channels are C5, C6, CP5, CP6, FT7, FT8, P7, P8, T7, T8, TP7 and TP8.
Additionally, to enable two-dimensional valence and arousal measurements, Ramirez et al. [33] have established two equations with the measurements from four locations in the prefrontal cortex:

$$Arousal = \frac{\beta F3 + \beta F4 + \beta AF3 + \beta AF4}{\alpha F3 + \alpha F4 + \alpha AF3 + \alpha AF4}$$

$$Valence = \alpha F3 - \alpha F4$$

3.2.2 Heart Rate Variability HRV

Alongside the human brain, the heart is another major player in the measurement of emotions. Originally developed for medical purposes, Electrocardiography (ECG) is used to determine the electrical activity of the heart and through this the beat-to-beat temporal variation of the heart rate (HR), called the heart rate variability (HRV), can be calculated.
This is quite complex with conventional procedures and difficult to apply in everyday life. Although there are garments with corresponding sensors, the method of photoplethysmography (PPG) is easier to use, as often seen in smartwatches [4].
PPG utilizes a light source emitting red or near infra-red light and a photodetector to measure changes in the reflection by or transmission trough the skin , from which volumetric variations of blood circulation in the underlying tissue are measured [34]. This is called the blood volume pulse (BVP).
For the transmission technique earlobes and fingertips are most commonly used, while for the reflection technique the measurement device can be attached to wrists, ankles, the forehead or torso. From this monitoring of the BVP the heart rate can be determined

in a non-invasive, mobile and cost effective way. Nonetheless, it must be considered that HRV is dependent on the position of the body during monitoring [4].
The heart rate is a good indicator of arousal. For example, the heart rate increases during stress or excitement, but is lower during a relaxed state [35].

3.2.3 Galvanic Skin Response GSR

The terms electrodermal activity (EDA), galvanic skin response (GSR) or skin conductance (SC) all describe the electrical conductivity of the skin. The GSR is influenced by salts in the sweat. The sweat glands on the palms, fingers and soles of the feet are particularly noteworthy. Arousal is directly linked to the skin's conductivity. With increased arousal, sweat production is stimulated and with more salts in the skin, its conductivity also increases [22]. However, the GSR lacks relation to valence and therefore it is usually used in combination with other physiological signals. The body's sweat reaction is regulated by the autonomous nervous system (ANS). This is advantageous for emotion recognition. Since humans can not actively control the ANS, the GSR can occur as a reaction to stimulation the subject is unaware of and can not deliberately inhibit it [36].
To measure the skin conductivity, two sensors are attached to the skin. Ag/AgCl electrodes are used for these sensors and they are best attached to the palm of the hand [37]. Here, devices are often attached to the fingers and can be incorporated into gloves, demonstrated in the study by Domínguez-Jiménez et al.[38].

3.2.4 Skin Temperature SKT

Just like the sweat production for the GSR, the skin temperature is also controlled by the ANS and is therefore beyond conscious control. Thus, his physiological parameter is also suitable for use in emotion recognition.
The emotional states in relation to stress cause the blood vessels to constrict and this causes the skin temperature to decrease. Similarly, relaxed emotional states lead to dilated blood vessels and therefore a higher skin temperature [39].
However, this process takes time and the rather big latency has to be considered for the utilization. The stimulus for this type of signal is required to persist for a longer duration and has to evoke intense emotions, unlike to stimuli for EEG signals for instance [22]. To measure the skin temperature either contact-less thermal imaging with infrared cameras can be used, or it is measured with sensors directly on the skin. Here, the sensors could be integrated into a wearable device, like a glove or wristband, as demonstrated in [40].

3.2.5 Respiration Rate RSP

The respiratory process is a complex and vital procedure and affects numerous parts of the human body. Respiration depth and velocity are also affected by emotional states. According to Zhang et al. [41] deep and fast breathing shows excitement that is accompanied by happy, angry or fearful emotions. They also state that shallow and fast breathing show tension, while relaxed people often have deep and slow breathing. The work also states that the respiration rate roughly doubles during excitement compared to calm states.
There are numerous different methods to measure changes in the respiration pattern. The most straight forward method would be simply counting breaths over a duration of time, using timers or software applications for assistance in the process of counting.
To avoid this manual labor, automatic methods exist. These include the measurement of the air humidity or the temperature of exhaled breaths, as well as changes in the air pressure. Furthermore, movement and changes in circumference of the torso, as well as respiratory sounds can be measured. ECG and PPG can also be utilized to indirectly measure the respiratory rate [42].
However, breathing is heavily influenced and dependent on the level of body activity and fatigue. It has to be noted that it can be very easily consciously influenced.

In addition, it is possible to measure the electrical muscle activity with electromyography (EMG) and thus to measure higher levels of valence through higher muscle tension. Major limitations of this method are the lack of sensitivity to the emotion intensity and the fact that muscle movement can be easily influenced by external factors and stimuli, e.g. changes in light conditions can cause the subject to squint their eyes and therefore interfere with the measurements.
Another way to measure brain activity is magnetoencephalography (MEG), the measurements of magnetic brain activity produced by the electrical currents in the brain.
It is also possible to measure eye movements with electrooculography (EOG), but this thesis does not discuss those last two topics. Facial expressions, gestures and body posture are also modalities that can be used, but they are not considered physiological signals and therefore are not included in this thesis.

3.3 Elicitation and Evaluation Methods

3.3.1 Elicitation Methods

The previous sections discussed what has to be measured, but in order to do so the desired emotions to measure have to be present. If they are not, they have to be deliberately elicited. Different methods exist for this purpose.
It is possible to use experienced actors who pretend to feel an emotion, but the goal of the usage of physiological signals is to avoid fabricated responses and emotional expressions. Moreover, physiological responses are hard, if not impossible, to willfully influence. Consequently, using actors is not a feasible approach in this case.
Emotions can be subject-elicited or event-elicited [1]. They are subject-elicited when related to memories of the subject, while emotions evoked through stimuli independent from a subject are called event-elicited. The latter is most widely used. For this, external visual, auditory or audiovisual stimuli are presented to the subject and the evoked emotions and physiological reactions are recorded. For purely visual stimuli, pictures are shown to the subjects. The International Affective Picture System (IAPS) [43] states it offers standardized, emotionally-evocative, internationally accessible, color photographs that include contents across a wide range of semantic categories. Meanwhile, the International Affective Digitized Sounds (IADS) database [44] provides acoustic emotional stimuli. Additionally, a very common approach is to utilize videos as audiovisual stimuli. For example the MAHNOB-HCI database [45] used segments from popular movies to elicit emotions in their subjects.
Furthermore, a small study with 10 test participants exists that shows that odors can influence the elicitation of emotional states positively or negatively [46].

3.3.2 Evaluation Methods

With the required emotions elicited, there is still the need to annotate the recorded physiological signals with the emotions the subjects experienced during the stimulus. There are two main approaches to this. Self-reporting and expert-labeling [47].
With self-reporting the subjects themselves report their affective state when exposed to the presented stimulus. This can be done with questionnaires, either in a textual or visual way. The SAM[48] is a pictorial assessment technique that measures pleasure, arousal and dominance over the experienced situation, as seen in Figure 3.3. Most studies concentrate only on the assessment of arousal and valence.
However, a drawback of self-reporting is the possibility that subjects misunderstand the questionnaires and report falsely. Alternatively, they could forget to report some emotions if not reporting immediately when the emotion is felt [47].

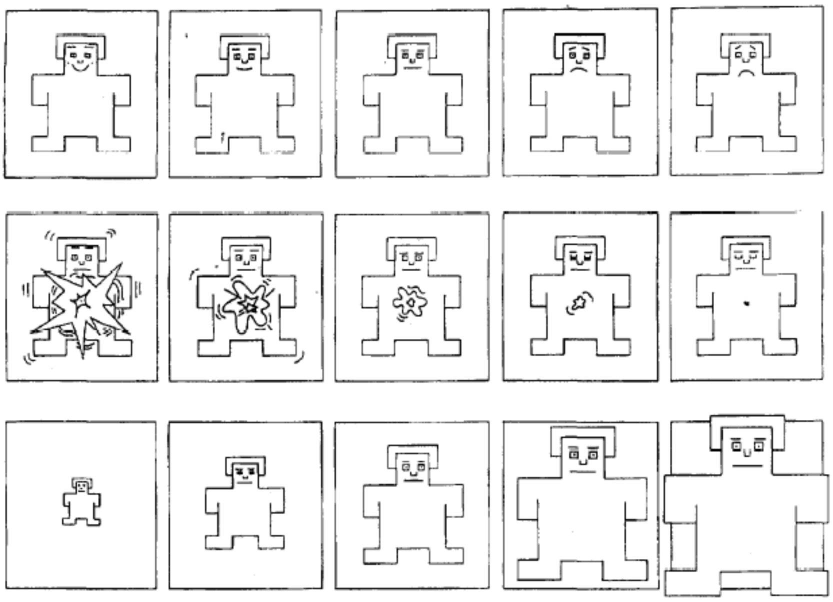

Figure 3.3: The SAM pictures as presented in [48].

It is also possible to involve experts to examine the data to annotate emotions. This can be done either by analyzing the measured physiological signals and label those according to the experts opinion, or by utilizing the facial and bodily expressions of the subjects. This raises the issue that multiple experts can disagree on perceived emotions [47]. Additionally, the issue of emotions being influenced by personality, age and experience, cultural bias, as well as environmental attributes persists with both approaches to evaluating emotions [17], [4].

3.4 Commonly Used Data Sets

It is not always needed to collect a completely new data set for research or the time and resources to design and conduct new experiments with human test subjects. There are several different publicly available data sets that capture different signals and assign them to different emotion models. In the following, a few of the most commonly used data sets are examined in more detail.

3.4.1 SEED

The *SJTU Emotion EEG Dataset* (SEED) contains data from 15 subjects, of which 7 are male and 8 female. Each participant performs the experiment thrice, with an interval of one week or longer. The goal of the experiments were to elicit emotions described as positive, negative or neutral. For each emotion 5 film clips with a duration of 4 minutes each were chosen.
Simultaneously EEG signals and facial expressions of the participants were recorded. The EEG signals were recorded using a 62-channel electrode cap with a sampling rate of 1000 Hz. After each clip the participants filled in a questionnaire to assess the felt emotion and its intensity during the clip. A more detailed description of the data set and its methods can be found in [49].

3.4.2 DEAP

The *Database for Emotion Analysis Using Physiological Signals* (DEAP) consists of the EEG data and peripheral physiological signals of 32 participants, of which 16 are male and 16 female. Additionally, facial expressions of 22 participants were recorded as well.
Each participant was shown 40 music videos with a duration of 60 seconds each. The EEG signals were recorded at a sampling rate of 512 Hz using 32 electrodes. Peripheral signals include GSR, skin temperature, blood volume and respiration pattern.
After each video the participants performed a self-assessment of their arousal, valence, liking, dominance and familiarity of the songs.
A more detailed description of the data set and its methods can be found in [50].

Other popular data sets are *A dataset for Multimodal research of affect, personality traits, and mood in Individuals and GrOupS* (AMIGOS) [51] which includes two experiments with short and long elicitation videos, ASCERTAIN [52], DECAF [53] which includes MEG instead of EEG, DREAMER [54], or MAHNOB-HCI [45]. Table 3.1 shows a compact overview of some key aspects of these data sets, with ”*N*” describing the number of participants and a listing of only physiological modalities.

Name	**N**	**Stimuli**	**Modalities**	**Emotion Model**	**Assessment Type**
AMIGOS	40, 37	16 Videos (51-150 s), 4 Videos (14 - 24 min)	EEG, ECG, GSR	Discrete, Continuous	Self, Expert
ASCERTAIN	58	36 Videos (51-127 s)	EEG, ECG, GSR	Continuous	Self
DEAP	32	40 Videos (60 s)	EEG, ECG, GSR, RSP, SKT, BVP, EMG, EOG	Continuous	Self
DECAF	30	36 Videos (51-128 s)	MEG, ECG, EMG, EOG,	Continuous	Self
DREAMER	23	18 Videos (65 - 393 s)	EEG, ECG	Continuous	Self
MAHNOB-HCI	27	20 Videos (35-117 s)	EEG, ECG, GSR, RSP, SKT	Discrete, Continuous	Self
SEED	15	15 Videos (240 s)	EEG	Discrete	Self

Table 3.1: Overview of Popular Data Sets.

There exist far more data sets, with some being more specialized, e.g. *Odor–Video Physiological Signal Database* (OVPD) [46], which incorporates odor stiumli into the emotion eliciation process, or RCEA-360VR [55], a database employing watching videos in a 360° virtual reality environment as their chosen stimulus. But those are up until now not often utilized by other studies.

3.5 Feature Extraction and Selection Methods

Algorithms process numerical input. Therefore, raw physiological signals have to be be transformed into numerical features, while preserving the information of the original data.

Features can be grouped into multiple different categories. Time-domain features, frequency-domain features, time-frequency-domain features and non-linear features. Time domain refers to the analysis of signals in respect to time, frequency domain states how a signal is distributed over different frequencies and time-frequency domain serves to analyze a signal in both dimensions simultaneously. Nonlinear features characterize chaotic behaviour or regularity in series of data. Prominent examples are Approximate Entropy or Sample Entropy, the Lyapunov Exponent and Hurst Exponent, amongst others.

Features that can be manually calculated, like the mean, median, maximum and minimum, or standard deviation of a signal or function, can be summarized under the term statistical features.

3.5.1 Fourier Transform

The Fourier transform analyzes signals in both time and frequency domain and decomposes a function into sines and cosines of different frequencies and amplitudes. The Fast Fourier Transform (FFT) is an algorithm to calculate the discrete Fourier transform. There exist multiple variations of the Fourier transform. For example, the Short-Time Fast Fourier Transform (STFT) uses a time window that is moved over the signal and the discrete Fourier transform is calculated for this specific frame. Another example is the Fractal Fast Fourier Transform (FrFT). This is a generalization of the ordinary Fourier transform and can be interpreted as the rotation of a signal in the time–frequency plane [56].

3.5.2 Wavelet Transform

The wavelet transform is a time-frequency analysis method as well. It allows to decompose a signal into a series of wavelet coefficients at different scales. This is sometimes favourable, especially for EEG signals, compared to the Fourier transform, since it captures global frequency information and EEG signals consist of short intervals of characteristic oscillation. A distinction is made between Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT).

DWT uses a finite set of wavelets, while CWT considers every possible wavelet over a range of scales. A more detailed explanation of wavelet transform generally, but especially for EEG signals can be found in [57].

After the process of feature extraction it is possible to end up with up to a few hundred different features. Using all of them would result in very long computing times and often not all features necessarily contain information important for emotion recognition. Therefore, it is beneficial to use strategies of feature selection to reduce the amount of features into an efficient set.

3.5.3 Principle Component Analysis PCA

Principal Component Analysis (PCA) is one commonly used technique for feature selection. It is used to reduce the dimensionality of data sets with minimal information loss. New uncorrelated variables are being created, the principal components, which successively maximize variance [58]. Then, PCA projects data onto those lower dimensions and selects which set of dimensions contain the best summary of the data set [59].

Of course there exist many other methods and strategies for feature extraction or selection, but these are a few predominantly used ones.

3.6 Classification Methods

The final step needed to calculate emotions from physiological signals, is the classification. For this purpose, classification models are trained to identify emotions based on the physiological data provided.
A distinction is made between supervised and unsupervised algorithms. Supervised means that the algorithm is trained with pairs of input and output data, while the unsupervised algorithms are trained with unlabeled data.
There are many different algorithms that are suitable for this task, e.g. Linear Discriminant Analysis (LDA), Long Short-Term Memory (LSTM), Linear Regression (LR), Decision Tree (DT) and Random Forest (RF), which is formed by combining decision trees. They all vary in their performance, but only the ones most commonly used in the reviewed papers are discussed below.

3.6.1 Support Vector Machine SVM

A Support Vector Machine (SVM) is a supervised model that uses hyperplanes to separate data into classes. A hyperplanes dimensionality is always lower by 1 than the amount of different features used for the classification. The data points closest to the hyperplane are the support vectors, they control the position of the hyperplane. The objective is to find a hyperplane with the maximum distance between the support vectors of different classes, thus ensuring better generalizability of the classifier.

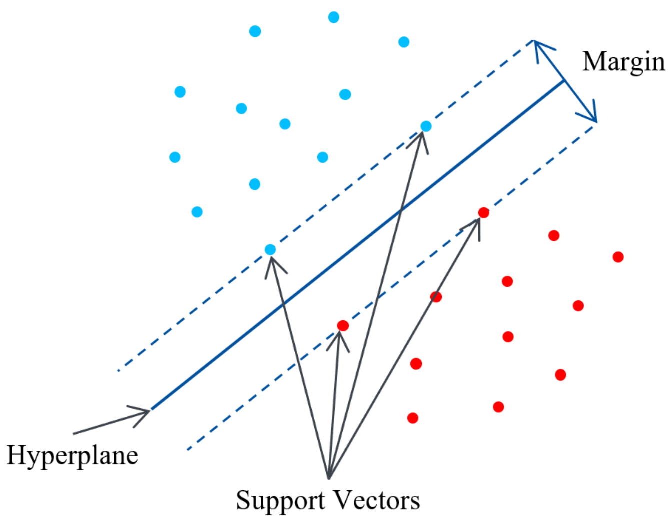

Figure 3.4: Example of a Hyperplane with Maximized Margin.

Figure 3.4 shows an example how such a hyperplane in a two dimensional space would look like without any training errors. It has to be considered that allowing training errors not only reduces the complexity of the resulting function, but also avoids overfitting [60]. This means the classification fits perfectly for training data, but when testing data is added it will not be classified correctly.

3.6.2 K-Nearest Neighbor KNN

The K-Nearest Neighbor (KNN) classifier is supervised as well and uses euclidean distances between test samples and training samples. For a test sample the k nearest training samples are determined. The classifier then sets the predicted class of the testing sample to be most frequent class of those nearest neighbors [61].

3.6.3 Naive Bayes NB

The Naive Bayes (NB) classifier is a probabilistic approach and works particularly well with data sets containing many different features [62]. It is based on Bayes Rule, for which the probability that a tuple of features $X = \{x_1, x_2, ..., x_n\}$ belongs to the class C is $p(C|X) = \frac{p(X|C)p(C)}{p(X)}$. It is assumed that all features contribute independently to the probability of a tuple X belonging to class C. This is usually a false assumption, but works surprisingly well [63].

3.6.4 Convolutional Neural Network CNN

Convolutional Neural Network (CNN) classifiers are prominently used for image recognition and are comprised of interconnected computational nodes, called neurons, that self optimize through learning [64]. They are compromised of five layers, namely the input layer, convolutional layers, pooling layers, fully-connected layers and output layer. The convolutional layer uses filter kernels on the input data to generate feature maps, while the pooling layer serves to reduce the spatial resolution of those feature maps, while maintaining important information throughout. The fully-connected layers perform the classification [65].

3.7 Validation Methods

In order to confirm that the results of the recognition process are correct, validation methods are used. Two very often used strategies are cross validation methods. For this the data is split into training and validation data, whereby both sets must cross over successively so that all data is validated.

3.7.1 k-Fold Cross Validation

One basic form of cross validation is k-fold cross validation. The data is separated into k equally sized sets. For k iterations the model gets $k-1$ alternating sets of training data in order to validate a different set in each iteration. In most reviewed papers k is set to 10.

3.7.2 Leave-One-Subject-Out Cross Validation

Leave-One-Subject-Out cross-validation (LOSOCV) is a special case of k-fold cross validation where k is set as the number of samples in the whole data set. This way all samples expect for one are used for training the classifier and the data of the remaining one test subject is used for validation. This is not feasible when working with very large data sets.

More information on cross validation can be found in [66].

4 Methodology

Torraco stated that the literature is the data of an integrative literature review [67]. And all data has to be collected and processed.
Google Scholar was chosen as the search engine for the whole research. At first keywords for the search were layed out. To get an overview of the topic of emotion recognition using physiological signals, "*emotion recognition*" and "*physiological signals*" were the initial search terms. Furthermore, to set the basic requirements of the search these were always used as the two main keywords.
The goal of this review is to provide an insight of the research work in emotion recognition based on physiological signals in the past five years, with an emphasis on the last two years. It would go beyond the scope of this work to review all research that has ever been there. Consequently, this restriction serves to limit the results to a more manageable amount and to highlight recent findings. Therefore the search results were limited to only entries published in a time frame from 2018 to (June) 2022.
To narrow the results further down and get more specific entries, search terms considering the different signal types were each added to those two main keywords, e.g. ”*EEG*”, ”*GSR*”, ”*ECG*”, ”*HRV*”, ”*respiration*”, ”*skin temperature*” or ”*multimodal*”.
Alternatively, in order to put an emphasis on the aspect of ”*mobile*”, ”*online*” and ”*remote*” options and developments, those keywords were included separately.
In addition, the search term "machine learning" was also used.
For the inital overview of the topic entries with the word ”*review*” or ”*survey*” in their title were preferred. Those provided the basis of a large part of the foundations chapter of this thesis. Table 4.1 shows a list of reviews used for general information and familiarization with the topic.

Authors	Year	Topic
Calvo and D'Mello [23]	2010	Examination of emotion theories from multiple areas, describing how researchers have incorporated psychological theories of emotion and how these theories affect research questions, methods, results, and their interpretations
Egger at al. [4]	2019	Overview of methods to recognize emotions and comparison of their applicability based on existing studies
Imani and Montazer [8]	2019	Review of different emotion theories, discussion of various emotions recognition methods, as well as their advantages and disadvantages for utilizing in e-learning systems
Chunawale and Bedekar [7]	2020	Survey of emotion recognition using physiological signals and their advantages and disadvantages
J. Zhang et al. [68]	2020	Review of emotion recognition methods based on multichannel EEG signals and multi-modal physiological signals, different feature extraction, feature reduction, and ML classifier design methods
Dzedzickis et al. [22]	2020	Analysis of sensors and physiological signals for automated emotion recognition
Hasnul et al. [17]	2021	ECG-based emotion recognition systems: observations of data collection, pre-processing, feature extraction, feature selection and dimensionality reduction, classification, and validation, analysis of the benefit of emotion recognition systems towards healthcare systems
Mejbri et al. [69]	2021	Systematic literature review of 27 papers, regarding the usage and efficiency of affective computing in e-learning

Table 4.1: List of Reviews used for Familiarization.

When it came to the core part of the research non-review studies were preferred. Many reviews either cover a broad subject in a more or less shallow way, or are very specific in one regard, but not all topics are covered equally. Therefore, it was considered a better way to directly identify which topics and approaches are actively pursued currently. Additionally, it allowed to precisely control the currentness of the collected and reviewed studies and experiments.
To select entries matching the desired themes, the titles and the two lines of highlighted content presented by Google Scholar were checked and excluded when the title suggested it was a review or survey. All entries were in english, so there was no need to actively exclude any search results based on the language.

The next step was to read the abstract to decide if the study was clearly associated with emotion recognition and only or at least primarily based physiological signals. If the abstract suggested otherwise or stated that the paper was a review, it was excluded. This way 25 papers were collected. The biggest challenge concerning this process was to find papers that were open access or, if they were locked behind a paywall, accessible through the University of Stuttgart's subscription from that publisher. This was challenging especially with the publisher Elsevier, since many of the papers were not included in the subscription. Nevertheless, six studies published with Elsevier could be retrieved through Sci-Hub. The complete distribution can be found in Table 4.2.

Publisher	Number of Papers
IEEE	11
Elsevier	6
MDPI	2
Springer Nature	6
Total	**25**

Table 4.2: List of Publishers and the Number of Reviewed Papers and Reviews.

Those papers then were analyzed by reading their conclusion and skimming their methodology and results chapters for relevant information. This information was gathered in a table, listing the author, the year the paper was published and under which publisher, as well as the country of the institution the authors are associated with. For information regarding their used data, the data set or respectively the number of participants when the authors gathered their own data was included in the results table, as well as the stimuli used to elicit emotions, the measured signal types and type of emotion model used. Additionally, the used feature extraction and feature selection methods, along with the classifier algorithms and their best achieved accuracy were added, as well as the chosen method of validation. The last column was filled with the particular focus of the study and their to be highlighted findings.
This structure was inspired by the tabular presentations of other reviews, which had a similar structure, e.g. [69],[17], [22]. Furthermore, this structure was conceived as a detailed and informative overview, which makes it possible to precisely list and retrieve individual key information. Similarly, noting the individual focus points and results for each paper helps facilitate the synthesis that forms the core aspect of an integrative literature review.
This table was then reprocessed into the tables containing meta information like publisher, publishing year and associated countries, listed in this chapter, as well as the results table in Chapter 5. Here, the last column containing the focus and findings was omitted, since those entries form the core content of Chapter 6.

Country	Number of Papers
Australia	2
Austria	3
Brazil	1
Chile	1
China	6
Colombia	2
Germany	3
India	4
Iran	3
Ireland	1
Lithuania	1
Malaysia	1
Norway	1
Netherlands	2
Pakistan	1
Portugal	1
Saudi Arabia	1
Taiwan	1
Tunisia	1
Turkey	2
UK	1
USA	2

Table 4.3: List of Countries Associated to Authors.

The documentation of associated countries showed that the topic of emotion recognition using physiological signals is one important to many different countries and cultures all over the world. Table 4.3 shows the various countries associated with the collected studies and reviews. Here, the numbers do not add up to a total of twenty-five studies and eight reviews, since some papers were written by authors associated with differing countries or even multiple countries.

Year	Number of Papers
2018	3
2019	3
2020	10
2021	7
2022	2
Total	**25**

Table 4.4: Overview of the Years the Selected Papers were Published.

The distribution of the years the papers were published is shown in Table 4.4. It should be mentioned that only papers were included that were not only published since 2018, but were also submitted for review at least in 2018.

5 Results

The collected and analyzed papers are presented in Table 5.1, Table 5.2 and Table 5.3. Listed are the author, year of publishing, the used data set and methods, as well as the best achieved classification accuracy and validation methods.
In cases where the paper did not state the used methods or algorithms, the cell sates "N/A". The classifier that achieved the highest accuracy is highlighted in bold letters. All methods are listed with their abbreviations to save space.

A wide variety of publicly available data sets and modalities was used, as well as many different feature extraction methods and classifiers. To get an overview how often those were used, Tables 5.4, 5.5, 5.6 and 5.7 list the differing sets, modalities, feature extraction methods and classifier sorted in descending order by frequency of use.

Table 5.6 and Table 5.7 only list algorithms that were used multiple times to stay more concise.

Regarding the emotion models, continuous models with the dimensions of arousal and valence are employed a total of nineteen times [70], [71], [72], [39], [73] [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [40]. Discrete emotion models are applied ten times [86], [70], [71], [38], [87], [88], [78], [89], [90], [91]. The studies either used distinctly labeled emotions [86], [71] [38], [88], [90] or a system of positive, negative [70], [91] and often neutral labelled emotions as well [87], [78], [89].

Almost half of all studies did not specify a method to validate their results. The two validation methods k-fold cross validation and LOSOCV were used equally often, both eight times respectively.

Author	Data Set	Stimuli	Signaltype	Emotion Model	Feature extraction	Feature Selection	Classification	Accuracy	Validation
Acevedo et al. [86]	own (25 participants)	University Exams	GSR, Electronic Nose to measure gas emitted from skin	Stress	N/A	N/A	LDA, KNN, **SVM**	96% (Electronic Nose) 100% (GSR)	5-fold cross validation
Alakus et al. [70]	GAMEEMO	Video Games	EEG	Valence & Arousal, Positive, Negative	DWT	N/A	SVM, KNN, **Multi-Layer Perceptron-neural Network (MLPNN)**	82% (F4, MLPNN, V/A) 87% (AF4, **MLPNN**, pos/neg)	10-fold cross validation
Ali et al.[71]	MAHNOB-HCI, own (6 participants)	Video Clips	EDA, ECG, SKT	Arousal & Valence 6 universal emotions	Statistical Features	N/A	**Cellular Neural Network**, KNN, NB, SVM, Artificial Neural Network (ANN)	89.38% (same environment) 71.05% (differing environment)	N/A
Awais et al. [72]	CASE	Video Clips	RSP, GSR, ECG, EMG, SKT, BVP	Arousal & Valence	N/A	N/A	**LSTM based**	N/A	N/A
Ayata et al.[39]	DEAP	Music Videos	SKT, RSP, BVP	Arousal & Valence	Statistical Features	Minimum Redundancy Maximum Relevance (mRMR)	**RF**, SVM, LR & subsequent decision level fusion	73.08% (Arousal) 72.18% (Valence)	LOSOCV
Balic et al. [73]	DEAP	Music Videos	EEG	Arousal & Valence	FFT	PCA, mRMR, Particle Swarm Optimization (PSO), Cuckoo Search (CS), Grey Wolf Optimizer (GWO)	**Bidirectional LSTM**	mRMR, 11 hours: 92.74% (Arousal), 92.36% (Valence) CS, 15 hours: 93.33% (Arousal), 93.67% (Valence)	N/A
Doma, Piruz [74]	DEAP	Music Videos	EEG	Arousal & Valence	Statistical Features	PCA	LDA, **SVM, KNN**, NB, LR, DT	70.41% (Valence, KNN) 73.75% (Arousal, SVM)	N/A
Domínguez-Jiménez et al. [38]	own (37 participants)	Video Clips	HR (PPG), GSR	Amusement, Neutral, Sadness	STFT	Random Forest Recursive Feature Elimination, Stepwise Regression, Genetic Algorithms	**SVM**, NB, DT, LDA, SLDA, Multinomial Regression (MN), Extreme Gradient Boost (XGBTREE), Boosted Logistic Regression (BLR), Lasso and Elastic-Net Generalized Linear Models (GLM-NET), Bagging Trees (TBAG)	96% Amusement, 86% Neutral, 91% Sadness	10-fold cross validation
Garg, Verma [75]	DEAP	Music Videos	EEG	Arousal & Valence	CWT	N/A	**CNN (GoogleNet)**	92.19% (Valence) 61.23 % (Arousal)	N/A
Gupta et al. [87]	SEED, DEAP	Video Clips, Music Videos	EEG	Pos/Neutral/Neg, Arousal & Valence	Flexible Analytic Wavelet Transfom & Information Potential Estimator	N/A	**RF**, SVM	90.48% (pos/neu/neg) 71.43% (V/A)	10-fold cross validation
He et al. [76]	DREAMER, AMIGOS	Video Clips	ECG	Arousal & Valence	Statistical Features	covered by domain adaptation	**unsupervised domain adaptation (UDA) with online data adaptation (ODA)**	72% (Valence, DREAMER) 71% (Arousal, DREAMER) 71% (Valence, AMIGOS) 72% (Arousal, AMIGOS)	LOSOCV

Table 5.1: Tabular Overview of all Reviewed Papers.

Author	Data Set	Stimuli	Signaltype	Emotion Model	Feature extraction	Feature Selection	Classification	Accuracy	Validation
Lutze, Waldhör [88]	own (2 participants)	Watching Soccer Games	HR	Stress	N/A	N/A	N/A	N/A	N/A
Panahi et al. [77]	ASCERTAIN	Video Clips	GSR, ECG	Arousal & Valence	FrFT	Wilcoxon signed rank test	**SVM**	78.32% (Valence, ECG) 76.83% (Arousal, ECG)	10-fold cross validation
Pandey, Seeja [78]	SEED, DEAP	Video Clips, Music Videos	EEG	Pos/Neutral/Neg, Valence (& Arousal)	CWT	N/A	**CNN**	61.5% (DEAP, Valence), 58.5% (DEAP, Arousal) 54.0% (cross database, Valence)	N/A
Sepúlveda et al. [79]	AMIGOS	Video Clips	ECG	Arousal & Valence	Statistical Features, Wavelet Scattering Algorithm	PCA	LDA, DT, NB, KNN, SVM, **Ensemble Bagged Tree**	89.3% (Valence) 89.1% (Arousal) 95.3% (two dimensional)	10-fold cross validation, LOSOCV
Sharma et al. [80]	CASE	Video Clips	ECG, BVP, EMG, GSR, RSP, SKT	Arousal & Valence	Statistical Features	N/A	N/A	N/A	N/A
Song et al. [81]	MPED	Video Clips	EEG, RSP, GSR, ECG	Arousal & Valence	Power Spectral Density (PSD), Hjorth, Higher Order Crossings, STFT, Hilbert-Huang Transform, FFT, Attention-LSTM	N/A	SVM, KNN, Deep Belief Network (DBN), Spatail-Temporal Recurrent Neural Network (STRNN), Dynmical Graph Convolutional Neural Network (DGCNN), LSTM, **A-LSTM**	72.93% (distiguish positive/negative)	N/A
Vijayakumar et al. [82]	DEAP	Music Videos	EEG, EMG, GSR, BVP, RSP, SKT	Arousal & Valence	Statistical Features, Sample entropy, Hurst exponent, detrended fluctuation analysis, Lyapunov exponent	PCA	**SVM**, KNN, RF, DT, LR, Gaussian NB (GNB), LDA	64.92% (EOG, SVM, Valence) 63.86% (all without feature extraction, Arousal)	10-fold cross validation
Wei et al. [89]	SEED	Video Clips	EEG	Pos/Neutral/Neg	Dual-Tree Complex Wavelet Transform (DT-CWT), Mean Absolute Value (MAV), PSD, Fractal Dimension (FD), Differential Entropy (DE)	**N/A**	**Simple Recurrent Units for Ensemble Learning (SRU)**, KNN, SVM, NB	80.02% (SRU + DE), 83.13% (SRU ensemble)	N/A
Yang et al. [90]	own(20 participants)	stimuli based on subject personal experiences	EEG, ECG, PPG	Happiness, Anger, Sadness, Neutral	STFT, Statistical Features	N/A	**CNN**	76.94% (EEG) 76.8% (ECG/PPG)	LOSOCV
Yang et al. [91]	own (45 participants)	Pictures	PPG, EDA, SKT & behavioural Signals	Positive/Negative	Attention-based LSTM	N/A	**modality fusion & decision making**	89.24% (Positive) 89.25% (Negative)	LOSOCV

Table 5.2: Tabular Overview of all Reviewed Papers continued.

Author	Data Set	Stimuli	Signaltype	Emotion Model	Feature extraction	Feature Selection	Classification	Accuracy	Validation
Zhang et al. [83]	DEAP, DECAF	Music Videos, Movie Clips	EEG, GSR, EMG, RSP, ECG, MEG	Arousal & Valence	Statistical Features	N/A	SVM, DT, NB, Feature-level fusion: SVM-FLF, AverageMultiple Kernel Learning (MKL), SimpleMKL, EasyMKL, Fully Connected Network (FCN) descision-level fusion: SVM-DLF, **Regularized Deep Fusion Kernel Machine (RDFKM)**	64.5% (Valence, DEAP) 63.1% (Arousal, DEAP) 71.2% (Valence, DECAF) 57.1% (Arousal, DECAF)	LOSOCV
Zhang et al. [85]	CASE, MERCA, CEAP-360VR	Video Clips	BVP & HR, EDA, SKT	Arousal & Valence	1D-CNN, Statistical Features	N/A	**Deep Multiple Instance Leaning based Emotion Recognition Algorithm (EDMIL)**	75.63% (Valence, CASE) 79.73% (Arousal, CASE)	LOSOCV
Zhang et al. [84]	CASE, MERCA, CEAP-360VR	Video Clips	BVP & HR, EDA, SKT	Arousal & Valence	Embedding Network	N/A	**Emotion Recognition based on Deep Siamese Network (EmoDSN)**	76.04% (Valence) 76.62% (Arousal) 57.62% (Arousal,Valence,Neutral)	N/A
Zhao et al. [40]	own (15 participants)	Video Clips	BVP, EDA, SKT	Arousal & Valence	FFT, Dynamic Threshold Difference (DTD), Butterworth Filter	Information Gain (IG), Sequence Forward Floating Selection (SFFS), PCA	**SVM**, RF, NB, Neural Network (NN)	75.65% (SFFS/SVM)	10-fold cross validation, LOSOCV

Table 5.3: Tabular Overview of all Reviewed Papers continued.

Table 5.4 shows the publicly available data sets that were used in the reviewed studies. By far, the data set most often used is DEAP, a data set which measured EEG, GSR, skin temperature, blood volume and respiration pattern, as described in Section 3.4. Seven papers state they conducted their own experiments with participant numbers ranging from 2 to 45 participants [86], [71], [38], [88], [90], [91], [40], and an additional three studies published their data sets, namely GAMEEMO [70], *Continuously Annotated Signals of Emotion* (CASE)[80] and *Multi-Modal Physiological Emotion Database* (MPED) [81], for other researchers to use.

Data Set	Amount	Studies
DEAP	8	[39], [73], [74], [75], [87], [78], [82], [83]
SEED	3	[87], [78], [89]
CASE	3	[72], [84], [85]
AMIGOS	2	[76], [79]
MERCA	2	[84], [85]
CEAP-360VR	2	[84], [85]
DREAMER	1	[76]
MAHNOB-HCI	1	[71]
ASCERTAIN	1	[77]
DECAF	1	[83]

Table 5.4: Overview of How Often each Data Set was Used.

As seen in Table 5.5, the signal types used the most are cardiac signals, i.e. ECG, HRV and BVP combined, with a total of sixteen papers using those. Three of these papers used cardiac signals as their only source of physiological data, while the other thirteen papers used them as part of a multi-modal approach. GSR and EDA combined are used a total of thirteen times, always part of a multi-modal approach. EEG signals are listed eleven times, whereby seven papers based their work only on EEG data. SKT is used nine times, RSP six times, both also only in multi-modal studies. The least often used are EMG signals in four papers.

Modality	Amount	Studies
ECG/HRV/BVP	17	[71], [72], [39], [38], [76], [88], [77], [79], [80], [81], [82], [90], [91], [83], [84], [85], [40]
GSR/EDA	13	[86], [71], [72], [38], [77], [80], [81], [82], [91], [83],[84], [85], [40]
EEG	11	[70], [73], [74], [75], [87], [78], [81], [82],[89], [90], [83]
SKT	9	[71], [72], [39], [80], [82], [91], [84], [85], [40]
RSP	6	[72], [39], [80], [81], [82], [83]
EMG	4	[72], [80], [82], [83]

Table 5.5: Overview of How Often each Modality was Used.

Feature Extraction	Amount	Studies
Statistical Features	11	[71], [39], [74], [76], [79], [80], [82], [90], [83], [84], [85]
FFT	3	[73], [81], [40]
STFT	3	[38], [81], [90]
CWT	2	[75], [78]
PSD	2	[81], [89]
A-LSTM	2	[81], [91]

Table 5.6: Overview of How Often each Feature Extraction Method was Used.

The feature extraction methods that were utilized in the reviewed studies are presented in Table 5.6. The most often mentioned features were not calculated with any special algorithms, but just statistical features listed by the researchers.

Table 5.7 lists SVM as the most used classifier by far, followed by KNN and NB. Thirteen of twenty-five papers applied SVM in their studies and in six it was the best performing classifier. KNN and NB were both listed eight times. Moreover, many more classifiers were applied, but each only once.

Classifier	Amount	Studies
SVM	14	[86], [70], [71], [39], [74], [38], [87], [77], [81], [79], [82], [89], [83], [40]
KNN	8	[86], [70], [71], [74], [79], [81], [82], [89]
NB	8	[71], [74], [38], [79], [82], [89], [83], [40]
LDA	5	[86], [74], [38], [79], [82]
DT	5	[74], [38], [79], [82], [83]
RF	4	[39], [87], [82], [40]
CNN	3	[75], [78], [90]
LSTM	3	[72], [73], [81]
LR	3	[39], [74], [82]

Table 5.7: Overview of How Often each Classifier was Used.

6 Discussion

The papers collected and displayed in Chapter 5 cover many different topics in the field of emotion recognition based on physiological signals. This Chapter takes a look at the relations between these studies and highlight important findings. Firstly, the findings regarding various topics in the experiment structure are discussed, ranging from different data sets and test subject dependencies to signal measurements and wearable measuring devices. Furthermore, the covered recognition methods and performances are examined as well.

6.1 Experiment Procedure

6.1.1 New Data Sets and Sample Sizes

Regularly, new data sets are created. Out of the set of reviewed papers, three aimed to present new unique data sets, namely CASE [80], GAMEEMO [70] and MPED [81]. Sharma et al. [80] demonstrated a novel joystick-based annotation interface in order to enable subjects to simultaneously report their perceived arousal and valence levels. This interface generates two dimensional recordings of emotion measurements alongside with multiple physiological signals. They published their data set for public usage under the name CASE and for example Zhang et al. utilized it in their studies [84] and [85]. Song et al. [81] collected multi-modal physiological data as well, on which they also successfully tested an attention based LSTM classifier.
Alakus et al. [70] presented their data set GAMEEMO containing EEG signals, for which they used video games as the emotion elicitation method. Their subjects played different videos games for five minutes to invoke emotions, namely boredom, relaxation, amusement or fear to be precise.
All three of these data sets did not have more than 30 subjects to gather their data from. This appears to be a pattern that continues through most of the reviewed papers. Most studies that conducted their own experiments had between fifteen [40] and thirty [80] participants. More than thirty subjects appeared only in the studies by Domínguez-Jiménez et al. [38] and Yang et al. [91]. The publicly available data sets with the highest

number of participants is ASCERTAIN [52] with fifty-eight participants. In contrast, two studies even had only two [88] and six [71] participants.

Overall, those participants were mostly healthy and young adults and of higher education levels. The only reviewed study that included people with health-risks was [90], which specifically concentrated on subjects with a high risk for heart diseases.
Pandey and Seeja highlighted that the larger of two used data sets provided better generalization [78].
It follows that such rather small and homogeneous sample sizes limit the generalizability of the commonly achieved results. It is therefore important to further expand the heterogeneity of data samples and to increase the size of data sets. Zhang et al. tried to actively combat this issue of small sample sizes in their study by providing an algorithm that can still achieve reasonable results on very small amounts of training data, i.e. data that contained less than ten samples [84].
However, this does not provide benefits towards the general generalizability.

6.1.2 Elicitation and Evaluation

When trying to elicit emotions for measuring, several approaches exist. The only study that used solely visual elicitation materials is [91]. All other studies utilized audiovisual material, being clips either from music videos or movies, or in the case of [70], video games. Yang et al. did not specify what kind of stimuli they used, only that it was based on the subjects personal experiences [90].

For the annotation of emotions during the experiments, self assessment is the preferred method. Out of ten studies that conducted their own experiments ([86], [70], [71], [38], [88], [80], [81] [90], [91], [40]) seven ([70], [71], [38], [80], [81], [91], [40]) reported using self-reporting assessment techniques to evaluate the emotions experienced by their participants. The other three studies ([86], [88], [90]) did not mention any assessment at all.
While the usage of SAM questionnaires is widely used, the only study that included the dominance ratings in their classification process was [74]. Additionally, the only used data set that utilized not only self assessments, but also external annotation is AMIGOS [51].

6.1.3 Subject Dependencies

Another important factor regarding test subjects and the measured signals is the inter- and intra-subject discrepancy. This means there exist differences not only between the

data gathered from different humans. But day-to-day differences in the measured signals of one single subject or over multiple hours in one day can be a major issue as well.
He at al. referred to this in their study and proposed a domain adaptation strategy to overcome this issue [76].
Meanwhile, Wei et al. [89] pointed out that neural patterns were relatively stable within and across days based on the performance of their proposed system. Inter-subject discrepancies are also considered in [78], where Pandey and Seeja tried to counter this problem by testing their approach in a cross database way.
Gupta et al. [87] addressed this topic by looking into EEG channel specifics and which individual channels provided the best results, with the goal of overcoming inter-subject difficulties.
A unique approach was presented by Ali et al. [71]. They created a calibration model, which is used to determine if a new test subject had similar responses to a subject from the training set. The calibration model identifies the most correlated training subject. Then the features of the test subject are transformed in such a way that afterwards they are calibrated from the perspective of the training set. Ali et al. stated that this improved the overall accuracy performance by more than 13%.

Vijayakumar et al. [82] noted in their work that male subjects achieved a significantly greater average valence recognition accuracy than female subjects. However, this was the only entry to contain gender-specific information.

6.1.4 Standardization

Additionally, regarding the experiment procedure, Lutze and Waldöhr [88] used the Graphical Assessment for RealLife Application-Focused Emotional Datasets (GARAFED) [47], a tool for standardization of the evaluation of emotion recognition experiments. This is done by looking into the dimensions of emotion origin, invasiveness, privacy, experimental days, experimental hours per day and number of subjects in an experiment. No other reviewed study applied this system for their own works, but it would be a helpful tool to more easily compare experiment approaches with each other.

6.1.5 Laboratory Settings and Wearable Measuring Devices

The measuring of physiological signals happens mostly in laboratory settings, as is the case for almost all studies reviewed in this thesis. Exceptions are the studies [84] and [85] by Zhang et al. and one study by Acevedo et al. [86]. Zhang et al. included a data set [92] that gathered data from people watching videos on smartphones while outdoors in both studies ([84] and [85]). Meanwhile, Acevedo et al. [86] conducted their

experiments in the homes of the participants, due to restrictions during the COVID-19 pandemic.
A large number of the reviewed papers were published after the start of the pandemic in 2020. One could assume more studies would be influenced and relate to the changing world environment, but only [86] and [72] acknowledged those circumstances.
This also brought forth an increased need for remotely functioning systems and mobile measuring devices. These are important research topics for the future and for comfortable day-to-day application as well. In this context, Awais et al.[72] proposed an Internet of Things based solution that transfers the measured data to cloud servers for processing.

Many of the collected studies use portable or wearable measuring devices instead of typical stationary clinical measuring equipment ([70], [88], [90], [91], [85], [40] [71] and [38]).
Sepúlveda et al. [79] reported plans for a shirt equipped with ECG electrodes and Domínguez-Jiménez et al. [38] created a wearable glove to measure PPG and GSR signals.
Lutze and Waldöhr [88] used a mainstream smartwatch to measure the heart rate of the test subjects, and concluded that only strong tension situations lead to unambiguous results. The studies [71], [91], [85] and [40] all utilized the Empatica E4 wristband with PPG and EDA sensors, as well as sensors for measuring skin temperature.
Zhang et al. [85] also stated that mobile sensors were showing promising results, although sensors of laboratory quality still achieve higher accuracy ratings. This does not yet include the difficulties that arise for everyday and mobile measurements. For example, the influences that movement has on the measurements of respiration and heart rate.

6.1.6 Uni-Modal vs. Multi-Modal

One way in which the collected studies differ is whether a uni- or multi-modal approach is used. Using multiple modalities offers a high variety of signals which can complement each other, while for example EEG signals are quite accurate on their own. The studies [70], [73], [74], [75], [87], [78], [89] all use solely EEG data, while [76], [88] and [79] used exclusively cardiac signals.
All other fifteen studies employed multi-modal approaches.
The studies using only ECG or EEG data showed a good performance over all, but not necessarily better or worse than the multi-modal approaches. EEG signals may be highly accurate on their own, but they are also more difficult to measure. Therefore, it is understandable that many studies chose to refrain from using EEG data at all and instead opted for a multi-modal approach using other physiological signals. One can

imagine the utilization of more diverse signal types together can improve the accuracy further. However, it has to be taken in account that the usage of differing types of signals also means more effort is needed to produce a coherent result.

6.1.7 EEG Channel Selection

As previously mentioned, Gupta et al. [87] examined the performance of different EEG channels. They reported that the channels FT7, FT8, T7, T8, C5 and TP7 performed best. These results coincide with a study by Zheng and Lu [28].
Contrary to that, Alakus et al. [70] and Pandey and Seeja [78] reported that frontal zones achieved the best results in their study. To be precise Alakus et al. [70] identified AF4 for high arousal and valence, F4 for low arousal and high valence, FC5 for high arousal and low valence and AF3 for low arousal and valence.
Wei et al. [89] noted that higher frequency bands were more favourable for emotion recognition. Regarding channel selection, Balic et al. [73] compared different algorithms for this purpose. They did not specify which channels performed best, instead they reported that they were able to significantly reduce computing time (from eighteen hours without channel selection down to eleven hours with channel selection) while only negligibly decreasing the accuracy. The used methods can be found in the results tables in Chapter 5.

6.2 Recognition Procedure

6.2.1 Feature Extraction Methods

The different signals in raw form are not really applicable for classification purposes, so features that provide more distinct information are needed to be calculated from those raw signals accordingly. Acevedo et al. [86] and Awais et al. [72] both did not mention any used features or extraction methods at all. Meanwhile, [83], [80], [76], [74], [39], [71] only specified the usage of statistical features, but not if those features were extracted with any algorithms.
When mentioned, various forms of the Wavelet Transform or Fourier Transform were the most commonly applied methods. Garg and Verma [75] and Pandey and Seeja [78] both created scalograms, a visual representation of a wavelet transfrom in form of an image, to feed into their CNN classifiers. Therefore, the implementation of CNNs is evident, since they are primarily used for pattern recognition in images [64].
Zhang et al. [85] stated that "feature extraction using an end-to-end structure can

improve recognition accuracy compared with manual feature extraction as well as unsupervised learning feature extraction methods".

6.2.2 Feature Selection Methods

Since physiological signals can provide a great number of different features, it is beneficial to employ strategies to reduce the amount of actually used features for minimizing computing time and effort. Only nine papers ([39], [73] [74], [38], [76], [77], [79],[82], [40]) specified the use of any type of such feature selection methods.
Since there exist multiple different approaches for feature selection, [73], [38], [40] each compared a different set of methods. Balic et al. [73] achieved the best results with mRMR and CS, while Domínguez-Jiménez et al. [38] reported good results with Random Forest Recursive Feature Elimination and Genetic Algorithms, but Stepwise Regression with Forward Selection apparently did not perform well. Furthermore, Domínguez-Jiménez et al. noted that PPG features were not relevant for the recognition of amusement or sadness in their study. Zhao et al. [40] noted their best accuracy was achieved with SFFS, when combined with the SVM classifier.
The most often used feature selection method was PCA, mentioned by five studies ([73], [74], [79], [82], [40]).

6.2.3 Fusion Methods

When working with bigger sets of data it can also be beneficial to combine this information via fusion methods. Six studies ([39], [81], [82], [91], [83], [84]) employed modality level or decision level fusion methods. The five studies [81], [82], [91], [83] and [84] utilized modality level fusion to combine feature sets for further processing in classifiers. All of those studies reported improvements in accuracy compared to approaches without modality fusion. Two studies, [39] and [91], used decision level fusion, to combine results provided by their classifiers for a more reliable outcome.

6.2.4 Comparison of Classifiers

Way more focus was put on the comparison of classifiers, with thirteen studies applying multiple classifiers and comparing them in terms of accuracy [86], [70], [71], [39], [74], [38], [87], [79], [81], [82], [89], [83], [40]. Alakus et al. stated that the KNN classifier did not perform well on their denoised data set due to the fact that it is a algorithm well suited for data with artifacts [70].

However, there is no general consensus on which classifier is definitely superior to the others. The closest would probably be SVM, as five studies identified it as the best performing classifier when compared with others ([86], [74], [38], [82], [40]). It was also overall the most used classifier when including the studies that only used one.
Other classifiers that were considered the best performing out of a set of different algorithms were RF [39] and [87], Cellular Neural Networks [71], KNN [74], MLPNN [70], Ensemble Bagged Tree [79], Attention-LSTM [81] and SRU [89].
Hereby, Ali et al. reported that emotion recognition is a "highly nonlinear dynamical system" and "the history of inputs might affect the outputs". They achieved reasonable results with their proposed classification model based on Cellular Neural Networks and concluded that it performed better because of the utilization of a memory to consider the history of inputs [71].

Ten studies employed only one classifier [72], [73], [75], [76], [77], [78], [90], [91], [84], [85]. Additionally, four studies in total proposed new classifiers [76] [83],[84], [85]. He et al. [76] combined Unsupervised Domain Adaptation UDA from [93] with their own online data adaptation approach. Zhang et al. [83] proposed a classifier called RDFKM, which they compared to other classifiers. It performed best, but also was the only classifier out of ten that was receiving all modalities available in the used data sets. Zhang et al. proposed two different new methods, EDMIL and EmoDSN, where EDMIL combines preprocessing, feature extraction and classification [85], and EmoDSN "learns the difference between samples instead of building the precise mapping between samples and emotion labels" [84].

6.2.5 System Performance

The performance of emotion recognition systems is the key point of their success. Almost all studies measured the performance of their presented system by its accuracy. Balic et al. were the only ones to also include the computing time in their performance assessment [73].
The only exception to the performance measuring by accuracy were Awais et al. [72]. They argued that the f-score provides better insight across unbalanced data sets and did not offer the accuracy ratings their system achieved.

None of the studies arguing with their achieved accuracy ratings reported an value under 50%. Four studies merely achieved accuracy results under 65% ([78], [82], [83], [84]). Meanwhile, six studies ([86], [73], [38], , [75], [87], [79]) reached results higher than 90%: [86] and [73] for all their results and [87] only for discrete emotions, [38] for amusement and sadness recognition, and [79] for a two dimensional approach. However, the ratings for arousal and valence on their own were still very close to 90% accuracy.

The study by Garg and Verma exceeded this threshold only for valence recognition [75]. Zhang et al. [83] also declared significantly higher accuracy values for valence than arousal. Contrary to that, all other studies working with the continuous emotion model achieved very similar results for both dimensions [39], [73], [74], [76], [77], [78], [79], [85], [84].

Since some papers utilized two different emotion models, comparisons can be made in this regard. Two studies reported performance for both continuous and discrete emotion models and reported that their systems performed better for discrete models than continuous.
In the case of [87] this might be because the data set SEED provided more EEG channels than DEAP and therefore the algorithms had more information for the discrete model to work with. Alakus et al. [70] conducted their study only on one data set and produced the same result. This leads to the conclusion that it is also likely that it is easier to recognize less dimensions of emotions, i.e. one dimension from positive to negative compared to two dimensions of arousal and valence or multiple distinct emotion labels.

6.2.6 Validation

The data sets the reviewed studies are working with are rather small. And because the validation methods k-fold cross validation and LOSOCV are easy to apply when working with smaller bodies of data, they are often utilized.
As stated in Chapter 5 the validation methods k-fold cross validation ([86], [70], [38], [87], [77], [79], [82], [40]) and LOSOCV ([39], [76], [79], [90], [91], [83], [85], [40]) were used equally often, eight times to be precise. Sepúlveda et al. [79] and Zhao et al. [40] employed both methods. However, this also means that multiple studies did not report any used validation method at all ([71], [72], [73], [74], [75], [78], [81], [89], [84]). Sharma et al. [80] created a new data set without classifying their measurements and Lutze and Waldöhr [88] also did not apply any classification to their data. Therefore they did not need to validate the results of any classifiers. Nonetheless, this lack of validation should not be the case in such a large proportion.

7 Conclusion

7.1 Summary

Emotion recognition is a very broad field of research, even if the focus limited to studies that concentrate only on systems working with only or mostly physiological signals. Many studies exist covering different approaches to recurring subjects. Twenty-five of such studies published in the years 2018 to 2022 were gathered and reviewed in this work.

The main findings include that many different data sets exist and regularly new approaches and methods in this context are tested. Particularly worthy of mention is the data set CASE and the joystick-based annotation interface to simplify simultaneously annotating emotions, while experiencing stimuli. The most commonly used types of stimuli were videos and self assessment, prominently in the form of SAM, was always the chosen approach for emotion annotation.
Differences between subjects and for the same subject at different times are often investigated issues, as are the channel specific performances for EEG signals. Studies in laboratory environments seem to be the norm, but in eight studies wearable measurement devices are employed instead of medical grade sensors. Ten out of twenty-five studies worked with a single physiological modality, while the rest used multi-modal approaches.
Feature extraction and feature selection are both topics that do not get much in-depth attention. Feature fusion was mentioned a couple of times and was perceived as beneficial. Contrary to that, comparison of classifiers and the development of new techniques is very actively worked on. SVM is a regularly good performing classification method, but other novel approaches showed promising results as well.
To evaluate the success of their system, almost all researchers presented the accuracy values their systems achieved and validated their results with either k-fold or leave-one-subject-out cross validation, both of which work very well on little data.

7.2 Limitations and Future Work

As previously mentioned, a prominent limitation is the lack of large sets of data. The increase of sample sizes would offer more training data, which is beneficial for the learning process of classifiers and lead to better generalizability of the results. This can also be improved by purposefully increasing the diversity of the test subjects, by including more older people, people with illnesses or disabilities and of different ethnicity and cultural backgrounds.
It should be a goal for the future to contribute a data set with at least more than 100 participants. This would also help with issues like the inter- and intra-subject discrepancies, due to a opportunity to investigate those. Maybe patterns can be found in a large set of different subjects, or it would simply be beneficial for the training of classifiers in consideration of this issue.
Such bigger training data sets may also entail longer and more complex training processes. The computing duration for classifiers was mentioned in only one paper [73], therefore no further conclusions can be drawn in this regard from the reviewed papers. But this can not be a hindrance, even if it means more effort.
This leads to the next point, the lack of standardization. The study design can be evaluated with GARAFED and such standardization would enable other researchers to more easily compare data sets and their creation. This way they can make informed decisions more quickly. It would also be beneficial if all studies published their collected data along with their documented results for replicability and transparency, as well as improve documentation overall, since quite a few papers were excluded in the collection process of this work due to missing methodologies or descriptions of their process.

The daily applicability of emotion recognition is one prominent issue that requires further research. Emotions specifically elicited in a laboratory setting are very different from the complex, most often simultaneously occurring multitude of emotions in daily life. Those laboratory emotions may be easier to work with, but the more complex real life emotions ultimately have to be the final goal of research.
Another persisting problem is the main difficulty to measure physiological signals in a non-clinical environment. Movement has a great influence on many physiological signals, e.g. heart rate or respiration. Therefore, a system trained in a clinical environment can not just be transferred to a real life application, since the classifier would misinterpret the changed signals wrongly.
Also, wearable or mobile sensors are not as precise as sensors of laboratory quality, but the results are promising [83]. Consequently, the development of non-intrusive precise wearable sensors is an important objective, from which not only the field of emotion recognition profits, as well as field studies investigating and considering real life situations for their measurements.

Additionally, EEG signals are difficult to measure in daily life. Multi-modal approaches that omit EEG modalities and instead concentrate on other, more easily measurable, signal types seem to have a more promising future for application in the wild, so to speak.
Such multi-modal approaches require more effort to process and for example synchronisation of those multiple signals is very important. But, as is the need for larger data sets, this is an obstacle that has to be overcome in the future.

Features are a great tool to condense the raw signals into distinct important information points. Not all researchers mention using features for their systems and if they do, the extraction of these is not always thoroughly documented. Sensible selection of features is also a point that provides advantages. When only features that definitely provide helpful information are fed to the classifier, it can increase the accuracy of the results, as seen with feature fusion. Using less but more informative features also reduces the computing time, a metric that is completely neglected in most studies. A faster training process has to be weighted against the accuracy of a system, but is an important factor that future commercial applications would consider.

In Chapter 6 it was mentioned that there is no general consensus on a ultimately superior classifier. But, the comparison of classifiers or whole systems is made difficult by the fact that the various studies utilized different data sets. Those use different stimuli, employ different measuring devices and use different emotion models for their annotation process.
Often when employing discrete emotion models, only positive and negative emotions are distinguished instead of the six basic emotions by Ekman [27]. Since real life emotions are more complex than just positive and negative it is necessary to use the more complex approaches and put in the effort to meet the demands.
Additionally, different features and their complexity influence the selection and performance of classifiers as well. All these factors play into great differences between the various emotion recognition systems and complicate attempts to compare them with each other.
Additionally, accuracy ratings are not as helpful in determining the performance of a system [72] when using unbalanced data sets so, this has to be considered in future research. The validation of the results also leaves a lot to be desired when nine papers are not reporting any validation methods.

To conclude the most important points, it is important to improve the overall documentation of studies and to provide bigger and more heterogeneous sets of data. Wearable sensors seem promising and real life environments are necessary to investigate further.

Bibliography

[1] R. Picard, E. Vyzas, J. Healey. "Toward machine emotional intelligence: analysis of affective physiological state." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.10 (2001), pp. 1175–1191. DOI: 10.1109/34.954607. URL: https://doi.org/10.1109%2F34.954607 (cit. on pp. 17, 30).

[2] B. Reeves, C. Nass. *The media equation*. en. Stanford, CA: Centre for the Study of Language & Information, May 1998 (cit. on p. 17).

[3] R. W. Picard. *Affective Computing*. The MIT Press. London, England: MIT Press, July 2000 (cit. on p. 17).

[4] M. Egger, M. Ley, S. Hanke. "Emotion Recognition from Physiological Signal Analysis: A Review." In: *Electronic Notes in Theoretical Computer Science* 343 (May 2019), pp. 35–55. DOI: 10.1016/j.entcs.2019.04.009. URL: https://doi.org/10.1016%2Fj.entcs.2019.04.009 (cit. on pp. 18, 21, 27, 28, 31, 40).

[5] "A Survey on Emotion Recognition for Human Robot Interaction." In: *Journal of Computing and Information Technology* 28.2 (June 2021), pp. 125–146. DOI: 10.20532/cit.2020.1004841. URL: https://doi.org/10.20532%2Fcit.2020.1004841 (cit. on p. 21).

[6] L. Pepa, L. Spalazzi, M. Capecci, M. G. Ceravolo. "Automatic emotion recognition in clinical scenario: a systematic review of methods." In: *IEEE Transactions on Affective Computing* (2021), pp. 1–1. DOI: 10.1109/taffc.2021.3128787. URL: https://doi.org/10.1109%2Ftaffc.2021.3128787 (cit. on p. 21).

[7] A. Chunawale, M. Bedekar. "Human Emotion Recognition using Physiological Signals: A Survey." In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3645402. URL: https://doi.org/10.2139%2Fssrn.3645402 (cit. on pp. 21, 40).

[8] M. Imani, G. A. Montazer. "A survey of emotion recognition methods with emphasis on E-Learning environments." In: *Journal of Network and Computer Applications* 147 (Dec. 2019), p. 102423. DOI: 10.1016/j.jnca.2019.102423. URL: https://doi.org/10.1016%2Fj.jnca.2019.102423 (cit. on pp. 21, 40).

[9] A. G. and. "Speech Emotion Recognition Using Deep Learning Techniques." In: *ABC Journal of Advanced Research* 5.2 (2016), pp. 113–122. DOI: 10.18034/abcjar.v5i2.550. URL: https://doi.org/10.18034%2Fabcjar.v5i2.550 (cit. on p. 21).

[10] S. G. Koolagudi, K. S. Rao. "Emotion recognition from speech: a review." In: *International Journal of Speech Technology* 15.2 (Jan. 2012), pp. 99–117. DOI: 10.1007/s10772-011-9125-1. URL: https://doi.org/10.1007%2Fs10772-011-9125-1 (cit. on p. 21).

[11] B. Ko. "A Brief Review of Facial Emotion Recognition Based on Visual Information." In: *Sensors* 18.2 (Jan. 2018), p. 401. DOI: 10.3390/s18020401. URL: https://doi.org/10.3390%2Fs18020401 (cit. on p. 21).

[12] E. Kodhai, A. Pooveswari, P. Sharmila, N. Ramiya. "Literature Review on Emotion Recognition System." In: *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*. IEEE, July 2020. DOI: 10.1109/icscan49426.2020.9262389. URL: https://doi.org/10.1109%2Ficscan49426.2020.9262389 (cit. on p. 21).

[13] A. De, A. Saha. "A comparative study on different approaches of real time human emotion recognition based on facial expression detection." In: *2015 International Conference on Advances in Computer Engineering and Applications*. IEEE, Mar. 2015. DOI: 10.1109/icacea.2015.7164792. URL: https://doi.org/10.1109%2Ficacea.2015.7164792 (cit. on p. 21).

[14] A. Landowska, A. Karpus, T. Zawadzka, B. Robins, D. Erol Barkana, H. Kose, T. Zorcec, N. Cummins. "Automatic Emotion Recognition in Children with Autism: A Systematic Literature Review." In: *Sensors* 22.4 (2022). ISSN: 1424-8220. DOI: 10.3390/s22041649. URL: https://www.mdpi.com/1424-8220/22/4/1649 (cit. on p. 21).

[15] S. Wioleta. "Using physiological signals for emotion recognition." In: *2013 6th International Conference on Human System Interactions (HSI)*. IEEE, June 2013. DOI: 10.1109/hsi.2013.6577880. URL: https://doi.org/10.1109%2Fhsi.2013.6577880 (cit. on p. 21).

[16] M. Ali, A. H. Mosa, F. A. Machot, K. Kyamakya. "Emotion Recognition Involving Physiological and Speech Signals: A Comprehensive Review." In: *Recent Advances in Nonlinear Dynamics and Synchronization: With Selected Applications in Electrical Engineering, Neurocomputing, and Transportation*. Ed. by K. Kyamakya, W. Mathis, R. Stoop, J. C. Chedjou, Z. Li. Cham: Springer International Publishing, 2018, pp. 287–302. ISBN: 978-3-319-58996-1. DOI: 10.1007/978-3-319-58996-1_13. URL: https://doi.org/10.1007/978-3-319-58996-1_13 (cit. on p. 22).

[17] M. A. Hasnul, N. A. A. Aziz, S. Alelyani, M. Mohana, A. A. Aziz. "Electrocardiogram-Based Emotion Recognition Systems and Their Applications in Healthcare—A Review." In: *Sensors* 21.15 (July 2021), p. 5015. DOI: 10.3390/s21155015. URL: https://doi.org/10.3390%2Fs21155015 (cit. on pp. 22, 31, 40, 41).

[18] M. Hamada, B. B. Zaidan, A. A. Zaidan. "A Systematic Review for Human EEG Brain Signals Based Emotion Classification, Feature Extraction, Brain Condition, Group Comparison." In: *Journal of Medical Systems* 42.9 (July 2018). DOI: 10.1007/s10916-018-1020-8. URL: https://doi.org/10.1007%2Fs10916-018-1020-8 (cit. on p. 22).

[19] S. Saganowski, B. Perz, A. Polak, P. Kazienko. "Emotion Recognition for Everyday Life Using Physiological Signals from Wearables: A Systematic Literature Review." In: *IEEE Transactions on Affective Computing* (2022), pp. 1–1. DOI: 10.1109/taffc.2022.3176135. URL: https://doi.org/10.1109%2Ftaffc.2022.3176135 (cit. on p. 22).

[20] H. Z. Wijasena, R. Ferdiana, S. Wibirama. "A Survey of Emotion Recognition using Physiological Signal in Wearable Devices." In: *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*. 2021, pp. 1–6. DOI: 10.1109/AIMS52415.2021.9466092 (cit. on p. 22).

[21] K. Tzafilkou, A. A. Economides, N. Protogeros. "Mobile Sensing for Emotion Recognition in Smartphones: A Literature Review on Non-Intrusive Methodologies." In: *International Journal of Human–Computer Interaction* 38.11 (Sept. 2021), pp. 1037–1051. DOI: 10.1080/10447318.2021.1979290. URL: https://doi.org/10.1080%2F10447318.2021.1979290 (cit. on p. 22).

[22] A. Dzedzickis, A. Kaklauskas, V. Bucinskas. "Human Emotion Recognition: Review of Sensors and Methods." In: *Sensors* 20.3 (Jan. 2020), p. 592. DOI: 10.3390/s20030592. URL: https://doi.org/10.3390%2Fs20030592 (cit. on pp. 22, 28, 40, 41).

[23] R. A. Calvo, S. D'Mello. "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications." In: *IEEE Transactions on Affective Computing* 1.1 (Jan. 2010), pp. 18–37. DOI: 10.1109/t-affc.2010.1. URL: https://doi.org/10.1109%2Ft-affc.2010.1 (cit. on pp. 22, 40).

[24] P. Marrero-Fernández, A. Montoya-Padrón, A. Jaume-i-Capó, J. M. B. Rubio. "Evaluating the Research in Automatic Emotion Recognition." In: *IETE Technical Review* 31.3 (May 2014), pp. 220–232. DOI: 10.1080/02564602.2014.906863. URL: https://doi.org/10.1080%2F02564602.2014.906863 (cit. on p. 22).

[25] D. Keltner, J. J. Gross. "Functional Accounts of Emotions." In: *Cognition &amp Emotion* 13.5 (Sept. 1999), pp. 467–480. DOI: 10.1080/026999399379140. URL: https://doi.org/10.1080%2F026999399379140 (cit. on p. 23).

[26] P. Ekman, D. Cordaro. "What is Meant by Calling Emotions Basic." In: *Emotion Review* 3.4 (Sept. 2011), pp. 364–370. DOI: 10.1177/1754073911410740. URL: https://doi.org/10.1177%2F1754073911410740 (cit. on p. 24).

[27] P. Ekman, W. V. Friesen. "Constants across cultures in the face and emotion." In: *Journal of Personality and Social Psychology* 17.2 (1971), pp. 124–129. DOI: 10.1037/h0030377. URL: https://doi.org/10.1037/h0030377 (cit. on pp. 24, 61).

[28] W.-L. Zheng, B.-L. Lu. "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks." In: *IEEE Transactions on Autonomous Mental Development* 7.3 (Sept. 2015), pp. 162–175. DOI: 10.1109/tamd.2015.2431497. URL: https://doi.org/10.1109%2Ftamd.2015.2431497 (cit. on pp. 24, 27, 55).

[29] J. A. Russell. "A circumplex model of affect." In: *Journal of Personality and Social Psychology* 39.6 (1980), pp. 1161–1178. DOI: 10.1037/h0077714. URL: https://doi.org/10.1037%2Fh0077714 (cit. on pp. 24, 25).

[30] H. Jasper. "The Ten-Twenty Electrode System of the International Federation." In: *Electroencephalography and Clinical Neurophysiology* 10 (1958), pp. 371–375 (cit. on p. 26).

[31] M. R. Nuwer, G. Comi, R. Emerson, A. Fuglsang-Frederiksen, J.-M. Guérit, H. Hinrichs, A. Ikeda, F. J. C. Luccas, P. Rappelsburger. "IFCN standards for digital recording of clinical EEG." In: *Electroencephalography and Clinical Neurophysiology* 106.3 (Mar. 1998), pp. 259–261. DOI: 10.1016/s0013-4694(97)00106-5. URL: https://doi.org/10.1016%2Fs0013-4694%2897%2900106-5 (cit. on p. 26).

[32] *Electroecephalography*. accessed on 03 June 2022. Oct. 2017. URL: https://www.britannica.com/print/article/183075 (cit. on p. 26).

[33] R. Ramirez, M. Palencia-Lefler, S. Giraldo, Z. Vamvakousis. "Musical neurofeedback for treating depression in elderly people." In: *Frontiers in Neuroscience* 9 (Oct. 2015). DOI: 10.3389/fnins.2015.00354. URL: https://doi.org/10.3389%2Ffnins.2015.00354 (cit. on p. 27).

[34] J. Allen. "Photoplethysmography and its application in clinical physiological measurement." In: *Physiological Measurement* 28.3 (Feb. 2007), R1–R39. DOI: 10.1088/0967-3334/28/3/r01. URL: https://doi.org/10.1088%2F0967-3334%2F28%2F3%2Fr01 (cit. on p. 27).

[35] A. S. Andreas, A. Haag, S. Goronzy, P. Schaich, J. Williams. "Emotion Recognition Using Bio-Sensors: First Steps Towards an." In: *In André et al (Eds.): Affective Dialogue Systems, Proceedings of the Kloster Irsee Tutorial and Research Workshop on Affective Dialogue Systems, Lecture Notes in Computer Science 3068, Springer-Verlag*. 2004, pp. 36–48 (cit. on p. 28).

[36] R. A. McCleary. "The nature of the galvanic skin response." In: *Psychological Bulletin* 47.2 (Mar. 1950), pp. 97–117. DOI: 10.1037/h0059810. URL: https://doi.org/10.1037%2Fh0059810 (cit. on p. 28).

[37] M. Sharma, S. Kacker, M. Sharma. "A Brief Introduction and Review on Galvanic Skin Response." In: *International Journal of Medical Research Professionals* 2.6 (Dec. 2016). DOI: 10.21276/ijmrp.2016.2.6.003. URL: https://doi.org/10.21276%2Fijmrp.2016.2.6.003 (cit. on p. 28).

[38] J. Domınguez-Jiménez, K. Campo-Landines, J. Martınez-Santos, E. Delahoz, S. Contreras-Ortiz. "A machine learning model for emotion recognition from physiological signals." In: *Biomedical Signal Processing and Control* 55 (Jan. 2020), p. 101646. DOI: 10.1016/j.bspc.2019.101646. URL: https://doi.org/10.1016%2Fj.bspc.2019.101646 (cit. on pp. 28, 45, 46, 48–52, 54, 56–58).

[39] D. Ayata, Y. Yaslan, M. E. Kamasak. "Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems." In: *Journal of Medical and Biological Engineering* 40.2 (Jan. 2020), pp. 149–157. DOI: 10.1007/s40846-019-00505-7. URL: https://doi.org/10.1007%2Fs40846-019-00505-7 (cit. on pp. 28, 45, 46, 49, 50, 55–58).

[40] B. Zhao, Z. Wang, Z. Yu, B. Guo. "EmotionSense: Emotion Recognition Based on Wearable Wristband." In: *2018 IEEE SmartWorld, Ubiquitous Intelligence &amp Computing, Advanced &amp Trusted Computing, Scalable Computing &amp Communications, Cloud &amp Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, Oct. 2018. DOI: 10.1109/smartworld.2018.00091. URL: https://doi.org/10.1109%2Fsmartworld.2018.00091 (cit. on pp. 28, 45, 48–52, 54, 56–58).

[41] Q. Zhang, X. Chen, Q. Zhan, T. Yang, S. Xia. "Respiration-based emotion recognition with deep learning." In: *Computers in Industry* 92-93 (Nov. 2017), pp. 84–90. DOI: 10.1016/j.compind.2017.04.005. URL: https://doi.org/10.1016%2Fj.compind.2017.04.005 (cit. on p. 29).

[42] A. S. Ginsburg, J. L. Lenahan, R. Izadnegahdar, J. M. Ansermino. "A Systematic Review of Tools to Measure Respiratory Rate in Order to Identify Childhood Pneumonia." In: *American Journal of Respiratory and Critical Care Medicine* 197.9 (May 2018), pp. 1116–1127. DOI: 10.1164/rccm.201711-2233ci. URL: https://doi.org/10.1164%2Frccm.201711-2233ci (cit. on p. 29).

[43] P. J. Lang, M. M. Bradley, B. N. Cuthbert, et al. "International affective picture system (IAPS): Technical manual and affective ratings." In: *NIMH Center for the Study of Emotion and Attention* 1.39-58 (1997), p. 3 (cit. on p. 30).

[44] M. Bradley, P. Lang. "The international affective digitized sounds: affective ratings of sounds and instruction manual (Technical Report No. B-3). University of Florida." In: *NIMH Center for the Study of Emotion and Attention, Gainesville, FL* (2007) (cit. on p. 30).

[45] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic. "A Multimodal Database for Affect Recognition and Implicit Tagging." In: *IEEE Transactions on Affective Computing* 3.1 (Jan. 2012), pp. 42–55. DOI: 10.1109/t-affc.2011.25. URL: https://doi.org/10.1109%2Ft-affc.2011.25 (cit. on pp. 30, 33).

[46] J. Xue, J. Wang, S. Hu, N. Bi, Z. Lv. "OVPD: Odor-Video Elicited Physiological Signal Database for Emotion Recognition." In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–12. DOI: 10.1109/tim.2022.3149116. URL: https://doi.org/10.1109%2Ftim.2022.3149116 (cit. on pp. 30, 33).

[47] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, L. S. Mattos. "Toward Emotion Recognition From Physiological Signals in the Wild: Approaching the Methodological Issues in Real-Life Data Collection." In: *Frontiers in Psychology* 11 (July 2020). DOI: 10.3389/fpsyg.2020.01111. URL: https://doi.org/10.3389%2Ffpsyg.2020.01111 (cit. on pp. 30, 31, 53).

[48] M. M. Bradley, P. J. Lang. "Measuring emotion: The self-assessment manikin and the semantic differential." In: *Journal of Behavior Therapy and Experimental Psychiatry* 25.1 (Mar. 1994), pp. 49–59. DOI: 10.1016/0005-7916(94)90063-9. URL: https://doi.org/10.1016%2F0005-7916%2894%2990063-9 (cit. on pp. 30, 31).

[49] W.-L. Zheng, J.-Y. Zhu, B.-L. Lu. "Identifying Stable Patterns over Time for Emotion Recognition from EEG." In: *IEEE Transactions on Affective Computing* 10.3 (July 2019), pp. 417–429. DOI: 10.1109/taffc.2017.2712143. URL: https://doi.org/10.1109%2Ftaffc.2017.2712143 (cit. on p. 32).

[50] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras. "DEAP: A Database for Emotion Analysis Using Physiological Signals." In: *IEEE Transactions on Affective Computing* 3.1 (Jan. 2012), pp. 18–31. DOI: 10.1109/t-affc.2011.15. URL: https://doi.org/10.1109%2Ft-affc.2011.15 (cit. on p. 32).

[51] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, I. Patras. "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups." In: *IEEE Transactions on Affective Computing* 12.2 (Apr. 2021), pp. 479–493. DOI: 10.1109/taffc.2018.2884461. URL: https://doi.org/10.1109%2Ftaffc.2018.2884461 (cit. on pp. 33, 52).

[52] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, N. Sebe. "ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors." In: *IEEE Transactions on Affective Computing* 9.2 (Apr. 2018), pp. 147–160. DOI: 10.1109/taffc.2016.2625250. URL: https://doi.org/10.1109%2Ftaffc.2016.2625250 (cit. on pp. 33, 52).

[53] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, N. Sebe. "DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses." In: *IEEE Transactions on Affective Computing* 6.3 (July 2015), pp. 209–222. DOI: 10.1109/taffc.2015.2392932. URL: https://doi.org/10.1109%2Ftaffc.2015.2392932 (cit. on p. 33).

[54] S. Katsigiannis, N. Ramzan. "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices." In: *IEEE Journal of Biomedical and Health Informatics* 22.1 (Jan. 2018), pp. 98–107. DOI: 10.1109/jbhi.2017.2688239. URL: https://doi.org/10.1109%2Fjbhi.2017.2688239 (cit. on p. 33).

[55] T. Xue, A. E. Ali, T. Zhang, G. Ding, P. Cesar. "RCEA-360VR: Real-time, Continuous Emotion Annotation in 360° VR Videos for Collecting Precise Viewport-dependent Ground Truth Labels." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, May 2021. DOI: 10.1145/3411764.3445487. URL: https://doi.org/10.1145%2F3411764.3445487 (cit. on p. 33).

[56] B. Ren, D. Liu, L. Qi. "Emotion recognition based on multiple order features using fractional Fourier transform." In: *SPIE Proceedings*. Ed. by C. M. Falco, X. Jiang. SPIE, July 2017. DOI: 10.1117/12.2281534. URL: https://doi.org/10.1117%2F12.2281534 (cit. on p. 34).

[57] J. Zhang, Z. Yin, P. Chen, S. Nichele. "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review." In: *Information Fusion* 59 (July 2020), pp. 103–126. DOI: 10.1016/j.inffus.2020.01.011. URL: https://doi.org/10.1016%2Fj.inffus.2020.01.011 (cit. on p. 35).

[58] I. T. Jolliffe, J. Cadima. "Principal component analysis: a review and recent developments." In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (Apr. 2016), p. 20150202. DOI: 10.1098/rsta.2015.0202. URL: https://doi.org/10.1098%2Frsta.2015.0202 (cit. on p. 35).

[59] J. Lever, M. Krzywinski, N. Altman. "Principal component analysis." In: *Nature Methods* 14.7 (July 2017), pp. 641–642. DOI: 10.1038/nmeth.4346. URL: https://doi.org/10.1038%2Fnmeth.4346 (cit. on p. 35).

[60] D. A. Pisner, D. M. Schnyer. "Support vector machine." In: *Machine Learning*. Elsevier, 2020, pp. 101–121. DOI: 10.1016/b978-0-12-815739-8.00006-7. URL: https://doi.org/10.1016%2Fb978-0-12-815739-8.00006-7 (cit. on p. 36).

[61] L. Peterson. "K-nearest neighbor." In: *Scholarpedia* 4.2 (2009), p. 1883. DOI: 10.4249/scholarpedia.1883. URL: https://doi.org/10.4249%2Fscholarpedia.1883 (cit. on p. 36).

[62] H. Rajaguru, S. K. Prabhakar. “Non linear ICA and logistic regression for classification of epilepsy from EEG signals.” In: *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, Apr. 2017. DOI: 10.1109/iceca.2017.8203602. URL: https://doi.org/10.1109%2Ficeca.2017.8203602 (cit. on p. 37).

[63] K. P. Murphy. *Naive Bayes classifiers*. 2006. URL: https://www.ic.unicamp.br/~rocha/teaching/2011s1/mc906/aulas/naive-bayes.pdf (cit. on p. 37).

[64] K. O’Shea, R. Nash. *An Introduction to Convolutional Neural Networks*. 2015. DOI: 10.48550/ARXIV.1511.08458. URL: https://arxiv.org/abs/1511.08458 (cit. on pp. 37, 55).

[65] P. Thiam, P. Bellmann, H. A. Kestler, F. Schwenker. “Exploring Deep Physiological Models for Nociceptive Pain Recognition.” In: *Sensors* 19.20 (Oct. 2019), p. 4503. DOI: 10.3390/s19204503. URL: https://doi.org/10.3390%2Fs19204503 (cit. on p. 37).

[66] P. Refaeilzadeh, L. Tang, H. Liu. “Cross-Validation.” In: *Encyclopedia of Database Systems*. New York, NY: Springer New York, 2016, pp. 1–7 (cit. on p. 38).

[67] R. J. Torraco. “Writing Integrative Literature Reviews: Guidelines and Examples.” In: *Human Resource Development Review* 4.3 (Sept. 2005), pp. 356–367. DOI: 10.1177/1534484305278283. URL: https://doi.org/10.1177%2F1534484305278283 (cit. on p. 39).

[68] J. Zhang, Z. Yin, P. Chen, S. Nichele. “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review.” In: *Information Fusion* 59 (July 2020), pp. 103–126. DOI: 10.1016/j.inffus.2020.01.011. URL: https://doi.org/10.1016%2Fj.inffus.2020.01.011 (cit. on p. 40).

[69] N. Mejbri, F. Essalmi, M. Jemni, B. A. Alyoubi. “Trends in the use of affective computing in e-learning environments.” In: *Education and Information Technologies* 27.3 (Oct. 2021), pp. 3867–3889. DOI: 10.1007/s10639-021-10769-9. URL: https://doi.org/10.1007%2Fs10639-021-10769-9 (cit. on pp. 40, 41).

[70] T. B. Alakus, M. Gonen, I. Turkoglu. “Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO.” In: *Biomedical Signal Processing and Control* 60 (July 2020), p. 101951. DOI: 10.1016/j.bspc.2020.101951. URL: https://doi.org/10.1016%2Fj.bspc.2020.101951 (cit. on pp. 45, 46, 48–52, 54–58).

[71] M. Ali, F. Machot, A. Mosa, M. Jdeed, E. Machot, K. Kyamakya. “A Globally Generalized Emotion Recognition System Involving Different Physiological Signals.” In: *Sensors* 18.6 (June 2018), p. 1905. DOI: 10.3390/s18061905. URL: https://doi.org/10.3390%2Fs18061905 (cit. on pp. 45, 46, 48–50, 52–58).

[72] M. Awais, M. Raza, N. Singh, K. Bashir, U. Manzoor, S. U. Islam, J. J. P. C. Rodrigues. "LSTM-Based Emotion Detection Using Physiological Signals: IoT Framework for Healthcare and Distance Learning in COVID-19." In: *IEEE Internet of Things Journal* 8.23 (Dec. 2021), pp. 16863–16871. DOI: 10.1109/jiot.2020.3044031. URL: https://doi.org/10.1109%2Fjiot.2020.3044031 (cit. on pp. 45, 46, 49, 50, 54, 55, 57, 58, 61).

[73] S. Balic, L. Kleybolte, C. Märtin. "A Swarm Intelligence Approach: Combination of Different EEG-Channel Optimization Techniques to Enhance Emotion Recognition." In: *Human-Computer Interaction. Technological Innovation*. Ed. by M. Kurosu. Cham: Springer International Publishing, 2022, pp. 303–317. ISBN: 978-3-031-05409-9 (cit. on pp. 45, 46, 49, 50, 54–58, 60).

[74] V. Doma, M. Pirouz. "A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals." In: *Journal of Big Data* 7.1 (Mar. 2020). DOI: 10.1186/s40537-020-00289-7. URL: https://doi.org/10.1186%2Fs40537-020-00289-7 (cit. on pp. 45, 46, 49, 50, 52, 54–58).

[75] D. Garg, G. K. Verma. "Emotion Recognition in Valence-Arousal Space from Multichannel EEG data and Wavelet based Deep Learning Framework." In: *Procedia Computer Science* 171 (2020), pp. 857–867. DOI: 10.1016/j.procs.2020.04.093. URL: https://doi.org/10.1016%2Fj.procs.2020.04.093 (cit. on pp. 45, 46, 49, 50, 54, 55, 57, 58).

[76] W. He, Y. Ye, Y. Li, T. Pan, L. Lu. "Online Cross-subject Emotion Recognition from ECG via Unsupervised Domain Adaptation." In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine &amp Biology Society (EMBC)*. IEEE, Nov. 2021. DOI: 10.1109/embc46164.2021.9630433. URL: https://doi.org/10.1109%2Fembc46164.2021.9630433 (cit. on pp. 45, 46, 49, 50, 53–58).

[77] F. Panahi, S. Rashidi, A. Sheikhani. "Application of fractional Fourier transform in feature extraction from ELECTROCARDIOGRAM and GALVANIC SKIN RESPONSE for emotion recognition." In: *Biomedical Signal Processing and Control* 69 (Aug. 2021), p. 102863. DOI: 10.1016/j.bspc.2021.102863. URL: https://doi.org/10.1016%2Fj.bspc.2021.102863 (cit. on pp. 45, 47, 49, 50, 56–58).

[78] P. Pandey, K. R. Seeja. "Subject independent emotion recognition system for people with facial deformity: an EEG based approach." In: *Journal of Ambient Intelligence and Humanized Computing* 12.2 (July 2020), pp. 2311–2320. DOI: 10.1007/s12652-020-02338-8. URL: https://doi.org/10.1007%2Fs12652-020-02338-8 (cit. on pp. 45, 47, 49, 50, 52–55, 57, 58).

[79] A. Sepúlveda, F. Castillo, C. Palma, M. Rodriguez-Fernandez. "Emotion Recognition from ECG Signals Using Wavelet Scattering and Machine Learning." In: *Applied Sciences* 11.11 (May 2021), p. 4945. DOI: 10.3390/app11114945. URL: https://doi.org/10.3390%2Fapp11114945 (cit. on pp. 45, 47, 49, 50, 54, 56–58).

[80] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, F. Schwenker. "A dataset of continuous affect annotations and physiological signals for emotion analysis." In: *Scientific Data* 6.1 (Oct. 2019). DOI: 10.1038/s41597-019-0209-0. URL: https://doi.org/10.1038%2Fs41597-019-0209-0 (cit. on pp. 45, 47–52, 55, 58).

[81] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, Z. Cui. "MPED: A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition." In: *IEEE Access* 7 (2019), pp. 12177–12191. DOI: 10.1109/access.2019.2891579. URL: https://doi.org/10.1109%2Faccess.2019.2891579 (cit. on pp. 45, 47–52, 56–58).

[82] S. Vijayakumar, R. Flynn, N. Murray. "A Comparative Study of Machine Learning Techniques for Emotion Recognition from Peripheral Physiological Signals." In: *2020 31st Irish Signals and Systems Conference (ISSC)*. IEEE, June 2020. DOI: 10.1109/issc49989.2020.9180193. URL: https://doi.org/10.1109%2Fissc49989.2020.9180193 (cit. on pp. 45, 47, 49, 50, 53, 56–58).

[83] X. Zhang, J. Liu, J. Shen, S. Li, K. Hou, B. Hu, J. Gao, T. Zhang, B. Hu. "Emotion Recognition From Multimodal Physiological Signals Using a Regularized Deep Fusion of Kernel Machine." In: *IEEE Transactions on Cybernetics* 51.9 (Sept. 2021), pp. 4386–4399. DOI: 10.1109/tcyb.2020.2987575. URL: https://doi.org/10.1109%2Ftcyb.2020.2987575 (cit. on pp. 45, 48–50, 55–58, 60).

[84] T. Zhang, A. E. Ali, A. Hanjalic, P. Cesar. "Few-shot Learning for Fine-grained Emotion Recognition using Physiological Signals." In: *IEEE Transactions on Multimedia* (2022), pp. 1–1. DOI: 10.1109/tmm.2022.3165715. URL: https://doi.org/10.1109%2Ftmm.2022.3165715 (cit. on pp. 45, 48–53, 56–58).

[85] T. Zhang, A. E. Ali, C. Wang, A. Hanjalic, P. Cesar. "Weakly-supervised Learning for Fine-grained Emotion Recognition using Physiological Signals." In: *IEEE Transactions on Affective Computing* (2022), pp. 1–1. DOI: 10.1109/taffc.2022.3158234. URL: https://doi.org/10.1109%2Ftaffc.2022.3158234 (cit. on pp. 45, 48–51, 53–55, 57, 58).

[86] C. M. D. Acevedo, J. K. C. Gómez, C. A. A. Rojas. "Academic stress detection on university students during COVID-19 outbreak by using an electronic nose and the galvanic skin response." In: *Biomedical Signal Processing and Control* 68 (July 2021), p. 102756. DOI: 10.1016/j.bspc.2021.102756. URL: https://doi.org/10.1016%2Fj.bspc.2021.102756 (cit. on pp. 45, 46, 48–50, 52–58).

[87] V. Gupta, M. D. Chopda, R. B. Pachori. "Cross-Subject Emotion Recognition Using Flexible Analytic Wavelet Transform From EEG Signals." In: *IEEE Sensors Journal* 19.6 (Mar. 2019), pp. 2266–2274. DOI: 10.1109/jsen.2018.2883497. URL: https://doi.org/10.1109%2Fjsen.2018.2883497 (cit. on pp. 45, 46, 49, 50, 53–58).

[88] R. Lutze, K. Waldhör. "Practical Suitability of Emotion Recognition from Physiological Signals by Mainstream Smartwatches." In: *Human-Computer Interaction. Technological Innovation*. Ed. by M. Kurosu. Cham: Springer International Publishing, 2022, pp. 362–375. ISBN: 978-3-031-05409-9 (cit. on pp. 45, 47–49, 52–54, 58).

[89] C. Wei, L.-l. Chen, Z.-z. Song, X.-g. Lou, D.-d. Li. "EEG-based emotion recognition using simple recurrent units network and ensemble learning." In: *Biomedical Signal Processing and Control* 58 (Apr. 2020), p. 101756. DOI: 10.1016/j.bspc.2019.101756. URL: https://doi.org/10.1016%2Fj.bspc.2019.101756 (cit. on pp. 45, 47, 49, 50, 53–58).

[90] C.-J. Yang, N. Fahier, C.-Y. He, W.-C. Li, W.-C. Fang. "An AI-Edge Platform with Multimodal Wearable Physiological Signals Monitoring Sensors for Affective Computing Applications." In: *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, Oct. 2020. DOI: 10.1109/iscas45731.2020.9180909. URL: https://doi.org/10.1109%2Fiscas45731.2020.9180909 (cit. on pp. 45, 47–50, 52, 54, 57, 58).

[91] K. Yang, C. Wang, Y. Gu, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, J. Goncalves. "Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition." In: *IEEE Transactions on Affective Computing* (2021), pp. 1–1. DOI: 10.1109/taffc.2021.3100868. URL: https://doi.org/10.1109%2Ftaffc.2021.3100868 (cit. on pp. 45, 47–52, 54, 56–58).

[92] T. Zhang, A. E. Ali, C. Wang, A. Hanjalic, P. Cesar. "RCEA: Real-time, Continuous Emotion Annotation for Collecting Precise Mobile Video Ground Truth Labels." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Apr. 2020. DOI: 10.1145/3313831.3376808. URL: https://doi.org/10.1145%2F3313831.3376808 (cit. on p. 53).

[93] X. Chai, Q. Wang, Y. Zhao, Y. Li, D. Liu, X. Liu, O. Bai. "A Fast, Efficient Domain Adaptation Technique for Cross-Domain Electroencephalography(EEG)-Based Emotion Recognition." In: *Sensors* 17.5 (May 2017), p. 1014. DOI: 10.3390/s17051014. URL: https://doi.org/10.3390%2Fs17051014 (cit. on p. 57).

All links were last followed on July 31, 2022.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

Esslingen, 31.07.2022 Evelin Grande

place, date, signature