Institute for Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Masterarbeit

Correlating Facial Expressions and Contextual Data for Mood Prediction Using Mobile Devices

Robin Reutter

Course of Study: Informatik

Examiner: Prof. Dr. Albrecht Schmidt

Supervisor: Thomas Kosch M.Sc.
Mariam Hassib M.Sc.

Commenced: May 3, 2017

Completed: November 3, 2017

CR-Classification: I.5.4, H.5.m

Abstract

Facial recognition can nowadays be achieved by any casual smartphone with a camera. Sophisticated systems and methods allow extracting information from facial data such as connected emotions depending on a person's current expressions. In a mobile setting, information about emotions can be constantly grasped from a person using the smartphone front camera and annotated with various types of contextual data. This master thesis introduces *OpenFaceAndroid* – an Android application based on the existing facial analysis frameworks *OpenFace* and *OpenFace++*. The system allows to gather and process facial expression information as well as contextual data in real-time on a smartphone using the front camera device and various sensors. The output is a prediction of seven different emotions which is – compared to pure facial data extraction – improved through annotation with context. In two conducted studies first off data from several participants is collected, assessed for their usefulness in terms of this master thesis and afterward utilized to learn classifier models taking live emotion values and context information as training data. Subsequently, these models are evaluated for their accuracy in general emotion prediction and noticing affective mood changes – supported by findings of participant interviews. In conclusion, possible improvements and general limitations of this work are discussed as well as suggestions for future work proposed.

Kurzfassung

Heutzutage kann Gesichtserkennung über jedes gewöhnliche Smartphone mit Kamera durchgeführt werden. Fortgeschrittene Systeme und Methoden erlauben das Extrahieren von gesichtsbasierten Daten, wie beispielsweise den damit verbundenen Emotionen, von denen die Gesichtsausdrücke der jeweiligen Person abhängen. In einer mobilen Umgebung können durch die Frontkamera durchgehend Informationen über die Gefühle des Nutzers gesammelt werden und es gibt zahlreiche Arten von kontextbezogenen Daten mit denen man diese annotieren kann. In dieser Masterarbeit wird *OpenFaceAndroid* – eine Android Applikation, die auf den existierenden Gesichtsanalyse-Frameworks *OpenFace* und *OpenFace++* basiert – vorgestellt. Das System ermöglicht das Sammeln und Verarbeiten von Informationen über Gesichtsausdrücke sowie kontextbezogenen Daten. Dies geschieht in Echtzeit mithilfe der Frontkamera eines Smartphones und verschiedenen Sensoren. Die Ausgabe ist eine Prognose von sieben verschiedenen Emotionen, die, verglichen mit reiner Extraktion aus gesichtsbasierten Daten, durch Kontextannotation verbessert wird. In zwei durchgeführten Studien werden zunächst Daten von mehreren Teilnehmern gesammelt und deren Eignung im Rahmen dieser Masterarbeit bewertet. Danach werden die Daten zur Erstellung eines Klassifikators verwendet, welcher Echtzeitwerte über Emotionen und Kontextinformationen als Trainingsdaten benutzt. Anschließend wird die Genauigkeit der generierten Modelle bezüglich der Emotionsprognose sowie dem Erkennen von affektiven Stimmungsänderungen ausgewertet. Dies wird durch Erkenntnisse aus Teilnehmerbefragungen gestützt. Zum Abschluss werden Verbesserungsmöglichkeiten und allgemeine Einschränkungen im Zusammenhang mit dieser Arbeit diskutiert und zukünftige anknüpfende Forschungstätigkeiten vorgeschlagen.

Contents

List of Figures

List of Tables

List of Listings

List of Abbreviations

API Application Programming Interface. 29

ARFF Artificial File. 48

AU Action Unit. 19

CLM Constrained Local Model. 19

CLNF Conditional Local Neural Fields. 19

CPU Central Processing Unit. 43

CSV Comma-Separated values. 35

EDA Electrodermal Activity Sensor. 24

EEG Electroencephalography. 15

FPS frames per second. 20

GPS Global Positioning System. 33

GUI Graphical User Interface. 29

ID identifier. 29

IDE Integrated Development Environment. 29

MB megabytes. 35

OS operating system. 24

RQ research question. 16

RSS Rich Site Summary. 35

SMS Short Message Service. 58

SVM Support Vector Machine. 24

SVR Support Vector Regression. 20

Weka Waikato Environment for Knowledge Analysis. 48

XML Extensible Markup Language. 29

1 Introduction

There are many different ways emotions are expressed, might it be speech, body language, or facial expressions – the latter providing superior abilities of reflecting emotions [EFE13]. Facial expression behavior has been examined for a long time reaching back to Darwin [Dar99] discovering generally valid similarities for humans. These aspects have been continuously seized and used in further research work by Ekman and Friesen [EF70]. More than a hundred years after Darwin and having evolved as an important topic in both science and industry, solutions for analyzing and extracting information have come up in both intrusive – for instance by tracking of Electroencephalography (EEG) waves – and non-intrusive approaches, such as facial expression analysis by facial data extraction and facial expression recognition from pictures and videos. Since many years evaluating facial expressions can be accomplished live and in real-time using the video stream of camera devices [BLFM03; JY02; OPB00]. With the increasing amount of smartphone users [Pou16; Sta17a; Sta17b] this can be performed for a large user base at any time and in any situation using smartphone cameras, provided that the user is facing the camera for a large number of people.

McDuff et al. [MKK+12] have built a system which visualizes data collected by bio-sensors together with contextual data of one's smartphone. Automatic annotation by using context data in terms of a user's stress level has been researched upon as well by Ayzenberg and Picard [AP14]. Therefore they developed a system using custom stress recognition software upon extracted data in order to give information about context and corresponding stress level. Their research results show that context information becomes an important factor for annotating signal data.

Correlation between facial expressions and contextual data might enable to give better predictions about current emotions of a person. However, no approach on correlating emotions and context in real-time and in a mobile setting, providing mobile contextual data, has been performed yet.

Recently, the face analysis framework *OpenFace++* [Ban17a] was developed, allowing to extract emotions such as happiness, sadness, and anger by evaluating facial expressions in both stationary desktop as well as mobile settings. The continuously growing usage of mobile devices offers a great data source and application field. In this master thesis, possible correlations between extracted facial information using smartphones and their context, such as current smartphone application usage, geolocation, calendar events, or weather are investigated. This will be done by conducting user studies, with simultaneous consideration of these properties' suitability. Therefore technical aspects regarding the implementation, the respective study duration, and sample size are observed. Identifying

correlations could lead to both a derivation and possible conceptual mood prediction based on gathered data.

A literature research will give an overview of existing approaches and results of different mood derivation as well as prediction methods. Based on the ideas of existing work a concept for the technical implementation is established and research questions (RQs) are defined. Afterward, a prototype based on the existing Android implementation of *OpenFace++* is developed and extended to work in the background of one's smartphone. A controlled study will then help to evaluate its effectiveness of mood derivation and prediction. Finally, the evaluation of the study results is intended to provide answers to the predefined research questions.

There are countless possible use cases for such systems. Starting on the device itself, smartphone applications and interaction could be observed and classified whether causing positive or negative mood. For instance, social media (and thus their respective applications) have been analyzed in previous work [DVD12; FZCX13], yielding different discussions in their study results in which way they might have an effect on certain emotions. Thus, a system which actively and accurately monitors and predicts mood could certainly help smartphone users to better understand and handle emotional effects caused by smartphone-related context such as the used applications on them. Or think of emotion analysis for increasing self-awareness about caused affective mood changes, enabling people to identify situations and happenings actively affecting their emotions. Commercial use might appear in terms of advertisement addressing to provoke certain emotions when being watched.

Their effectiveness could be easily quantified having confident propositions about their actual mood. Outreached even further certain negative situations connected to people's mood could possibly be avoided if a predictive system recognized these and warned its user beforehand. With the seamless integration of smartphones in everyday life, this could be a vital improvement and help in daily life.

1.1 Outline

This thesis is structured as follows:

Chapter 2 – Background: In this chapter, basic information about emotions and emotional analysis is given. Furthermore, facial data extraction methods used in this work and their characteristics are sketched.

Chapter 3 – Related Work: This chapter shows and discusses research that has been performed on emotional analysis using different technical approaches.

Chapter 4 – Concept: In this chapter, the basic concept for the technical implementation of this master thesis is presented, defining its scope of application and requirements.

Chapter 5 – Technical Implementation: This chapter introduces the technical implementation of the Android application, followed by a discussion about connected limitations.

Chapter 6 – Data Collection and Model Prediction Studies: In this chapter both user studies and the steps in between are presented. Methods used to process data and built predictive models are explained.

Chapter 7 – Evaluation: In this chapter, the evaluation of all gathered data during the user studies is presented, including both information obtained through the application as well as the answers the study participants gave in the interviews.

Chapter 8 – Conclusion: This chapter summarizes the results of this thesis. Based on the insights obtained possible improvements are suggested for future work.

2 Background

This chapter covers basics of different facial emotions, followed by facial data extract methods and related emotion classification in particular utilized in this thesis. Its possibilities and limitations are discussed afterward.

2.1 Basic Emotions

The different emotion types attended in this thesis are a subset of the eight basic emotions defined by Plutchik and Conte [PC97], visualized in the so-called *Plutchik's wheel of emotions* (see Figure 2.1 in the second inner circle). Different studies conducted to find out which emotions can be expressed by the face have been examined by Ekman and Friesen [EF70] resulting into:

1. *Anger*
2. *Disgust*
3. *Fear*
4. *Joy (called happiness in this thesis)*
5. *Sadness*
6. *Surprise*

These are according to Ekman, Friesen, and Ellsworth [EFE13] the basic emotions that can be derived from facial expressions.

2.2 Facial Data Extraction

The predictive mood system created in this master thesis utilizes the work of Banzhaf [Ban17a] and its library *OpenFace++*. It is based on the facial data extraction library *OpenFace* of Baltrušaitis, Robinson, and Morency [BRM16]. *OpenFace* enables to perform real-time facial behavior analysis upon picture and video feed. Therefore new approaches for detecting and tracking facial landmarks are exploited [BRM13], using an advanced method of Constrained Local Model (CLM) [CC06] called Conditional Local Neural Fields (CLNF) which was also further extended to amongst others support face validation. Furthermore, head pose translation and orientation, as well as eye gaze, are estimated. The latter is performed separately to the actual facial landmarks and is provided through the 3D position of the eye and a 3D vector originating from the pupil [WBZ+15]. Action Units

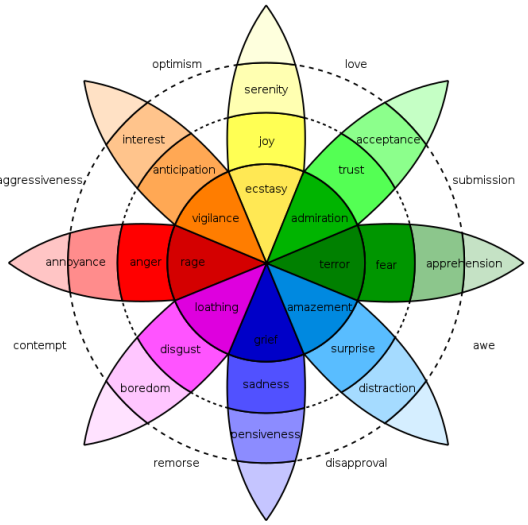


Figure 2.1: Plutchik's wheel of emotions [Plu91]

(AUs) such as "lips part" or "upper lid raiser" allow to analyze facial expressions and are recognized automatically.

Banzhaf [Ban17a] extends *OpenFace* with features such emotion classification derived from AUs, attention estimation, and Android support as well as further implementation changes. Emotion classification is realized through machine learning using Support Vector Regression (SVR) on a training sample provided by Lucey et al. [LCK+10]. In addition, an attention score ranging from zero to one is calculated through the eye gaze vector *OpenFace* supplies. If looked directly in the respective camera device, the attention score is maximal. Minimum attention score is yielded if a certain threshold angle is reached.

The emotion classification ability and Android support are in particular point of interest. They establish a basis for emotion evaluation in a mobile setting. *OpenFace++* allows detecting the six basic facial expressible emotions in real-time from pictures. In terms of this thesis, they are provided by a camera device's video feed. Therefore image data is processed and probabilities computed for each emotion as visualized in Figure 2.2. The six emotions are extended by a seventh emotion "Neutral" which is reported back if none of the six others is detected. Further information as the attention score expressing how directly the eyes gaze to the front camera as well as face position and orientation might possibly help in later data evaluation as well.

2.2.1 Limitations of Visual Analysis

Banzhaf [Ban17a] discusses the limitations of emotion classification upon AUs. As for the technical side, AUs might not be extracted correctly from the image, thus classification is also wrong or they are completely missing as not all AUs are supported by *OpenFace*.

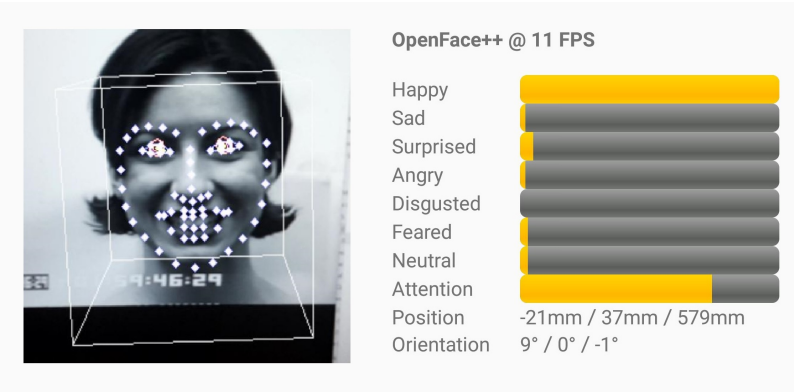


Figure 2.2: Exemplary evaluation of emotions derived from facial data of a picture using *OpenFace++*. In this demonstration, probability values from zero to one are visualized as orange bars. Additional information for frame rate (in frames per second (FPS)), attention score, 3D face position, and 3D face orientation is given [Ban17b] (edited).

(a) Anger **(b)** Disgust **(c)** Fear

(d) Happiness **(e)** Sadness **(f)** Suprise

Figure 2.3: CK+ database [LCK+10] examples of the six basic emotions (modified version of [SYLT14]). The intensity of these facial expressions such as surprise in Figure 2.3f and disgust in 2.3b give rise to discussion how applicable they are in the wild.

Furthermore, results might be better if a series of frames is evaluated in coherent context, instead of picture by picture. Usually facial expressions do not follow an ideal representation as sample pictures do (see Figure 2.3), but, for instance, combinations of feelings, which might lead to ambiguous and wrong results. Banzhaf [Ban17a] also mentions that pure visual evaluation does not give full information about emotions. An issue which is directly addressed in this thesis by trying to connect contextual data to classified emotions *OpenFace++* yields.

3 Related Work

This chapter gives an overview of research that has already been performed on related topics.

For this master thesis in particular scientific work dealing with facial expression analysis and work addressing emotion analysis using technical interfaces are of peculiar interest. Also, work aiming at specific topic areas such as social media are taken account of.

One of the earlier research work before smartphones were introduced was performed by Iso, Kwasaki, and Kurakake [IKK05]. They equipped a folding cellphone with seven additional sensor types for personal context determination (Figure 3.1). Standing out from related work until then, all sensor types used were mounted on the device, the user didn't have to wear any additional sensors.

Some information gathered by sensor types used in this work is covered by nowadays smartphones, such as acceleration and angular rate sensors or microphones allowing to record and process audio data. These sensors can form possible information sources for this master thesis in particular for contextual data described as "Physical User Context" by Iso, Kwasaki, and Kurakake [IKK05], including information such as the "user's physical action state such as running, walking, or sitting".

Hernandez and Picard [HP14] suggest using "Google Glass" as a wearable device to amplify its user's ability to sense emotions. The benefit of various sensors provided by Google Glass

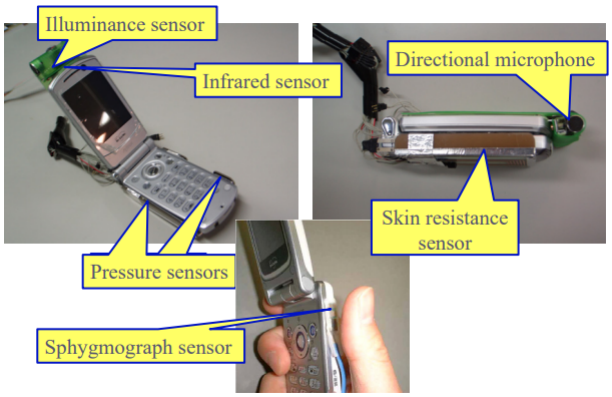


Figure 3.1: A regular mobile phone equipped with multiple additional sensors for obtaining "Phyisical User Context" [IKK05].

Emotion	Percent correct
Anger	82.2%
Disgust	84.6%
Fear	71.7%
Joy (Happiness)	93.0%
Sorrow (Sadness)	85.4%
Surprise	99.3%

Table 3.1: The results of Michel and El Kaliouby [ME03] show that joy (called happiness in this thesis) and surprise yielded best classification accuracies. Overall also anger, disgust, and sorrow (called sadness in this thesis) reached scores higher than 80%, leaving fear with the lowest score, but still above 70%.

is utilized in their work by integrating smile recognition software and directly displaying gained information on the device.

Hernandez and Picard [HP14] also point out limitations, mentioning battery and storage issues when acquiring data from many sensors at the same time and possible privacy violation inflictions. These limitations might arise in the context of this master thesis as well and should be examined in detail.

Ayzenberg and Picard [AP14] worked on connecting data from bio-physiological sensors and respective context. In a study, they collected data their participants provided captured by Electrodermal Activity Sensor (EDA) sensors. Mobile contextual data acquired by a smartphone application included location and calendar events and were annotated with sensor data. In their developed system "FEEL" they allowed users to annotate these recordings to receive full context information later on. The guidelines that came along with their work in terms of systems working with contextual information and corresponding mobile application architecture give good directions for this master thesis. Unobtrusive data acquisition and the choice of a widely spread operating system (OS), supporting development of applications and easy access to various data sources are just a few keywords here.

Michel and El Kaliouby [ME03] addressed facial expression recognition in their research work, trying to apply a Support Vector Machine (SVM) classifier on extracted facial features of video streams. Table 3.1 shows results of their classifier training. An average accuracy of 86% was achieved. This leads to assuming SVM as a possibly well-suiting model for training a classifier in the context of this master thesis.

McDuff et al. [MKK+12] introduced "AffectAura", a system for tracking and predicting emotions. Three features "valence, arousal, and engagement" were computed by the input of several sensors in a desktop setting (illustrated in Figure 3.2). Gathered data from of a user study with five participants was later on processed by machine learning methods training a nearest neighbor classifier and yielding predictive models. [MKK+12] Results from classifier validation by using a leave-one-out analysis showed an average

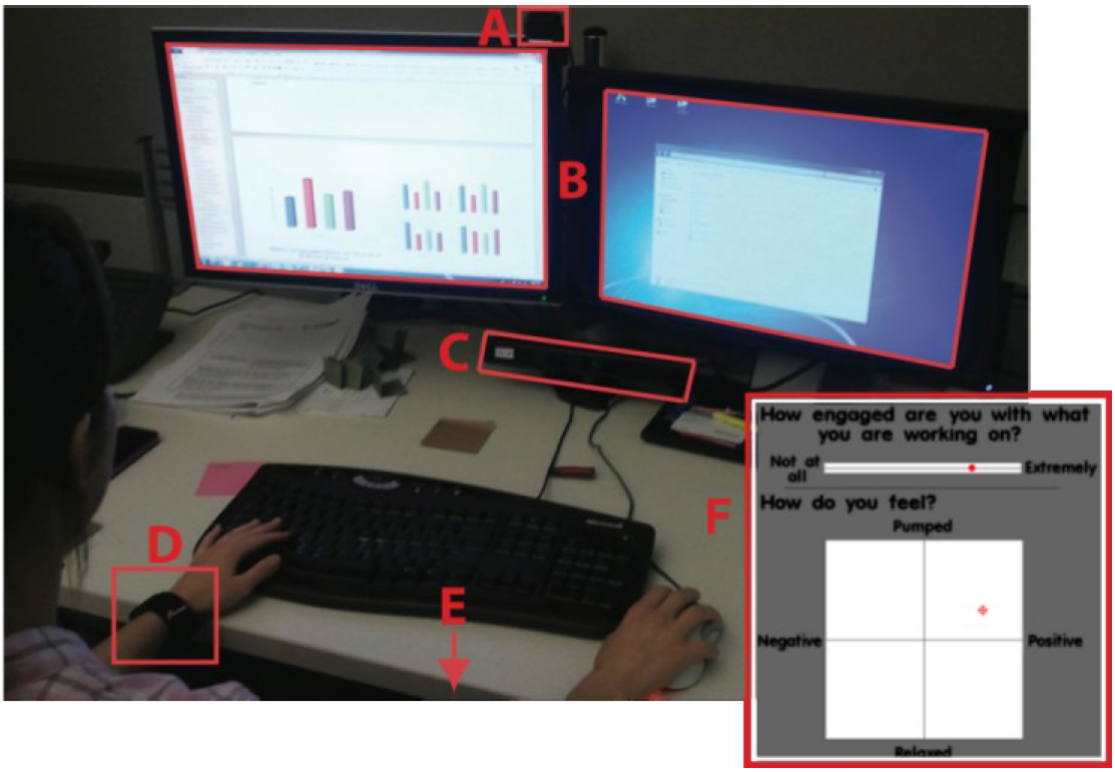


Figure 3.2: The desktop setup of McDuff et al. [MKK+12] with its sensors: A webcam (A), file activity and calendar monitor (B), Microsoft Xbox Kinect (C), EDA sensor (D), GPS (E), and self-report interface (F).

correctness of 68% and a top correctness of 71%. User studies proved that these numbers were sufficient indicating a possible range for this thesis.

Coherences between emotions and mobile phone interactions as well as mutual influencing have recently been investigated by Mehrotra et al. [MTHM17] with the aim to explore causal links. Therefore they used several common metrics and also newly introduced ones such as special phone usage metrics. The emotional states examined contained activeness, happiness, and stress. An in-the-wild study yielded over 5000 mood questionnaire responses which were analyzed and evaluated afterward. Their observed time periods included the preceding, current, and next hour period – which also forms possible time frames for this master thesis. The results showed amongst others causality between users' emotional states and application usage as well as correlations between general smartphone usage and happiness. These aspects are also to be further examined in context of this thesis.

With the insight gained from related work in terms of contextual data and emotion analysis, in this thesis instead of local desktop setups mood prediction of all six basic emotions introduced in Chapter 2 will be evaluated in a mobile setting with real-time facial expression analysis and mobile contextual information.

4 Concept

This chapter discusses the concept of a mobile mood prediction system which is created in this thesis. It serves as a foundation for the implementation followed in the next chapter.

4.1 Basic Idea

For a mood prediction system using facial data extraction and contextual data in a mobile setting, an application has to be developed conflating diverse information gathered from various sources on a mobile device. The application collects and stores this information in a controlled study. Analyzing and processing the data will give insight about which contextual data is suitable for annotation. Afterward, the selected information is used as attributes in order to train models using machine learning techniques. Eventually, these predictive models are evaluated in a subsequent study.

4.2 Mobile Application Requirements

In this thesis, the face analysis framework "OpenFace++" [Ban17a] is utilized to gather and store information about the basic emotions listed in Section 2.1 using facial data extraction.

The mobile application has to integrate the framework and is intended to work in the background of its running device, while the device itself can just be used normally during daily operation. As *OpenFace++* processes image data (see Section 2.2), for this master thesis, mobile devices capable of sufficient computation power providing front cameras are mandatory. Contextual information available on a mobile device such as the currently used application, user interaction, geolocation and weather data, or calendar events will be additionally collected and stored. Study participants have to be able to rate their emotions based on their actual feelings. As different research [LF04; WH89] suggests, multiple emotions can be experienced at the same time. Thus it should be possible to combine emotion ratings in the application. The input is later on taken as class labels for training the classifier.

4.3 Predictive Model Requirements

Using machine learning techniques, a predictive model will be derived from both the data collected as well as the participant's emotion rating. A reasonable accuracy score is intended to be achieved which can be adjusted by picking suitable variables as classifier attributes. The yielded models must be integrated into the mobile application for the second study allowing to predict emotion states during daily smartphone usage. Eventually, model prediction can be compared against user rating and facial expression data derived from *OpenFace++*.

4.4 Research Questions

Based on the concept we define the following RQs for this thesis:

RQ 1 Are there certain patterns in affective emotion state changes and emotions in general connected to applications used at that time?

RQ 2 Does annotating contextual information to face extraction data work in the scope of our sample size and duration and yield reasonable models for emotion prediction? Which contextual information does work and which does not?

a) Does a trained predictive model possibly yield better results than pure facial extraction?

RQ 3 Can affective emotion changes be discovered by a classifier?

5 Technical Implementation

This chapter describes the technical implementation used in the following data collection study. Based on the concept in Chapter 4 the mobile application *OpenFaceAndroid* is developed integrating all necessary Application Programming Interfaces (APIs) for data assembling. Its capabilities and also emerging limitations are pictured in detail.

5.1 Operating System and Device Selection

As operating system for the mobile application Google's *Android* is chosen. Its market share of about 85 % [Cor17] allows a high range of possible study participants and later on a large application field. Its support for developing applications enables implementing necessary functionalities for this thesis, such as machine learning techniques.

Nowadays, smartphones are capable of intensive computation, equipped with high-resolution cameras (both front and back), thus well-suiting the requirements stated in Section 4.2.

Due to two technical limitations, smartphones for this study are required to be capable of Android version 5.0 or higher. Processing and data storing have to stop for the time the user opens another application using the camera (for example when taking a picture or video) to avoid crashes, as two processes cannot access the camera device at the same time. Android 5.0 introduces a new camera API *camera2*[1] being able to detect other applications on the smartphone using any camera. In addition, the currently open application has to be logged. Android's *UserStatsManager*[2] allows collecting this data for all devices using Android 5.0 or higher without any compatibility issues.

As the background job is very power and battery consuming, it requires current mid to high-end smartphones to perform properly (see also sections 5.4, 6.1.3, and 6.3.4), which is compliant to the OS version requirements since current smartphones tend to have latest OS versions installed [Net17].

[1] https://developer.android.com/reference/android/hardware/camera2/package-summary.html
[2] https://developer.android.com/reference/android/app/usage/UsageStatsManager.html

5.2 Android Application

The application is solely written in Java (for all classes) and Extensible Markup Language (XML) for the layout. Android Studio[3] is used as Integrated Development Environment (IDE). As the application is based on the framework *OpenFace++* it is accordingly named *OpenFaceAndroid*, indicating the target OS. There are three Graphical User Interfaces (GUIs) the user is confronted with, two Android Activities and a notification. The first activity shown in Figure 5.1 is the main screen of the application which allows to start and stop the background process for the study. A small info box displays whether the background process is running ("Cam Status"), the study's starting date and time, and a unique participant identifier (ID) ("Study ID"). The ID is created once when initially opening the application and persists throughout both user studies. The entire data package stored on a user's device is marked with this ID which is later on carried through the data processing and machine learning step. For the second study, each participant thus can receive their individually trained classifier model identified by their unique ID.

To make sure the application is not stopped through the GUI by the participant, a password is asked for when tapping "Stop Study" (Figure 5.2).

When the background process is running, a notification (Figure 5.3) pops up in a set time interval of 15 minutes after the user has submitted a rating, asking the user to rate their current emotions.

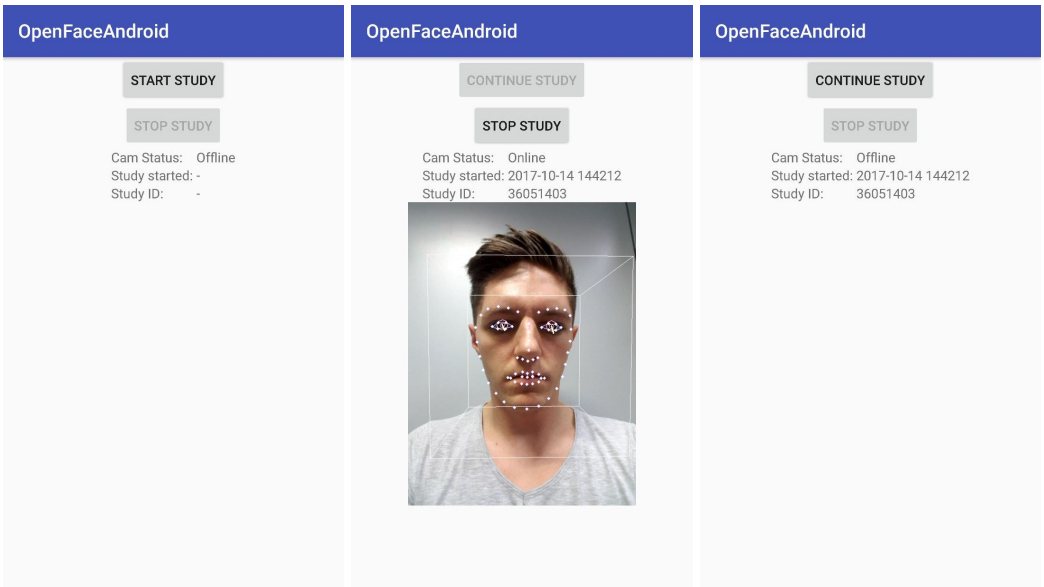
The second activity shown in Figure 5.4 is the emotion rating screen and accessible by tapping on the notification. All emotions yielded by *OpenFace++* can be rated in this screen. Neutral is included as the "default" emotion if none of the other six is experienced at that moment. Single as well as multiple selection (see Chapter 4) is allowed in this activity.

Android Broadcasts ensure the process keeps running throughout the study. Broadcasts include but are not limited to allowing to receive certain action triggers by events such as the device is booted or the screen is turned on or off [Goo17c]. Smartphone booting is monitored by a broadcast and triggers to start *OpenFaceAndroid*. When the smartphone screen is turned on, the background process initializes the front camera and video feed is processed. As soon as it's turned off, the camera stops recording.

OpenFaceAndroid requires four permissions in order to work [Goo17f]:

- android.permission.CAMERA: Camera permission is necessary to access the front camera device and its video feed.
- android.permission.RECEIVE_BOOT_COMPLETED: Allows to receive the boot completed event happening when the smartphone has booted. This is necessary for restarting the application in case the smartphone is turned off for example when the user shuts it down or the battery is empty.

[3] https://developer.android.com/studio/index.html



(a) Initial screen **(b)** Screen after study start **(c)** Screen after application stop

Figure 5.1: The main screen in its initial state when opening the application for the first time (5.1a). As soon as the study is started and the background service running "Cam Status" is set to "online" (5.1b) for easy identification if the user is asked to check the status. An imagebox displays the video feed of the front camera processed by *OpenFace++* for initial examination if both camera and framework are working on the smartphone. This imagebox is invisible once the main window is closed. If the application is stopped it can be easily identified by changed button appearances and restarted by tapping "Continue Study" (Figure 5.1c).

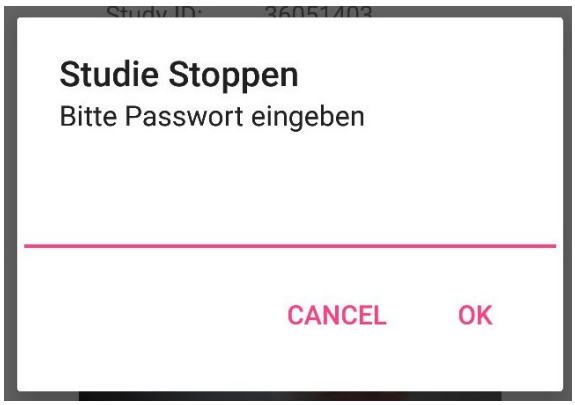


Figure 5.2: A password is asked for when tapping "Stop Study".

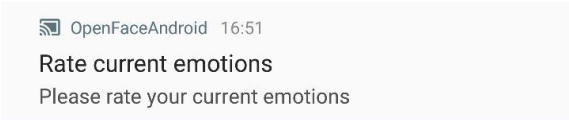


Figure 5.3: This notification pops up in a time span of 15 minutes after the last submitted user rating and can be tapped in order to access the emotion rating activity shown in Figure 5.4. Alternatively, in case of an inappropriate moment, it can also be dismissed by swiping it away or kept for later access.

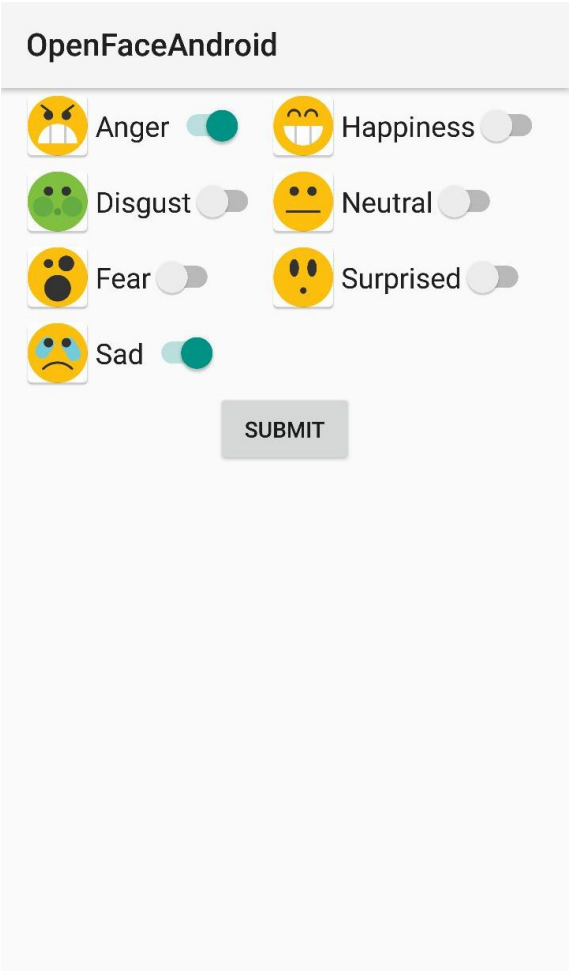
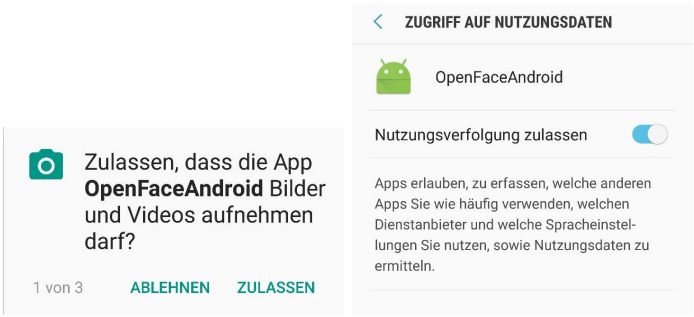


Figure 5.4: The emotion rating dialog allows users to rate their current emotions by tapping on the respective icon, text, or switch. In this case, “Anger” and “Sadness” are selected. Neutral works as the default emotion. The information is stored by clicking “Submit”. The dialog is automatically closed afterward.



(a) Initial screen **(b)** Screen after study start

Figure 5.5: For all permissions except usage statistics, the default permission dialog is shown (Figure 5.5a). To be able to evaluate usage statistics, the user must set this permission separately in the smartphone settings. (Figure 5.5b).

- android.permission.WRITE_EXTERNAL_STORAGE: Required to read and write files outside the application-specific directories. As *OpenFaceAndroid* directly stores data in the primary storage directory this permission is used.
- android.permission.PACKAGE_USAGE_STATS: Allows to evaluate usage statistics. This is the only permission that needs to be set in *Settings application* instead of the usual permission dialog (Figure 5.5).
- com.google.android.gms.permission.ACTIVITY_RECOGNITION: Necessary for using the Google Activity API to receive activity states such as running and walking.
- android.permission.ACCESS_FINE_LOCATION and android.permission.ACCESS_COARSE_LOCATION: Allow to access both Global Positioning System (GPS) as well as network location data.

The application uses internet connection to query weather data. However, with the introduction of Android 6, internet access is automatically given and not necessary to be set by permission anymore [Goo17g].

5.3 Integration of APIs and Logged Variables

In order to log all data required, several APIs have to be integrated to work with *OpenFaceAndroid* within a background job. The most important ones are listed as follows.

5.3.1 OpenFace++

OpenFace++ is designed to work cross-platform. For Android, there is a separate fork available hosted on GitHub[4]. This fork is taken and modified to work with Android's camera API *camera2* and the front camera of a user's smartphone.

OpenFace++ requires image data to be forwarded as raw byte arrays in NV21 (YUV) format. This is Android's standard format for camera preview images used in the initial camera API[5] [Goo17e]. In camera2 images are by default in a multi-plane YUV 420 format and can be transformed to NV21 by converting the respective planes to a raw byte array. Front camera images are provided in landscape mode. To allow correct computation of *OpenFace++* emotions, when used in portrait mode, image rotation by 90 degrees has to be applied in the conversion process. This can easily be done by aligning array positions accordingly.

Furthermore, the frame rate and image processing frequency is adjusted. Eventually, the frame rate is set to circa one FPS (depending on the phone's computation time) as a trade-off between power consumption and number of generated data instances.

5.3.2 Google Play Services

Google Play services API allows the integration of Google Play services and receiving data they provide required for the studies of this thesis.

Activity Recognition

Google's *Activity Recognition* API provides a list of possible user activities gathered from smartphone sensor data. It contains eight different types of activities expressed as confidence scores. Table 5.1 shows the activities and their explanations:

The recognition mechanism uses solely low power sensors and reads their data at regular intervals for little battery usage, which is important in the context of the upcoming study (see also Section 5.4). Activity detection is performed irregularly as it depends on the smartphones current state and samples from sensors. The API only allows passing parameters for the preferred minimum boundary of detection frequency [Goo17a; Goo17d]. In *OpenFaceAndroid* the highest frequency possible is used.

[4] https://github.com/cyberjunk/OpenFace
[5] https://developer.android.com/reference/android/hardware/Camera.html

Activity	Description
In Vehicle	The device is in a vehicle, such as a car.
On Bicycle	The device is on a bicycle.
On Foot	The device is on a user who is walking or running.
Running	The device is on a user who is running.
Still	The device is still (not moving).
Tilting	The device angle relative to gravity changed significantly.
Unknown	Unable to detect the current activity.
Walking	The device is on a user who is walking.

Table 5.1: List of activity types Google Activities can detect. Activities are returned with a confidence score between 0 and 100. As in the context of this application, activities are only monitored when the device screen is turned on, it is assumed that the device is always on a user [Goo17d]

.

Location Awareness

Google Play services location APIs allow identifying the last known location of a device without actively enabling its GPS sensor. They use all possible location providers of a smartphone (e.g. GPS and network information) while keeping use of battery power optimized. In context of the upcoming studies and the requirement to keep battery consumption as low as possible, this method is used.

5.3.3 Yahoo Weather

For annotation of weather data, *Yahoo Weather API*[6] is chosen due to its open source availability and sufficient reliability. It uses Rich Site Summary (RSS) feed allowing to receive weather information by HTTP GET requests. Weather can be queried with geolocation data originating from Google Play services (see section above). The result is detailed information about current weather (such as temperature, pressure, wind speed), astronomy and atmosphere details, three-day forecast, and the location this information is taken from. YWeatherGetter4a[7] is used as the corresponding wrapper for Android.

5.3.4 Collecting and Storing Data

Video feed from the front camera gets forwarded to *OpenFace++*. Data, in general, is only stored when a face is detected by *OpenFace++* and thus emotion values derived.

[6] https://developer.yahoo.com/weather/
[7] https://github.com/zh-wang/YWeatherGetter4a

Data privacy is an existent issue for smartphone applications. Participants should have as little concerns as possible taking part in the upcoming studies. Having a logging rate of lower-equal to one instance per second resulted in a data volume lower than 10 megabytes (MB) after a one week test period. Such factors allow storing data directly and entirely on the smartphone. Hence in both studies, all information is logged locally on the respective device storage without any data exchange to remote servers. Data is stored in a Comma-Separated values (CSV) format due to its ease of use and wide distribution [Gro17]. It is shared over different files and for evaluation personally collected at the end of the study. The separation in multiple files takes place according to which information being stored:

1. Main file for all directly accessible values shown in Listing 5.1. These values are continuously available and do not require waiting for a reply from an additional background service:

   - *OpenFace++* variables: anger, disgust, fear, happiness, sadness, surprise, neutral, attention

   - latest known position: longitude and latitude

   - currently active foreground application

2. Google Activity variables, reported back occasionally.:

   - probabilities: in vehicle, on bicycle, on foot, running, still, tilting, unknown, walking

   The respective CSV file is shown in Listing 5.2.

3. Weather information, queried each hour:

   - weather description, e.g. “Mostly Cloudy”

   - weather information location: country code, postal code, locality, sub-locality, street, house number

4. User rating file containing all emotion ratings (see Listing 5.3):

   - anger, disgust, fear, happiness, sadness, surprise, neutral - either 0 or 1

5. Debug file:

   A pure informative file containing certain debug information in case issues arise during study. Function calls and their respective parameters are logged in this file. Each day a separate log file for each category is created.

Each data entry within the files is extended with a timestamp to allow joining the data afterward. For convenience within German-speaking world, a semicolon (“;”) is used as separation character instead of a comma (“,”) as commas appear in all floating point values when parsing data later on.

Listing 5.1 Exemplary CSV file for main data. The initial timestamp located in all files is followed by *OpenFace++* results from front camera. They range between 0 and 1 for each emotion. An attention score expresses how directly the eyes gaze to the front camera. The currently used application as well as longitude and latitude are also stored in this file.

```
Time;Anger;Disgust;Fear;Happiness;Neutral;Sadness;Surprise;Attention;app;long;lat
2017-08-04 180657;0.016100273;0.034954034;0.27402997;0.0927334;0.80558246;0.0151441125;
    0.0012901387;0.726638;com.instagram.android;9.3376905;48.7846016
2017-08-04 180658;0.05515717;0.063931465;0.22971965;0.14125969;0.7955613;0.0;0.035134986;
     0.589438;com.instagram.android;9.3376905;48.7846016
2017-08-04 180659;0.04955804;0.07453113;0.27852213;0.107499406;0.7082434;
    0.018113915;0.032722205;0.6546314;com.instagram.android;9.3376905;48.7846016
```

Listing 5.2 Exemplary CSV file for Google Activities. The timestamp is followed by confidence scores for all detectable activities ranging from 0 to 100.

```
Time;STILL;ON_FOOT;WALKING;RUNNING;ON_BICYCLE;IN_VEHICLE;TILTING;UNKNOWN
2017-08-05 013942;100;0;0;0;0;0;0;0
2017-08-05 014223;100;0;0;0;0;0;0;0
```

5.4 Technical Limitations

During test phase (and later on also in the study, see Section 6.1.3 and 6.3.4), different technical problems were encountered. These included distinct behavior of diverse Android smartphones and hardware limitations in terms of battery power.

5.4.1 Smartphone OS characteristics

While a few smartphone manufacturers run stock android on their devices, many have their own Android OS implementation.

During testing, lots of possible participants' smartphones being incompatible with the application were discovered. Custom manufacturer's battery and application monitoring software terminate the background service, which can not be intercepted or avoided without acquiring root access.

In particular "Samsung Galaxy" devices do not show this behavior, thus becoming preferred in participant selection.

Table 5.2 shows a list of all smartphones tested, their respective Android version, and further notes. One of two Huawei P9 devices suspends the background service irregularly.

Listing 5.3 Exemplary CSV file for user ratings. In this case, the timestamp is followed by the user rating of each emotion (0 for not selected, 1 for selected).

```
Time;Anger;Disgust;Fear;Happiness;Neutral;Sad;Surprised
2017-08-04 180438;0;0;0;0;1;0;0
2017-08-04 182035;0;0;0;0;1;0;0
```

Smartphone	OS	Application works	Notes
Huawei P9	Android 6	(yes)	Background service is partially terminated
Google Nexus 5X	Android 5	yes	-
Google OnePlus 5	Android 6	no	Background service is terminated
Samsung Galaxy J5	Android 6	yes	-
Samsung Galaxy S5 mini	Android 6	yes	-
Samsung Galaxy S6 (Edge) (Plus)	Android 6	yes	-
Samsung Galaxy S7	Android 6	yes	-
Sony Experia E5	Android 6	no	Camera2 API not supported
Sony Experia Z3	Android 6	yes	-

Table 5.2: This is the list of devices tested before the study conduct. Devices marked with "yes" work without problems, "(yes)" indicates problems with a part of devices. Phones marked with "no" are not compatible with the application. While for Galaxy S5 only the "Mini" version is tested, the record "Galaxy S6" includes its variants "Edge" and "Plus".

The other device does not show any issues. While Google's Nexus 5X works fine, the more recent OnePlus 5 suspends the background service after a few seconds when the screen is turned off. Sony's Experia E5 does not support Android's camera2 API, thus does not receive image feed from the front camera. The more recent Z3 is compatible with the API. Notably, all Samsung devices from Galaxy S5 mini up to S7 show no issues regarding the implementation.

5.4.2 Battery consumption

There are multiple reasons why battery drain is higher when using the mobile application for the study. The front camera is continuously recording whenever the screen is turned on, forwarding picture data to *OpenFace++*. *OpenFace++* then computes emotion values for each picture as well, requiring CPU and thus battery power. The mentioned contextual information being logged triggers sensors (activity recognition and location data) and data exchange to external services (weather). All information is continuously stored on the smartphone.

Each component of the application runs within the background service. Usual daily use of smartphones with possibly highly power consuming applications such as mobile games

or social media applications together with the expensive background service affect power consumption.

This leads to specific choices and trade-offs collecting information, summarized as follows:

- Battery consumption highly increases together with raising FPS rate of the camera and thus computation for each frame when processed by *OpenFace++*. One frame per second is a compromise generating sufficient data instances and keeping battery consumption acceptable.
- Every sensor activated increases battery drain. Requesting geolocation passively, for instance without activating GPS sensors, yields as little consumption as possible.
- Weather is supposed to change over a longer time period than sensor data mentioned above. For the study, weather data is thus requested every hour and considered having no relevant impact on battery drain.
- Testing showed that Google activity recognition has the least effect on battery consumption. In this case, the highest request rate is selected.

For more details on the outcome of the implementation see the respective section 6.1 and 6.3.

6 Data Collection and Model Prediction Studies

In this chapter, the preparation, demographics, procedure, and evaluation of the data collection and subsequent model prediction study are described. This is followed a detailed look at processing and information yielding of the data for later evaluation in Chapter 7.

In the first study, emotions derived from facial expression data by *OpenFace++* and contextual data are collected from daily used smartphones. Resulting data is evaluated, processed, and used to train a classifier. In the second study, the most suiting contextual data together with *OpenFace++* emotion values and the classifier prediction are logged in order to compare predictive results against *OpenFace++*. Furthermore, classifier data raises user interaction under certain conditions to find out whether affective mood changes can be predicted by the classifier.

6.1 Data Collection Study

The developed implementation *OpenFaceAndroid* explained in Chapter 5 is used on smartphones during this study. The same application is installed and started on each user's smartphone. In this study, *OpenFaceAndroid* collects *OpenFace++* and contextual data continuously in the background during every participant's daily use of their smartphone whenever the screen is turned on and stopped when the screen is turned off. Training data is appended by the users rating grasped from the pop-up notification dialog. Every 15 minutes the user is asked to rate their current emotions. Each emotion can be marked with *yes* or *no* depending on whether the user felt the respective emotion at that time, also allowing multiple selection. After a week, the application is stopped and the logged data is collected from each participant, followed by an interview (see Appendix A, Figure A.2).

6.1.1 Preparation and Study Conduct

Participants for this study are acquired by sending a "call for students" mail through the university's mailing list of computer science and software engineering, Facebook postings, and word-of-mouth advertising. There are no further restrictions except for possessing a mid-high end Android smartphone as explained in Chapter 4 and 5.

In order to ensure similar understanding and conduct of the study for all participants, a standard procedure is followed:

Each participant initially fills in a consent form to ensure their agreement in taking part in the study (see Appendix A, Figure A.1). The form attests that no risks and obligations would arise when participating. Brief information about the general goal and idea of the studies and the data being collected is listed in the consent form and explained in detail orally after that. As participation in the subsequent second study is only possible for those who take part in the first one, everyone is also asked if they are willing to participate in the second study as well.

Individual demographics information - for instance age, gender, and occupation - are noted down for later evaluation.

Each participant receives a detailed explanation about the data being collected. It contains information about how this data is processed and stored - especially indicating that neither pictures are stored at any time, nor any online data exchange takes place. Additionally, examples of produced log files are shown.

Further processing of the collected data as preparation for the second study is briefly outlined.

The installation file of *OpenFaceAndroid* is transferred to each smartphone and executed. General interaction with the application (for example start and - in case of an error - continue the application) and user-specific emotion rating within the graphical user interface dialog is shown to and once performed by each participant as a tutorial step. Furthermore, logging and notification frequency are explained. It is pointed out that the emotion dialog should be answered whenever possible to generate a maximum number of data instances.

Technical details about the application including possibly higher battery drain, low mobile data traffic evoked by requesting weather information, and triggers for data capturing (display screen is turned off / on) are discussed. As *OpenFaceAndroid* can be suspended through Android's task manager and application monitor, guidelines to avoid possible shut downs of the application due to user or operating system triggers complete the participant's instruction. To assure no external issue will interfere with the application possibly stopping it, all participants are reminded every two days during the study to check whether the application is still actively running, by sending a generic Whatsapp message. For convenience, the main view of the application displays this information for the user. The participants are explicitly asked to use their smartphone as they always do.

Directly following-up the installation and instruction the study is started and each participant uses the application continuously for the study duration of one week in the background whenever interacting with their smartphone.

After a week the application is stopped and uninstalled from each participant's smartphone as well as their generated data collected. Device type, battery consumption, storage, and memory usage as well as mobile data traffic are noted down as additional technical information provided by Android's *Application Info* screen. The participants are surveyed in a semi-structured interview afterward.

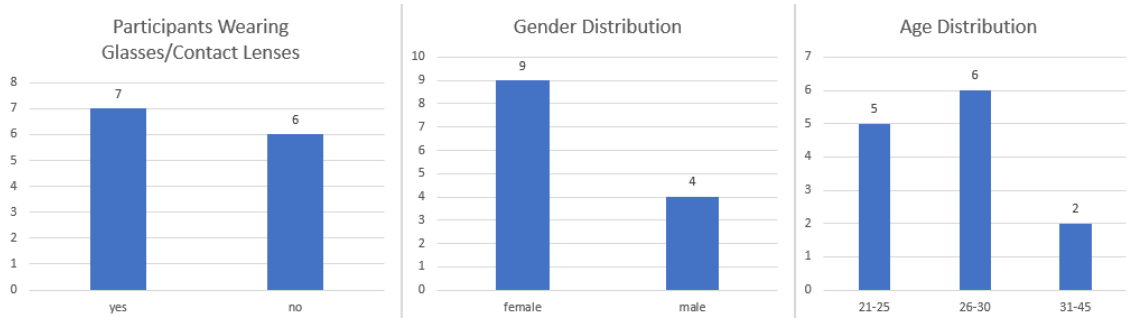


Figure 6.1: Demographics about participants of the first study: Optical aid, age, and gender distribution

A semi-structured interview completes each study process. In both studies, participants are asked about their subjective perception of the application in terms of battery usage and performance impact. Optionally they can tell about occurrences of special events that had an unusual effect on their emotions such as a birthday or a dispute. If answered, date and time as well as the respective emotions and optional comments are noted down. The actual event is intentionally left out for privacy reasons.

Eventually, each participant receives a compensation of 10 EUR for their contribution.

The first study took place in between August, 2nd and August, 21st 2017. Participants were split into two groups, one starting in calendar week 31, the other one in 32.

6.1.2 Participants and Devices

The first study originally consisted of 14 participants - one dropped out early due to technical issues caused by the device battery (see next section).

Out of the remaining 13 persons, over two third were female and one third male. The age varied between 21 and 45 (see Figure 6.1), mainly ranging from 21 to 25 and 26 to 30 while two of them were above 30. The median age was 26, the standard deviation amounts to 5.942. Nine participants were students, one in their apprenticeship, and three employees.

6.1.3 Interview Outcome and Feedback

Most participants explained that they experienced a higher battery drain than usual during the study. Figure 6.2 shows that more than two third of the participants state a subjectively higher and almost one third a much higher battery usage than normal. During the interview, seven participants stated that due to the higher battery drain recharging the phone more often than usual was necessary. The higher use of Central Processing Unit (CPU) power also increased the smartphone temperature for four users. One participant expressed their concerns as the smartphone subjectively felt very hot and this was not experienced

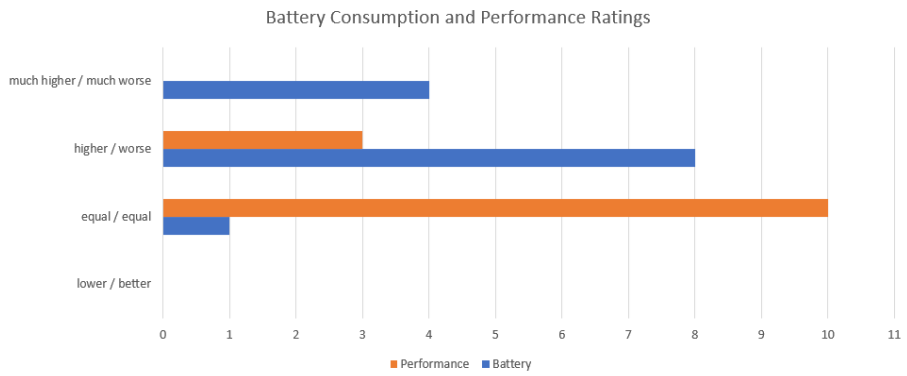


Figure 6.2: Number of ratings in categories battery consumption (left-handed terms going from lower to much higher) and performance (right-handed terms going from better to much worse).

before. For another user, Android OS signaled a warning and stopped recharging for safety reasons. This is also reflected by the application statistics directly obtained after the study. According to these, battery consumption ranged between 2.5% up to 34% of the entire battery capacity, with an average value of 11.04% and a standard deviation of 7.85. Memory usage was between 111 MB and 215 MB, with an average of 167.92 MB and a standard deviation of 35.46. Mobile traffic was for all participants below one megabyte.

This is captured and bore in mind for the second study implementation. Battery life is a vital aspect of any mobile application.

Opposed to that performance was for the greater part unaffected. Only three users noted a worse performance than usual, resulting in slower responding applications and input lag on touch gestures. This happened exactly on those devices with less CPU power. Notably, no user rated the performance as much worse while using the application.

Special occasions with unusual emotional states were reported by six participants including the emotions anger, happiness, and sadness.

Technical issues arose for two participants. On one Samsung Galaxy S7 device warning messages popped up making aware that the phone became too hot and thus could not be recharged until cool down anymore. The user suggested that weather conditions with an outside temperature of above 30 degrees Celsius and constant use of a highly power consuming mobile game might have led to this warning. Together with the computing-intense *OpenFaceAndroid* background application, this message came up for them for the very first time.

Another participant with a Sony Experia Z3 faced the problem of their smartphone battery draining unusually and more quickly than for any other participant. They reported an operating time of about an hour before the battery was completely empty and the phone had to be recharged. The participant assumed their rather old device and the battery being not fully functional anymore to cause these irregularities. This user had to discontinue participation and thus could not take part in the second study.

Type	Time Span
Rating	+/- 15 minutes (*)
Google Activity	+/- 3 seconds
Weather Information	+/- 30 minutes

Table 6.1: Selected time spans for data annotation. User rating is set to the same value as the initial time frame, while Google Activity and weather information is set to half of the value.

6.2 Data Processing and Machine Learning

Between the first and second study, results from the first run are used to build and train a classifier for eventual mood prediction. Therefore data has to be analyzed, appropriate variables picked, and parsed in a suitable format. Afterward, parsed and transformed data is used to train the classifier. The result is a model data later on being used within the application of the second study.

6.2.1 Parsing and Analyzing Data

For parsing data, at first, all files have to be joined by their timestamp for each participant separately. Directly accessible variables are logged continuously and have to be annotated with the closest data instance of context and user rating data records. Therefore certain time spans have to be defined, allowing to join data even if instances are not marked with the exact same timestamp. For best results, time spans to join within are kept as low as possible. Thus half of the shortest time frame between two data sets of each file is taken, except for user rating data which is set to the same value as the time frame (15 minutes) to produce more data records for training the predictive classifier. Google Activity report times vary strongly, ranging from a few seconds in between two instances up to minutes. Weather data was collected at most each hour (or as soon as the screen was turned on again), while user rating was performed each 15 minutes maximum (in case the user always rated as soon as the notification showed up). This resulted in time spans listed in Table 6.1. In case of multiple matches when joining data overlapping in these time spans, the closest nearby instance was taken. Additional information files for better analysis were created during parsing.

Joining itself was performed using small Java and Microsoft Visual C# programs. Both offer index structures enabling quick parsing and combination of data. As IDE Eclipse[1] for Java and Microsoft Visual Studio for C#[2] was used.

[1] https://www.eclipse.org/
[2] https://www.visualstudio.com/

	Anger	Disgust	Fear	Happiness	Neutral	Sad	Surprise
Anger	-			x			
Disgust	-	-		x			
Fear	-	-	-	x			
Happiness	-	-	-	-		x	
Neutral	-	-	-	-	-		
Sad	-	-	-	-	-		
Surprise	-	-	-	-	-	-	

Table 6.2: All emotions excluding themselves are marked with an 'x'. If these ratings occur less frequently than three times, the rating is assumed to be accidental and marked as invalid. Otherwise, it's assumed to be intended.

Next, combined results have to be analyzed and several different combinations yielded from them.

6.2.2 Results

In order to train individual classifiers for the second study, the data generated in the first study has to be evaluated for its usefulness. In sum over 300000 data instances were collected during the week of which in total nearly 120000 were annotated with a participant's emotion rating. Only these instances can be taken to train the classifier.

Data cleaning is performed in two steps. If the attention value (explained in Section 2.2) of a record is below 0.1, the record is removed. Furthermore, to avoid false or accidental emotion ratings, an exclusion matrix for possibly incompatible emotions is defined aiming at identifying those (see Table 6.2).

If such ratings occur within a defined boundary, these data instances are dismissed from evaluations. In total 315 data records were removed because of too low attention score and 1102 data records due to invalid ratings.

Taking a deeper look into the individual study data, one participant was detected who only rated two emotions: happy and neutral. As only those emotions the participant at least rated once can be used for training the classifier and later on predicted in the second study the minimum number of rated emotions was restricted to happiness, neutral, and either sadness or anger. Thus the mentioned participant could not take part in the second study.

Further steps are only performed on data of the remaining 12 participants. As illustrated in Table 6.3 less than half of the users rated emotions disgust and fear, about two third rated sadness and surprise, and every participant rated neutral and happiness at least once. Only one user did not rate anger during the first study. Looking at the number of data records annotated with these ratings reveals that neutral is by far the most frequent rated emotion. This indicates that participants sense neutral as a default rating if none of the six emotions

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Users	11	5	5	12	12	8	9
Records	11177	1942	4708	45223	74700	15604	1760

Table 6.3: This table shows the number of participants that rated the respective emotion at least once as well as the total number of data records annotated with the respective rated emotion.

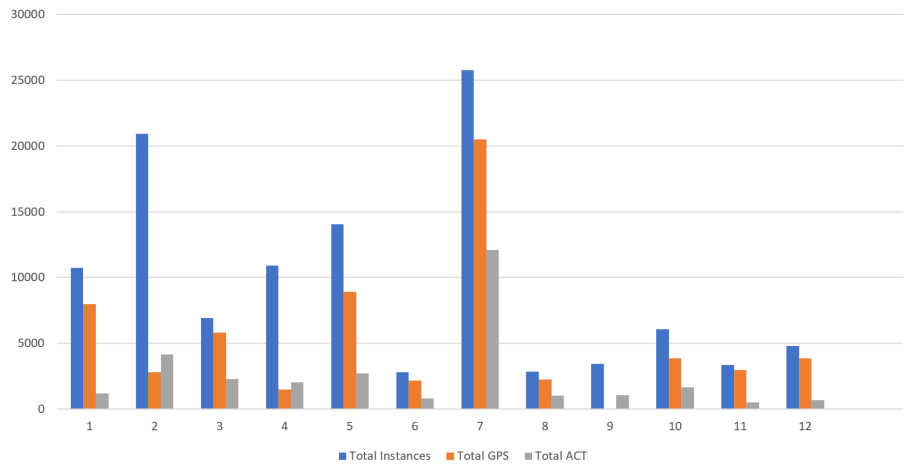


Figure 6.3: Total number of instances generated during study per participant in blue opposed to the number of instances annotated with GPS data in orange and with Google Activity in gray.

is really experienced at that moment. In an interview after the second study, participants are asked whether they feel that way. The second most rated emotion is happiness. There are above 10000 records annotated with sadness and anger while the number of records with disgust, fear, and surprise is for each less than 5000.

Analyzing the annotated data revealed that both Google Activity and passive geolocation data was strongly limiting the number of data instances that could be taken for the classifier. Figure 6.3 shows that both Google Activity and passive GPS are varying strongly amongst participants. GPS data annotation ranges from 13.3% for user 2 up to 88.6% for user 11. User 9 even generated no GPS data at all.

Both Google Activity and passive geolocation are not steady data resources but report their information back unpredictably. For the second study best case for contextual data is to be constantly available in order to always be able to yield predictions from the classifier. Furthermore, a highest possible participant number for the second study is desired. Users with a low percentage of annotated data would not be able to take part if this data is considered for classification. Hence the most suitable and steadily available contextual information is decided upon which turns out to be the currently open smartphone application.

Remarkably, special events reported in the interview had no measurable diversifying effect in the logged data, neither for positive nor negative emotions. This allows the supposition that emotion analysis can be performed independently of individual events.

6.2.3 Classifier

Training the classifier and deriving a predictive model is achieved with Waikato Environment for Knowledge Analysis (Weka)[3]. It is freely available and offers seamless integration for Android as it is fully implemented in Java.

Each participant's cleaned study data was converted to an artificial file [Uni17a]. Therefore joined CSV data was simply transformed in Artificial File (ARFF) notation, recasting and categorizing attributes, and reordering data records. In Weka four non-relational attribute types are supported: Numeric containing both integer and floating point values (real), nominal, string, and date [Uni17b]. This leads to the following mapping:

1. Emotion probabilities are turned from float values into real.
2. Currently open application and rated emotion as string attributes are categorized as nominal. This is necessary as certain classifier methods in Weka such as SVM can only work upon nominal data. Therefore each participant data has to be scanned a priori and lists of all applications appearing in their data as well as of all emotions rated have to be created individually. These lists serve as nominal value ranges data records can reside in.

These emotions never rated by the user were also deleted from the data *OpenFace++* generated. While emotion probabilities and application served as attributes, emotion ratings are used as class values. In case the user rated multiple emotions at once, the one with the higher *OpenFace++* value was used as class value.

As classifying algorithm both Weka's *Random Forest* implementation and *libSVM*[CL16] as SVM classifier were tested. As test option k-fold cross-validation was used containing nine subsamples as training and one subsample as validation data for each run. For Random Forest the number of iterations was set to 100 and seed to one. For SVM the same seed was used. Data of the first study yielded an average of ten percent higher accuracy using Random Forest over SVM, leading to this classification method as first choice. Random Forest uses multiple random arrangements of decision trees for classification, generated during training phase [BC17]. Compared to standard decision trees, the effect of overfitting is dampened by this random generation [Bre01; HTF09]. Random forests provide amongst other properties fast and efficient model training and evaluation.

Models were automatically generated in small Java programs using Weka library. The exact parameters used for classifier training are noted in Table 6.4. Therefore each ARFF file was parsed and the respective model stored under the participant's unique ID. Additionally classifier statistics such as the amount of correctly and incorrectly classified instances,

[3] https://www.cs.waikato.ac.nz/ml/weka/

Parameter	Value
Size of each bag, as a percentage of the training set size	100
Number of iterations	100
minimum variance for split	0.001
Seed for random number generator	1
Maximum depth of the tree	unlimited
Cross-validation folds	10

Table 6.4: Each Weka interface parameter and respective value which was set and used for classifier training.

Listing 6.1 In the header (line 1 to 9) attributes and their types are defined. Floating point attributes are turned to real. The nominal attribute "app" is indicated by "{}" surrounding all values they can take. For better visualization, only two applications are listed in this example. The class attribute, also indicated by surrounding curly braces, is in this case without emotion "fear" as this was not rated by the respective user. This is followed by all data records in the body starting at line 11. The order of data record values is similar to the order of attributes defined in the header.

```
1      @RELATION emotion
2
3      @ATTRIBUTE pbAnger      REAL
4      @ATTRIBUTE pbHappiness REAL
5      @ATTRIBUTE pbNeutral REAL
6      @ATTRIBUTE pbSad        REAL
7      @ATTRIBUTE pbSurprised REAL
8      @ATTRIBUTE app {com.sec.android.app.launcher,com.whatsapp,...}
9      @ATTRIBUTE class  {User_Anger,User_Happiness,User_Neutral,User_Sad,User_Surprised}
10
11     @data
12     0.1050642,0.1242275,0.4511586,0.2948778,0.103941,com.whatsapp,User_Neutral
13     0.1417169,0.1799573,0.1876743,0.3572525,0.2258321,com.whatsapp,User_Neutral
14     0.06777444,0.0698347,0.328446,0.435216,0.2575057,com.whatsapp,User_Neutral
15     ...
```

accuracy rates (precision, recall, and f-measure amongst others) by class, and the confusion matrix were stored in separate files under the respective ID. To automatically integrate the models into Android later on additionally a configuration file containing the ARFF header (see Listing 6.1) was created.

6.2.4 Training Model

Training the model yielded weighted F1 scores from 0.69 up to 0.97. Out of above 112000 data instances in sum, over 78% were classified correctly as indicated by Table 6.5.

Part. No.	Total	Correctly Cl.	Incorrectly Cl.	F-Measure Avg
1	10735	10394	341	0,9660
2	20909	16388	4521	0,7703
3	6905	5514	1391	0,7960
4	10922	9235	1687	0,8452
5	14049	10230	3819	0,7245
6	2777	2310	467	0,8281
7	25777	18121	7656	0,6929
8	2838	2512	326	0,8804
9	3425	2875	550	0,8302
10	6079	4324	1755	0,7082
11	3368	2788	580	0,8276
12	4775	4201	574	0,8762
Sum / Avg	112559	88892	23667	0,8121

Table 6.5: Total number of generated data records, the amount of correctly and incorrectly classified instances, and the average f-measure value per participant. On the bottom the sum of all data records and the average f-measure is listed.

A deeper look at the accuracy rates, in particular F1 scores, per emotion and participant gives further insight, visualized in Table 6.6.

Surprise performed worst with almost half of users who have the emotion in their data yielding zero accuracy, one with an F1 score below 0.1, two below 0.5, and a maximum of 0.85 for one user. It is followed by disgust with an average F1 score of 0.52. Anger and disgust perform better with scores close to 0.6, followed by sadness and happiness. Although the average score is higher, there are participants for which sadness performed second worse compared to all other emotions. Neutral yields the highest overall scores in average, minimum, and maximum values. This explains the high number of correctly classified instances, as neutral is predominate in ratings. The numbers furthermore tell that emotion accuracy is highly person dependent, best seen in emotions sadness and surprise where the difference between bottom and top F1 scores is above 0.6 and 0.8 respectively. Detailed classification results, accuracy information, and confusion matrices for each participant are given in Appendix A, Figures A.5 to A.16. The second study will reveal how accurate the models are when deployed in a real environment.

No.	Measure	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
	Precision	0.818	-	-	0.72	**0.978**	0.74	0
1	Recall	0.713	-	-	0.533	**0.991**	0.561	0
	F-Measure	0.762	-	-	0.613	**0.985**	0.638	0

2	Precision	0.76	**0.78**	**0.808**	0.705	0.812	*0.515*	0.308
	Recall	0.603	*0.209*	0.667	0.605	0.901	*0.146*	0.056
	F-Measure	0.673	*0.329*	0.73	0.652	0.854	*0.227*	0.094
3	Precision	0.722	0.707	-	0.809	0.699	**0.835**	**0.85**
	Recall	0.548	**0.651**	-	**0.859**	0.66	**0.855**	**0.776**
	F-Measure	0.623	0.678	-	0.833	0.679	**0.845**	**0.811**
4	Precision	**0.831**	-	-	0.823	0.869	-	*0*
	Recall	**0.771**	-	-	0.844	0.869	-	*0*
	F-Measure	**0.8**	-	-	0.834	0.869	-	*0*
5	Precision	0.672	0.768	-	0.774	*0.673*	0.667	*0*
	Recall	0.485	0.628	-	0.843	*0.659*	0.616	*0*
	F-Measure	0.563	**0.691**	-	0.807	*0.666*	0.641	*0*
6	Precision	*0.533*	-	-	0.783	0.855	-	-
	Recall	0.348	-	-	0.698	0.906	-	-
	F-Measure	0.421	-	-	0.738	0.88	-	-
7	Precision	0.55	-	*0.59*	0.653	0.722	0.736	*0*
	Recall	*0.264*	-	0.394	0.59	0.845	0.611	*0*
	F-Measure	*0.356*	-	0.473	0.62	0.779	0.668	*0*
8	Precision	0.559	-	-	*0.506*	0.928	-	-
	Recall	0.422	-	-	*0.465*	0.95	-	-
	F-Measure	0.481	-	-	*0.485*	0.939	-	-
9	Precision	0.68	0.692	0.6	**0.878**	0.878	-	0.579
	Recall	0.664	0.329	*0.261*	0.764	0.934	-	0.284
	F-Measure	0.672	0.446	*0.364*	0.817	0.905	-	0.382
10	Precision	0.628	*0.564*	0.734	0.64	0.757	0.764	0.438
	Recall	0.499	0.44	0.705	0.64	0.819	0.534	0.412
	F-Measure	0.556	0.494	0.719	0.64	0.787	0.628	0.424
11	Precision	-	-	0.741	0.552	0.84	0.831	-
	Recall	-	-	**0.76**	0.5	0.855	0.812	-
	F-Measure	-	-	**0.751**	0.525	0.847	0.822	-
12	Precision	0.737	-	-	0.832	0.916	0.721	0.707
	Recall	0.618	-	-	0.857	0.943	0.55	0.547
	F-Measure	0.672	-	-	**0.844**	0.93	0.624	0.617

Table 6.6: Classifier accuracy details per participant about the measures precision, recall, and F1-score listed per emotion. In case the emotion was not part of the model, it is marked with "-". For each measure, the maximum values are highlighted in blue and bold font, the minimum in red and cursive respectively.

6.3 Model Prediction Study

6.3.1 Android Application Changes

The individually trained classifiers shall be validated in the second study. Therefore the application is modified to automatically integrate the respective individual model file for each participant and to retrieve solely the contextual data selected from the results of the first study. Weather and Google Activity APIs are thus removed which also decreases workload, CPU, and battery usage of the application. The GUI remains nearly untouched compared to the initial application explained in Chapter 5. An additional configuration button is added to load the model and enable starting the background service. Within the emotion rating dialog, those emotions never rated by the user in the first study are not displayed and thus cannot be selected. Each data instance received from *OpenFace++* and annotated with contextual data is processed by the classifier, yielding probabilities for each basic emotion. Data is again continuously logged as in the first study.

Removing APIs yielding dismissed contextual data from the application also leads to a reduced need for permissions. Both Google Activity and location permissions mentioned in Section 5.2 have been removed in this implementation.

Referring to RQ 3 in Section 4.4 affective emotion changes assumed by the classifier shall invoke the mood questionnaire dialog to collate the user's emotions to the classifier prediction. Each emotion is provided with an internal counter. For each data instance, the highest predicted emotion's counter is incremented by one. As soon as the counter value reaches a certain boundary the mood questionnaire dialog appears, assuming this emotion is the most present one at that very moment. The respective emotion is marked as currently active and all counters are reset to zero. The next time a counter reaches the boundary, the respective emotion is compared against the last marked one. If it has not changed, counters are set to zero. If the emotion has changed an affective mood change is assumed and the dialog pops up to collate the user's perception. In this case, counters are also reset to zero.

As created data instances are directly connected to the fps rate used in the application computing a reasonable boundary for invoked predicted affective mood changes lead to the following formula:

$$boundary = fps * ts \tag{6.1}$$

where ts denotes the time span in seconds and fps the frame rate per second.

The fps rate for the second application is set to four. Testing different time spans show that a balance between fast enough recognition of a mood change against possible false positives using a too short time span is needed. A time span of 20 seconds, which is used for the second application, appears to hold this balance leading to the following boundary value:

$$boundary = fps * ts \tag{6.2}$$

$$boundary = 4 * \frac{1}{s} * 20s \tag{6.3}$$

$$boundary = 80 \tag{6.4}$$

To still have emotion ratings in case a mood change is never predicted by the classifier a fix frequency of 30 minutes is set on top evoking the dialog.

6.3.2 Preparation and General Instructions

Participants of the second study are a subset of those of the data collection study and directly contacted and asked for their participation. The standard procedure followed is most closely approximate as in the first study (Section 6.1.1). Additional diverging details within the study procedures are mentioned in the following sections. Another consent form has to be signed while the general topic and background of the study as well as demographics remain the same and thus are already common. Additional to the first study, the individually trained classifiers are explained to the participants. This time both the installation file similar for all participants as well as the individual model and configuration file are copied to each participants smartphone. The new behavior of the application, raising a notification when an emotion change is predicted and the possibly missing emotions to choose from (depending on the user's rating in the first study) are shown as part of the tutorial. Following this, the study is started and stopped after a week. Data is once more collected directly from each participant's phone and the application uninstalled. Again device type and technical information provided by Android's *Application Info* are noted down followed by a semi-structured interview. The interview contains similar questions about the subjective opinion of battery usage and performance as well as a detailed questionnaire about their general experience, privacy topics, and possible use cases of the application (see Appendix A, Figure A.4). This procedure is finished by compensating the participants with 10 EUR.

The second study was conducted between September, 7th and September, 19th 2017. Participants were split into two groups, one starting in calendar week 36, the other one in 38.

6.3.3 Participants

For the second study out of the original 13 participants during first study nine remained, reporting four drop-outs due to holidays, missing availability, and health issues. Gender distribution was almost equal this time while age variety was not heavily affected compared to the first study. In the range 31-45 one participant remained, 21-25 and 26-30 had the same distribution this time (see Figure 6.4). The median remained 26, the standard deviation changed to 3.083.

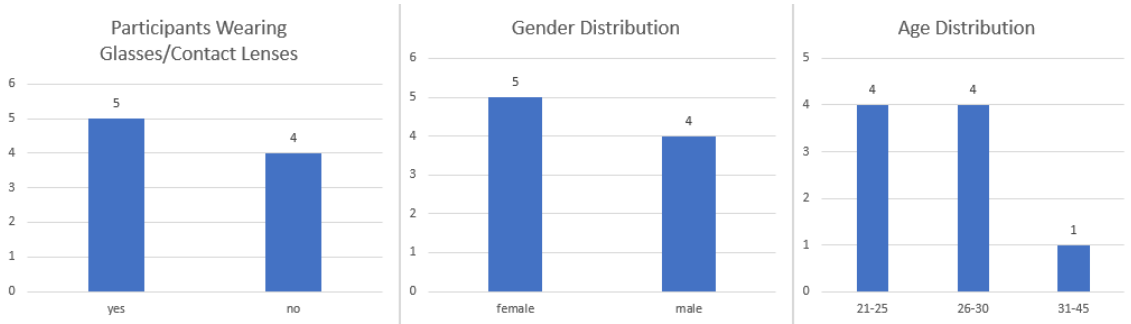


Figure 6.4: Demographics about participants of the second study: Optical aid, age, and gender distribution

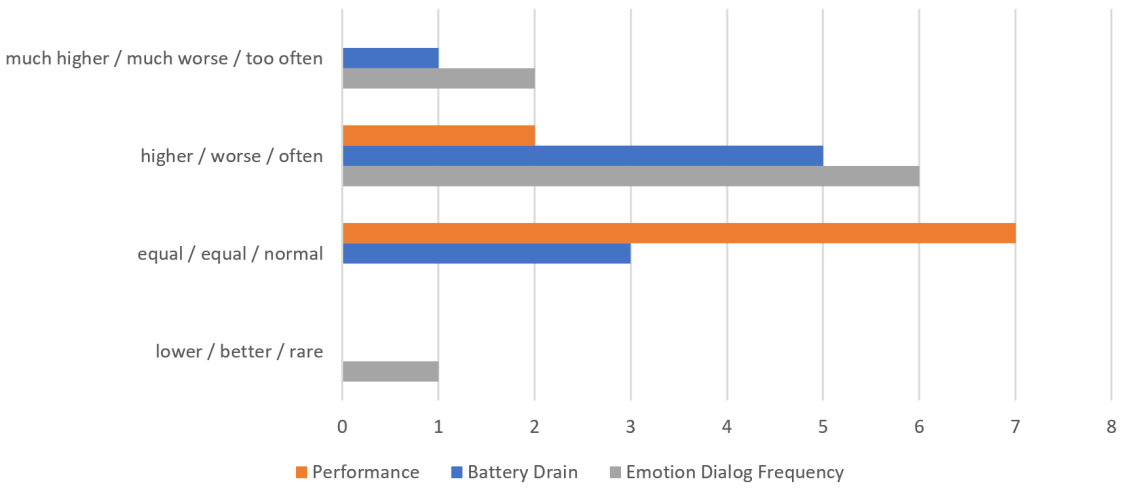


Figure 6.5: Number of users that rated the respective categories for battery consumption (left-handed terms going from lower to much higher), performance (middle terms going from better to much worse), and rating dialog frequency (right-handed terms going from rare to too often).

6.3.4 Interview Outcome and Feedback

The higher battery drain as in the first study remained in the feedback of the second study (Figure 6.5). However, seven out of nine participants reported their battery consumption had improved compared to the first study. Those users accounting temperature problems of their devices in the first study did not experience these issues anymore this time. One participant noticed differences in their smartphone performance between the first and second study and rated it this time as equal instead of worse. The other users rated it similar as in the first study. This is again reflected by the application statistics. Battery consumption was overall lower than in the first study, ranging from 7% up to 28%, with an average of 12.56% and a standard deviation of 7.17. Memory usage however changed from a range between 72 MB and up to 318 MB, the average being 163.44 MB and standard deviation 66.06 respectively.

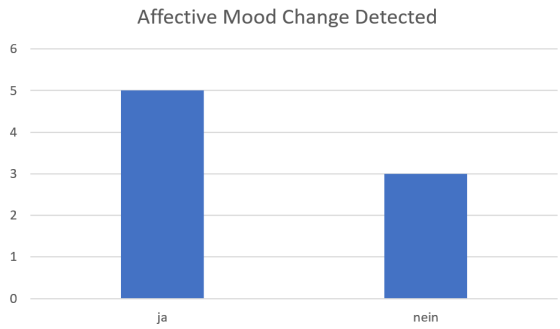


Figure 6.6: Number of users that detected an affective mood change close to the moment an emotion rating notification popped up.

Asking about their perception of the notification pop-up frequency for emotion rating, most participants agreed on a too high frequency. Only one user rated the frequency as rare, while three thought it is normal. Notably, almost two third of the users sensed an affective mood change at times the application asked to rate their emotions. (Figure 6.6).

One participant changed their device between the first and second study introducing a Samsung Galaxy S8 into the device pool, which was not tested during implementation phase. He reported occasional application issues displayed by the smartphone OS. This included infrequent messages that the process stopped and had to be restarted. He was still able to use the application and generated data instances. This issue was only observed on their very device, no other user was confronted with similar issues.

6.3.5 Data Processing

The first process step similar to the first study is joining continuously logged data with user ratings. Again this is done by using small Java and C# programs. Afterward, an in-depth analysis on joined data is performed.

During parsing it was discovered that one user, unfortunately, did not create any log data. It is presumed that this data was unintentionally deleted as parts of the files still existed. For data evaluation, this participant could not be taken into consideration, however their interview answers were still noted down. Out of the remaining eight participants, a total of above 64000 instances was logged of which more than 41000 were annotated with user ratings, yielding fractions from 35 up to 94 percent of rated data instances per participant. This time no false ratings according to the exclusion matrix (Table 6.2) occurred.

7 Evaluation

In this chapter results obtained from data processing within both studies are evaluated. Differences between results of the first and second study are discussed and correlations between emotions and applications are searched for. Comparisons between facial data extraction and classifier prediction are performed. With the gained insight of a detailed analysis of the results, the RQs formed in Chapter 4 are answered.

7.1 Emotions and Applications

For evaluation, each emotion in multiple selection ratings was counted individually. Consequently, in total unique emotion rating was performed 1144 times during the second study. Similar to the first study (cp. Section 6.2.2) the most prominent emotion was "Neutral" (see Table 7.1), followed by happiness. Sadness and anger are less frequently rated adding up to less than ten percent of all ratings. Disgust, fear, and surprise are each rated less than ten times over the entire study. This is analogous to the response from participants during the interview. All participants stated that the available emotions limited through the first study results were all they needed out of the original seven choices. As Table 7.2 points out, for two participants the range was restricted to just three emotion types: Happiness, neutral, and anger respectively sadness. Four users rated between three and four emotion types while the remaining two users rated all except for one. The least rated emotions were disgust and fear, not being rated once by six of the eight participants, followed by surprise which half of the participants included in their ratings. Anger and sadness were left out twice and once respectively, while all participants rated happiness and neutral in the second study. When asked about why they didn't rate certain emotions, participants mentioned that when feeling those, they most likely would not use a smartphone at that very moment or at least not rate their emotions in such a situation.

For evaluation of possible correlations between emotions and applications participants used during the study, initially, categories are defined to cluster applications of similar context.

Categories providing too few datasets (less than 100 unique ratings) or applications not of particular interest for this thesis (such as *Android Settings*) are excluded.

1. Browser: This group contains all Android browser applications, such as *Google Chrome*[1] or *Mozilla Firefox*[2] .

[1] https://play.google.com/store/apps/details?id=com.android.chrome
[2] https://play.google.com/store/apps/details?id=org.mozilla.firefox

Category	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Ratings	26	6	5	259	787	55	6
Browser	2	0	2	55	170	9	4
Games	0	0	2	28	150	6	0
Messaging	21	5	5	183	508	39	6
Social Media	12	8	1	95	350	45	1

Table 7.1: The top row contains the overall unique number of ratings per emotion. Below the number of ratings for a subset of four categories defined in Section 7.1 is listed.

No	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	Sum
1	-	-	-	Y	Y	Y	-	3
2	Y	-	-	Y	Y	Y	Y	5
3	Y	Y	-	Y	Y	Y	Y	6
4	-	Y	-	Y	Y	Y	-	4
5	Y	-	Y	Y	Y	Y	Y	6
6	Y	-	-	Y	Y	-	-	3
7	Y	-	Y	Y	Y	Y	-	5
8	Y	-	-	Y	Y	Y	Y	5
Sum	6	2	2	8	8	7	4	

Table 7.2: Emotion types rated by the respective participant (No) at least once are marked with a "Y". Column "Sum" describes the sum of emotion types rated by the respective participant, row "Sum" describes how many participants in total rated the respective emotion.

2. Games: In this category, all mobile games are listed.
3. Launcher: The main screen of Android called *Launcher* forms a separate group.
4. Messaging: All applications considered as messengers such as Whatsapp[3], text messaging via native Short Message Service (SMS) application, or Telegram[4] belong to this category.
5. Social Media: Amongst others, Facebook[5], LinkedIn[6], and Instagram[7] belong to this category.

[3] https://play.google.com/store/apps/details?id=com.whatsapp
[4] https://play.google.com/store/apps/details?id=org.telegram.messenger
[5] https://play.google.com/store/apps/details?id=com.facebook.katana
[6] https://play.google.com/store/apps/details?id=com.linkedin.android
[7] https://play.google.com/store/apps/details?id=com.instagram.android

Emotion	Instagram	Jodel	Facebook	Twitter	YouTube
Anger	2	0	9	0	1
Happiness	30	2	28	8	18
Sadness	16	9	16	0	3
Neutral	143	22	95	29	35

Table 7.3: The number of emotion ratings for a subset of the five most prominent applications grouped in social media.

As Table 7.1 shows, only categories messaging and social media show significant ratios of negative emotion ratings. Social media comprise most sadness ratings compared to the number of happiness ratings while messaging and games are relative to the total number of ratings less often rated as sad. Both also hold the highest ratio between anger and happiness of about one to nine. The lowest ratio between happy and sad is in browser category. *Games* is the only category not rated angry at any time.

Comparing on application level amongst the most frequently rated categories – browser, games, messaging, and social media – when not rated as neutral almost all data instances of games are rated as happy. Only two mobile games are also existing in six different instances rated as sad. Interestingly when comparing messaging applications the ratio between happiness and sadness ratings of Snapchat was the highest at 18:1 opposed to Whatsapp with 160 happiness and 34 sadness ratings. Table 7.3 illustrates how some of the social media applications differ amongst each other. Twitter and YouTube have the highest ratio of happiness compared to the negative emotions anger and sadness. For YouTube, this is fitting to statements of the participants, that YouTube is often used to quickly watch funny videos. The number of sadness ratings amounts to more than half of the number of happiness ratings for Facebook and Instagram. Jodel is the only application reporting more negative ratings than positive. One could argue this is due to quickly getting cheered up by reading Jodel messages when feeling negative.

At least in the scope of this thesis data seems to yield indeed certain patterns. Social media has the highest amount of sadness ratings and contains together with Messaging more ratings of anger and disgust than the other categories. Category Games has no anger and disgust ratings at all, which is compliant to the emotions these applications shall provoke.

This study, however, could not assert if the user's current emotion is due to the used application, or caused by other aspects beforehand. In the interview, participants were asked about applications that affect their emotional state. In some aspects, this is opposed to what results from data analysis show. All participants who mentioned games as applications having an impact on their emotions categorized them as causing happiness. Social media applications, such as Facebook, Instagram, and YouTube, were on the other hand mentioned as causing happiness as well, although, during the study, they had significant numbers of negative emotion ratings. This might indicate a disparity between self-perception of emotions and actual emotions provoked by social media applications. Messaging applications such as Whatsapp were in particular connected to any possible emotion, as it

strongly depends on context and person talked to. Two participants notably mentioned Whatsapp causing anger, in times when many messages would appear at the same time or people just didn't want to be bothered at the moment they received notifications.

7.2 Facial Data and Classifier Data Comparison

To answer the second research question (RQ 2, 2a) first off for each emotion the number of data records per emotion is compared for user ratings, facial recognition, and classifier prediction. Figure 7.1 shows that for emotions anger, disgust, fear, happiness, sadness, and surprise the number of data records which the classifier yields seems to approximate to the number of data records the user rated with the respective emotion. For the ambiguous state "neutral", however, the difference is above 20%. Taking a look at the behavior of facial recognition additionally shows that for all emotions except for disgust, the number of total records is closer to the respective number of rating records. This gives already a hint, that using the contextual information "application" might help to improve accuracy, however, it has to be examined further, as the actual comparison per record is not performed here.

To do so, an accuracy formula is defined:

$$d_{avg} = \frac{\sum_{i \in E}^{n} d_i}{n} \tag{7.1}$$

where d_{avg} denotes the average deviation. d_i the deviation of emotion i, E the group of emotions and n the number of emotions in E.

The sum of a participant rating is defined as "1" over all emotion types included. If the participant only rates a single emotion, this emotion is marked with a weight of one. If the participant rates two emotions, both get 0.5 equally. Three emotions lead to a distribution of 0.33 and so on. The deviation of a single emotion is determined by computing the deviation of its weight to the respective facial recognition and classifier prediction value.

This formula is applied to all records of each participant visualized in Figure 7.2. Deviation of evaluated emotions is reduced from almost 75% by facial expression to just above 50% using classifier prediction. This already shows that contextual information is helpful in predicting emotions. However, it is also person-dependent. While as average, the derivation is lower, there are two participants (no. 2 and 3) where pure facial extraction yielded less deviation to the actually rated emotions than classifier prediction and vice versa for instance for participant no. 1 classifier prediction performed up to more than 4 times better.

Results are similar when counting all records for which the facial recognition method yielded better results compared to those where model prediction performed better. This is shown in Figure 7.3. While overall in above two third of the records classifier prediction yielded a better result, again for participant 3 and 4 facial recognition performed better and for participant 7 both methods were almost equal.

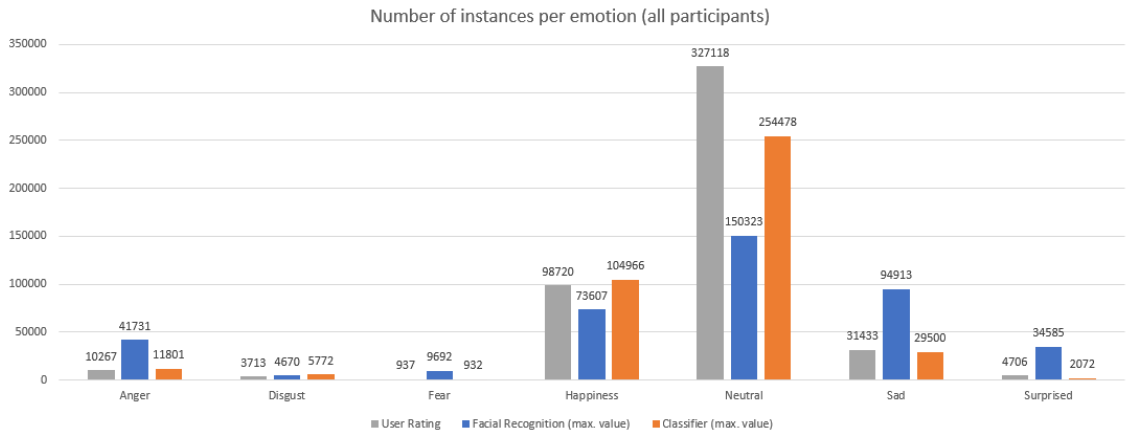


Figure 7.1: Total number of data instances of each emotion the user has rated (gray bars), the facial recognition has determined (blue bars), and the classifier has yielded (orange bars).

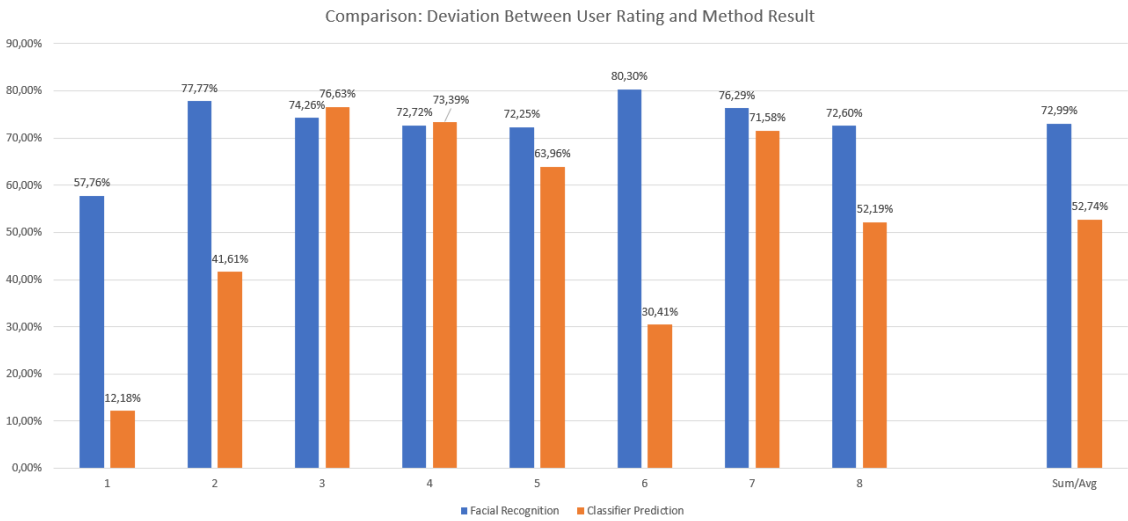


Figure 7.2: Deviation between facial recognition values (blue) and classifier prediction values (orange) compared to the user rating.

Thus this person dependency should be taken into consideration when trying to further improve the results of a predictive classifier.

7.3 Affective Mood Changes

Affective mood changes are measured by comparing the classifier prediction to the user's emotional rating closest to the potential mood change after its occurrence. Furthermore, as qualitative feedback, participants were explicitly asked in the interview whether they noticed a mood change when they were asked to rate their emotions.

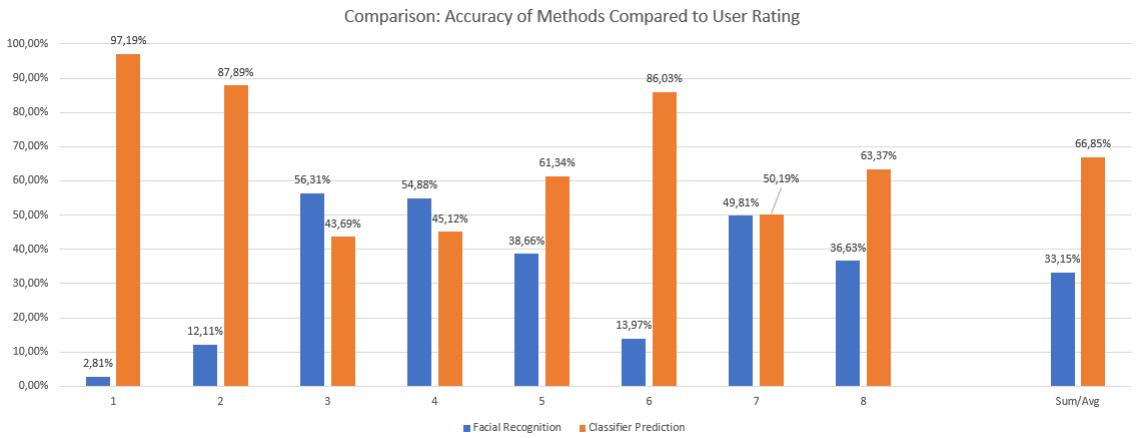


Figure 7.3: This graph illustrates the accuracy of pure facial recognition compared to the user rating in blue as well as the classifier prediction in orange.

Emotion	Total no. of predictions	Incorrectly Predicted	Correctly Predicted
Anger	27	27	0
Disgust	17	15	2
Fear	1	1	0
Happiness	244	191	53
Sadness	84	69	15
Surprise	2	2	0
Neutral	621	118	503
Sum	996	423	573

Table 7.4: Total number of performed predictions by the system during the second study and its correct as well as incorrect share.

Taking a look at the overview in Table 7.4 first off all predictions per emotion are compared against the user ratings.

As neutral has by far the highest number of predictions and also correct instances, it outweighs the other emotions in sum. Happiness and sadness were in more than 20% respectively 17% of times predicted correctly, while neutral is the only entry reporting more correct than incorrect predictions, with a positive share of over 80%. Under the remaining, anger performs worst with 27 or 100% incorrect predictions, followed by disgust with only about 12% correct ones. Neither anger nor fear nor surprise was predicted correctly at all. However, as for fear and surprise only one respectively two predictions existed qualifying these numbers and leading to disregarding these emotions.

Again a deeper look per participant shall give further insights into the correctness of recognizing and predicting affective mood changes. Table 7.5 proves the observation in the previous sections – affective mood changes are person-dependent as well. While disgust

ID	Neutral		Happiness		Disgust		Anger		Sad		Fear		Surprise	
	f	t	f	t	f	t	f	t	f	t	f	t	f	t
1	1	54	0	1	-		-		-		-		-	
2	25	159	10	1	7	0	-		-		-		-	
3	6	21	34	16	1	0	2	0	49	11	-		-	
4	13	49	84	3	7	2	1	0	6	3	-		-	
5	35	32	7	11	-		-		9	1	-		-	
6	12	104	9	1	-		20	0	-		-		-	
7	0	4	6	1	-		-		-		1	0	-	
8	26	80	41	19	-		4	0	5	0	-		2	0

Table 7.5: The total number of predicted affective mood changes per participant. Their correct (t) / incorrect (f) amount is determined by comparing against the closest user rating per participant and predicted emotion.

and anger performed for all users very poorly, participant no. 3 had an average of about one out of six and participant no.4 even 50% positive predictions for sadness whereas for the other participants it did not perform well. Opposed to that happiness prediction performed for participant no. 5 rather well, while for the others, false predictions were predominate. Also neutral shows distinctions between users. Overall it performed rather well including participant no. 1 only reporting one out of 55 predictions as false, however participant no. 5 had more false predictions than true ones.

In contrast, as mentioned in Section 6.3.4, five out of nine participants noticed an affective mood change at times the application asked them to rate their emotions, in particular in situations where they felt happy for a short time. In the data evaluation, this is reflected by a higher number of correct predictions for happiness than the other five general emotions.

The users also mentioned that for them it was hard to distinguish between affective mood changes and their general emotional state. Sometimes they would rate according to their affective mood, for example when watching a video they rated happy although, in general, they felt sad and vice versa, they would just rate according to their general feelings. Thus, the issue of being confused about the actual emotional state seems to also appear in case of distinguishing affective mood and general emotions and should also be taken into consideration in future work.

As well as accuracy, prediction is strongly user dependent and also affected by the high number of “neutral” ratings. These effects are enhanced due to the size and duration of both studies. Ways to improve prediction in future such as an increased number of participants and data records or adjustments in machine learning techniques are discussed in the next chapter.

7.4 Summary

Reviewing the results we could find mood patterns connected to the use of certain applications in Section 7.1 (RQ 1), which were partly confirmed by the participants in the qualitative feedback during the interview. As demonstrated in Section 7.2, emotion prediction using additional contextual information together with facial analysis was possible and, taken as a whole, the accuracy of the predictive models was higher compared to evaluating pure facial data (RQ 2, 2a). Affective mood changes can be captured by the system with varying accuracy depending on the respective emotions involved at that moment (RQ 3) as pointed out in Section 7.3.

7.5 User Experience with the Application

Another outcome of the interview does not directly tackle the RQs, but gives insight into general user experience and privacy concerns regard the application as well as possible suggested use cases by the participants.

7.5.1 General Experience and Side-Effects

There are two factors playing a role in the general experience with the application during the study. First off, the perception of the situation knowing that an application evaluating and predicting mood is running in the background of one's smartphone. Secondly, notifications actively asking for one's emotions constraining oneself to think and become aware of their own mood, possibly in inappropriate or unwelcome moments.

Four users stated that they felt observed during the study. When being asked they said the fact they were told that no private data and no images were stored could not change this feeling. For three participants, knowing that the system tried to evaluate and predict their mood furthermore caused a generally strange, unpleasant feeling for them. One person, on the other hand, found being asked about their mood as amusing, while the others did neither have a positive nor negative opinion about it.

All participants were asked if being continuously questioned about their emotions had an effect on their actual mood. The majority of five people stated it had no impact at all on their mood, however, one of them expressed that the self-awareness of their emotion highly increased: "It was necessary to think about my emotional state. This increased the awareness of my actual emotions, which was not the case before the study". The other four participants said that their mood was influenced. This turned out to happen in different forms. For one user the feeling of the currently experienced emotion was amplified. For instance, in case they were sad, they were reconfirmed in their feeling when answering the questionnaire. Another participant told that only by using the application, his emotions actually changed: "I had an exam coming up and was not nervous until the application asked me about my emotions. With every questionnaire notification, I became more

nervous". They also said that on the other hand, most of the time it was hard to be aware of the actual emotions, hence often "neutral" was selected. Completely to the contrary was the experience for another participant who said that the emotion questionnaire relativized their emotions, especially when experiencing negative ones: "It felt weird to enter feeling sad in the questionnaire when actually really feeling sad, which had a positive effect and kind of eased the situation".

Almost all participants described the frequent pop-up notifications as disturbing when currently being committed to something else on their smartphone. Amongst others they pointed out that when currently writing text to another person or watching videos, notifications particularly interrupted them. The change in the second study from silent notifications to vibrating/sound notifications was noticed by one participant and also mentioned as disturbing. Another user, however, added: "It did not impair my everyday life". Only three participants stated that – at least for the duration of one week – notifications did not bother them at all.

Subsequently, all were asked for reasonable questionnaire frequencies. Most of the users regarded one to two times per hour as an appropriate frequency. Longer periods of time were suggested by two other participants, every two hours and three times a day respectively. Overall two users said that additionally, the pop-up should appear when certain reactions are noticed, such as happiness while watching a video. Furthermore, two persons suggested the frequency should be adjustable by either the user or adaptively. For instance, during spare time or when smartphone usage is comparatively higher, the frequency could be higher than when working.

7.5.2 Privacy Concerns

People were asked if they had privacy concerns while participating in the second study, in fact, if there were situations in which they would not use their smartphone as the application was running in the background, their opinion regarding privacy, and whether they told others around them about the application running, as they could possibly be recorded by the front camera as well.

During the study, the application had almost no effect on the general usage of smartphones. Almost all participants used their phone in any – also private – situation just as they usually do. Only one of them mentioned that they had concerns taking their phone with them in private premises such as the bathroom but still took it with them. A single user pointed out that during the study they wouldn't take their smartphone with them anymore in private premises.

Asking about all data being collected from camera device and multiple sensors, three persons were concerned about the front camera being always active. One additionally pointed out that if pictures were stored, they wouldn't have taken part in the study. Notably, no participant was worried about other sensor data being stored and evaluated such as GPS whereat one person was surprised about the amount of data being collected: "The only

thing that worried me was that such a small application can collect and evaluated so much different data. This makes me concerned about other, especially commercial, apps".

Mentionable is also that there was no user at all who felt the obligation of telling people around them that an application constantly using the front camera was installed on their smartphone. Three persons told others out of interest or "just for fun" that the application was installed, while the six others did not tell anything. When being asked one explicitly answered "No, of course not!", although people around them could be exposed to their smartphones and its front camera as well.

7.5.3 Thoughts and Use Cases

Completing the interview the participants were asked for further thoughts on the application and possible use cases in which a mood prediction system could be applied and various ideas were put forward.

The majority of participants mentioned commercial application fields such as advertisements which often are aimed at provoking certain emotions. If emotions could be evaluated and predicted directly on the smartphone of a person watching the respective commercial, its effectiveness could be evaluated directly as well. Another user said that the smartphone itself could use predicted emotions to suggest applications accordingly: "For instance, if the user is sad, the device could propose playing games". Another idea that came up was to improve applications according to the predicted emotions they cause. If, for example, during the use of an application a person is throughout angry, either content or usability of this application could be improved. Regarding improvements in daily life, two users individually stated that such a system would strongly help to improve self-analysis. People could benefit from learning more about themselves and acquire awareness of their emotions. This is also reflected by participants telling that during the study they actually started thinking about their own emotions which they hadn't experienced before. Associated with life improvement, one user suggested that content-filtering could be applied according to a person's predicted emotions: "When being really sad, for instance, bad news about disasters could be excluded". Two users pointed out medical benefit people could profit from. They mentioned a highly sophisticated mood prediction system could detect health-damaging conditions such as depressions early enough to prevent corresponding risks: "In such cases, for example, the user and their doctor could be alarmed". A similar idea is also moved forward by Ma-Kellams et al. [MOBK16] who utilize Google search data to detect potential suicides evaluating search words put on Google. A sophisticated mood prediction system might show great potential in improving people's health.

8 Conclusion

This chapter summarizes the concept, implementation, and evaluation of *OpenFaceAndroid*. Afterward, ideas for possible future work and its concepts are outlined as well as general limitations regarding this thesis consolidated.

The primary goal of this master thesis was to find out whether contextual information can be correlated to more accurately evaluate emotions derived from facial expressions in order to predict mood in a mobile setting using smartphones. At first, an overview of previous research and basics concerning the thesis background were acquired. With this knowledge, a concept for a mobile mood prediction system was initially developed and RQs formed. The concept was then realized in the Android application *OpenFaceAndroid* gathering contextual data – physical activities, geolocation, current weather, and used application – together with facial expression data provided by *OpenFace++* from smartphones.

In a first study, data was continuously collected from 14 participants for a week and – together with a built-in ground truth put into effect by a periodically appearing questionnaire – prepared to train predictive classifiers for emotions. Gathered contextual data was evaluated and resulted in picking out preferable context information which was the currently used smartphone application. After corresponding adjustments in *OpenFaceAndroid*, a second study with nine participants was conducted again over the same time period. In this study, the application generated questionnaires asking for the current emotions whenever an affective mood change was predicted by the classifier. Furthermore, data was once more constantly logged to find out about correlations between used applications and emotions, the accuracy of mood prediction compared to pure facial expressions and of course the accuracy of predicted affective mood changes.

The outcomes of the study were analyzed post-hoc. As for RQ 1, application categories such as *Messaging* and *Social Media* were determined and results of consolidated applications combined. Evaluating these showed that categories like *Messaging* and *Social Media* had a significant share of negative emotions, while *Games* mainly were connected to positive ones. Comparing the accuracy of predicted emotions to those derived from facial extractions showed that contextual information can indeed improve the estimation of a person's emotion. It is person-dependent but overall, accuracy was twice better than using pure facial expressions. RQ 3 was answered by capturing affective mood changes. The results of the study data revealed that this performed worst. Affective mood changes were detected but the false positive detections were predominate. However, more than half of the participants noticed their affective mood change when being asked by *OpenFaceAndroid* at times.

Possible directions for future work referring to the outcome of and arisen limitations during this thesis are presented in the following.

8.1 Future Work

This master thesis has demonstrated how contextual information can be correlated with emotions derived from facial recognition in a mobile setting. It can be continued by evaluating further types of mobile contextual information. The developed application *OpenFaceAndroid* provides a good foundation to extend and improve mood prediction. Enhancements can be applied for instance in collecting ground truth. The frequency of the emotion questionnaire can either be adjusted or the way how to collect ground truth can be altered, for instance by capturing events during daily use and asking the user to rate their emotions once in the evening for multiple events that were marked. Feedback from the interview after the second study gives hints about what users in this context approved as appropriate. Further APIs for additional contextual data such as calendar events or different sensors like accelerometer can be included into *OpenFaceAndroid*. Methods can be reviewed and edited to grasp deeper information such as not only the currently open application but for instance, also the person chatted with, the website visited, the context in a chat and more. These, however, should be carefully taken into consideration and balanced against user privacy. Four participants in these two studies explicitly stated their concern whether both the website visited and the person chatted with or the text respectively would be logged.

In future, one or multiple of the suggested applications mentioned in Chapter 7 can be implemented.

By the looks of the results, “neutral” by far seems to be the top selection for users when rating their emotions. It would be very interesting to perform further research on that. Many users stated that most of the time they didn’t feel any specific emotion. This could be looked further into and evaluated in a separate study. Smartphone cameras constantly improve in quality. In future, higher detailed video feed allowing more sophisticated facial analysis methods such as evaluation of facial micro-expressions researched on by Pfister et al. [PLZP11] could play a big factor.

8.2 Limitations

One main factor for inferior results of accuracy was the minor number of user ratings during the first study. This caused loosing two third of collected data records as they could not be annotated with user ratings. Furthermore, the higher the time frame of joining data records with user ratings, the less accurate these records are. With a higher number of user ratings, the time frame can be reduced, improving accuracy for subsequent classifier training.

As for this technical implementation, it was difficult to find appropriate users ideally equipped with a Samsung Galaxy smartphone. The diverse and unpredictable behavior of different Android smartphone types complicates developing a mood prediction system running in background as a service. Introducing Android 8 this might become an even more relevant aspect, as it will once more reduce possible background activity and thus interfere with such systems [Goo17b].

The participant count of 14 in the first study and nine in the second was comparatively adequate, however, a higher number of participants might possibly lead to better results. Another vital issue is the duration of a week for gathering data when performing machine learning only once. If in that time a certain emotion rating outweighs, for instance, sadness, this strongly affects the model and is persistent throughout the application in the second study. Often, sadness is assumed although the person might be in a completely different emotional state then. This might be overcome by increasing the time frame of data collection studies and also using deep learning, for instance, to constantly improve the predictive model.

The other main factor is the number of contextual information and their corresponding technical limitations. There is various contextual information that can be collected from a smartphone, however, its availability is limited by the time frame of one week, the irregularities amongst user smartphones, and technical issues that arise. This is in particular due to the increased battery drain any additional activated information source provokes which leads to chosen trade-offs in the technical implementation of this thesis. If these trade-offs could be spared, realized more efficiently, or different approaches found for gathering contextual data, their value and usability would increment.

Of course, there are also always technical limitations when working with camera devices. As front cameras usually have no flashlight, at night without flash or under bad conditions of illumination no data can be derived. Facial data extraction does not work, if the user is not facing the front camera correctly, for instance when looking at the smartphone form a too wide angle.

Eventually, studies, as conducted in this thesis, are strongly dependent on their participants. Within this partition group, “neutral” was the overall overweighting emotion and participants stated that most of the time, they did not feel a particular emotion. This, of course, had strong impacts on the results. One way to avoid that could be by recording certain events or considered emotional peaks and asking the user specifically at such times to rate their emotions. Connected to that is furthermore the existing ambiguousness of public image versus self-image. People might rate emotions differently than they actually are for instance out of unconsciousness, denial, or on purpose. Feedback from the interview also suggested to introduce more subtle emotions Emotions were especially in the first study rated infrequently, hence many data records could not be used for classifier training. Additionally, accuracy would increase, the more times emotions were rated. Furthermore, smartphone applications can always be turned off and unlike a study in a fix laboratory setup, one is exposed to the smartphone owners decision when performing a study on

their devices. With more participants or implementations not allowing to terminate the application (for instance by exploiting root access) this issue might be avoided.

A Appendix

The appendix contains the following documents:

1. Consent form (Figure A.1)
2. Questionnaire first study (Figure A.2)
3. Questionnaire second study (Figure A.3)
4. Interview second study (Figure A.4)
5. Classification accuracy results (in detail per participant) (Figures A.5 to A.16)

University of Stuttgart
Germany

Human Computer Interaction Group (MCI), VIS
Prof. Dr. Albrecht Schmidt

Consent Form

DESCRIPTION: You are invited to participate in **a field study** for investigating **emotion detection in mobile settings.**

TIME INVOLVEMENT: Your participation will take approximately **one week.**

DATA COLLECTION: For this study, you will install an app which measures your facial emotional responses. Contextual data, such as current weather or your currently used app, will be collected during the study.

RISKS AND BENEFITS: No risk is associated with this study. The collected data is securely stored. We do guarantee no data misuse and privacy is completely preserved.

PAYMENTS: You will receive **10 Euro OR a participation point for the Human-Computer Interaction lecture** as payment for your participation.

PARTICIPANT'S RIGHTS: If you have read this form and have decided to participate in this project, please understand your **participation is voluntary** and you have the **right to withdraw your consent or discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled**. **The alternative is not to participate.** You have the right to refuse to answer particular questions. The results of this research study may be presented at scientific or professional meetings or published in scientific journals. Your identity is not disclosed unless we directly inform and ask for your permission.

CONTACT INFORMATION: If you have any questions, concerns or complaints about this research, its procedures, risks and benefits, contact following persons:
Robin Reutter (robin.reutter@gmail.com)
Thomas Kosch (thomas.kosch@vis.uni-stuttgart.de)
Mariam Hassib (mariam.hassib@vis.uni-stuttgart.de)

By signing this document I confirm that I agree to the terms and conditions.

Name: _____ *Signature, Date:* _____

Figure A.1: Consent Form

General Questions

Participant (Study ID) _____
Start-/End Date _____
Smartphone _____

1. Application Information
Battery _____
Memory _____
Storage _____
Mobile Data _____

2. Perception: Battery Drain

Less	Equal	Higher	Much higher

Comments:

3. Perception: Performance

Better	Equal	Worse	Much Worse

Comments:

4. Special occasions during study

Date/Time	Emotion	Comments

Figure A.2: Questionnaire First Study

General Questions

Participant (Study ID) _____
Start-/End Date _____
Smartphone _____

1. Application Information
Battery _____
Memory _____
Storage _____

2. Perception: Battery Drain

Less	Equal	Higher	Much higher

Comments:

3. Perception: Performance

Better	Equal	Worse	Much Worse

Comments:

4. Special occasions during study

Date/Time	Emotion	Comments

5. Perception: Popup Notification
Frequency (rare/normal/often/too often)
Comments:

Emotion change detected when surveyed?
__ yes __ no

Comments.

Figure A.3: Questionnaire Second Study

Interview Questionnaire

1. **GENERAL**
    a. What is your general experience with the app?
    b. How did you deal with being asked about your emotions continuously? Were there situations in which you felt this had an influence on your mood?
    c. How many times should the app ask you for your current emotion? What would be an acceptable frequency (e.g. per hour)?
    d. How often do you check your phone per day? When and in what context do you use it the most?
    e. What apps do you usually use? Name (1-3) apps that you think affect you and explain why.

2. **PRIVACY**
    a. Were there any moments in which you particularly deliberately disabled the app or did not take the phone somewhere with you or covered the camera? In case of yes, which and why?
    b. How do you feel about the app knowing that it uses your frontal camera and sensory data in particular?
    c. Did you tell the people who are around you about the app? Do you feel it makes a difference if they know of it or not? Did you feel obliged to tell them about it?

3. **USE CASES**
    a. As a second step of this project, we would like to use the information we have right now, i.e. that we can accurately know your current emotion, to help you in your daily life. What kind of applications (or features in apps you use) do you think would support you and make use of this information (e.g. Notifying you about the mood of other people nearby)?

Figure A.4: Interview Second Study

```
User: OpenFaceAndroid_11369403.arff - Attributes: pbAnger,pbHappiness,pbNeutral,pbSad,pbSurprised,app,class

Results
======

Correctly Classified Instances       10394               96.8235 %
Incorrectly Classified Instances       341                3.1765 %
Kappa statistic                          0.6815
K&B Relative Info Score             514792.3735 %
K&B Information Score                 2122.9318 bits      0.1978 bits/instance
Class complexity | order 0            4394.9545 bits      0.4094 bits/instance
Class complexity | scheme            49432.0709 bits      4.6048 bits/instance
Complexity improvement     (Sf)     -45037.1163 bits     -4.1954 bits/instance
Mean absolute error                      0.0176
Root mean squared error                  0.0978
Relative absolute error                 39.4136 %
Root relative squared error             65.5414 %
Total Number of Instances            10735

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,713    0,003    0,818      0,713    0,762      0,760    0,971     0,842     User_Anger
                 0,533    0,006    0,720      0,533    0,613      0,611    0,949     0,690     User_Happiness
                 0,991    0,360    0,978      0,991    0,985      0,708    0,959     0,996     User_Neutral
                 0,561    0,002    0,740      0,561    0,638      0,640    0,945     0,660     User Sad
                 0,000    0,000    0,000      0,000    0,000      0,000    0,832     0,191     User Surprised
Weighted Avg.    0,968    0,339    0,965      0,968    0,966      0,706    0,959     0,980

=== Confusion Matrix ===

     a     b     c     d     e   <-- classified as
   144     5    53     0     0 |     a = User_Anger
     5   152   117    11     0 |     b = User Happiness
    27    47 10024    15     0 |     c = User Neutral
     0     7    51    74     0 |     d = User Sad
     0     0     3     0     0 |     e = User Surprised
```

Figure A.5: Classifier Accuracy Details Subject 1

```
User: OpenFaceAndroid_39242486.arff - Attributes: pbAnger,pbDisgust,pbFear,pbHappiness,pbNeutral,pbSad,pbSurprised,app,class

Results
======

Correctly Classified Instances       16388               78.3777 %
Incorrectly Classified Instances      4521               21.6223 %
Kappa statistic                          0.5154
K&B Relative Info Score             903674.9731 %
K&B Information Score                11562.2167 bits      0.553  bits/instance
Class complexity | order 0           26720.4698 bits      1.2779 bits/instance
Class complexity | scheme           237493.0187 bits     11.3584 bits/instance
Complexity improvement     (Sf)    -210772.5489 bits    -10.0805 bits/instance
Mean absolute error                      0.0865
Root mean squared error                  0.2097
Relative absolute error                 62.6698 %
Root relative squared error             79.8307 %
Total Number of Instances            20909

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,603    0,002    0,760      0,603    0,673      0,675    0,937     0,688     User_Anger
                 0,209    0,001    0,780      0,209    0,329      0,401    0,856     0,316     User_Disgust
                 0,667    0,002    0,808      0,667    0,730      0,731    0,952     0,759     User Fear
                 0,605    0,095    0,705      0,605    0,652      0,536    0,869     0,741     User Happiness
                 0,901    0,411    0,812      0,901    0,854      0,525    0,857     0,914     User Neutral
                 0,146    0,004    0,515      0,146    0,227      0,264    0,828     0,268     User Sad
                 0,056    0,000    0,308      0,056    0,094      0,130    0,810     0,101     User_Surprised
Weighted Avg.    0,784    0,299    0,772      0,784    0,770      0,522    0,861     0,836

=== Confusion Matrix ===

     a     b     c     d     e     f     g   <-- classified as
   111     0     0     9    64     0     0 |     a = User Anger
     0    39     1    22   124     1     0 |     b = User Disgust
     0     0   168    38    43     1     2 |     c = User_Fear
     4     3    18  3471  2215    21     3 |     d = User_Happiness
    31     8    18  1248 12507    60     4 |     e = User_Neutral
     0     0     0   103   412    88     0 |     f = User Sad
     0     0     3    29    36     0     4 |     g = User Surprised
```

Figure A.6: Classifier Accuracy Details Subject 2

```
User: OpenFaceAndroid_46084488.arff - Attributes: pbAnger,pbDisgust,pbHappiness,pbNeutral,pbSad,pbSurprised,app,class

Results
======

Correctly Classified Instances        5514               79.8552 %
Incorrectly Classified Instances      1391               20.1448 %
Kappa statistic                          0.7149
K&B Relative Info Score             470064.6497 %
K&B Information Score                 9574.0459 bits      1.3865 bits/instance
Class complexity | order 0           14053.042  bits      2.0352 bits/instance
Class complexity | scheme            49115.6217 bits      7.1131 bits/instance
Complexity improvement     (Sf)     -35062.5797 bits     -5.0779 bits/instance
Mean absolute error                      0.0953
Root mean squared error                  0.2189
Relative absolute error                 39.9931 %
Root relative squared error             63.4231 %
Total Number of Instances             6905

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,548    0,016    0,722      0,548    0,623      0,606    0,945     0,692     User_Anger
                 0,651    0,002    0,707      0,651    0,678      0,675    0,971     0,708     User_Disgust
                 0,859    0,113    0,809      0,859    0,833      0,737    0,946     0,910     User_Happiness
                 0,660    0,053    0,699      0,660    0,679      0,622    0,934     0,780     User Neutral
                 0,855    0,092    0,835      0,855    0,845      0,759    0,954     0,926     User Sad
                 0,776    0,008    0,850      0,776    0,811      0,802    0,979     0,868     User Surprised
Weighted Avg.    0,799    0,083    0,796      0,799    0,796      0,720    0,949     0,876

=== Confusion Matrix ===

    a    b    c    d    e    f   <-- classified as
  262    3  131   29   37   16 |    a = User Anger
    2   41    5    3   12    0 |    b = User Disgust
   45    5 2123  119  156   23 |    c = User Happiness
   11    3  154  715  189   11 |    d = User Neutral
   31    5  165  152 2085    1 |    e = User_Sad
   12    1   47    5   18  288 |    f = User_Surprised
```

Figure A.7: Classifier Accuracy Details Subject 3

```
User: OpenFaceAndroid_47139796.arff - Attributes: pbAnger,pbHappiness,pbNeutral,pbSurprised,app,class

Results
======

Correctly Classified Instances        9235               84.5541 %
Incorrectly Classified Instances      1687               15.4459 %
Kappa statistic                          0.7439
K&B Relative Info Score             763193.7629 %
K&B Information Score                10983.1948 bits      1.0056 bits/instance
Class complexity | order 0           15711.0922 bits      1.4385 bits/instance
Class complexity | scheme            66733.1314 bits      6.11   bits/instance
Complexity improvement     (Sf)     -51022.0393 bits     -4.6715 bits/instance
Mean absolute error                      0.1055
Root mean squared error                  0.2347
Relative absolute error                 34.8206 %
Root relative squared error             60.2905 %
Total Number of Instances            10922

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,771    0,025    0,831      0,771    0,800      0,770    0,972     0,892     User_Anger
                 0,844    0,118    0,823      0,844    0,834      0,723    0,941     0,916     User_Happiness
                 0,869    0,116    0,869      0,869    0,869      0,753    0,949     0,941     User Neutral
                 0,000    0,000    0,000      0,000    0,000      0,000    0,493     0,000     User Surprised
Weighted Avg.    0,846    0,104    0,845      0,846    0,845      0,743    0,949     0,924

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 1146  197  144    0 |    a = User Anger
  143 3635  528    0 |    b = User Happiness
   90  580 4454    0 |    c = User Neutral
    0    3    2    0 |    d = User Surprised
```

Figure A.8: Classifier Accuracy Details Subject 4

```
User: OpenFaceAndroid_55374085.arff - Attributes: pbAnger,pbDisgust,pbHappiness,pbNeutral,pbSad,pbSurprised,app,class

Results
======

Correctly Classified Instances       10230               72.8166 %
Incorrectly Classified Instances      3819               27.1834 %
Kappa statistic                          0.5908
K&B Relative Info Score             772406.5111 %
K&B Information Score               14640.9238 bits      1.0421 bits/instance
Class complexity | order 0          26613.131  bits      1.8943 bits/instance
Class complexity | scheme          142107.9065 bits     10.1152 bits/instance
Complexity improvement     (Sf)    -115494.7755 bits     -8.2209 bits/instance
Mean absolute error                      0.1223
Root mean squared error                  0.2476
Relative absolute error                 54.0061 %
Root relative squared error             73.5955 %
Total Number of Instances            14049

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,485    0,011    0,672      0,485    0,563      0,554    0,922     0,606     User_Anger
                 0,628    0,012    0,768      0,628    0,691      0,677    0,939     0,749     User_Disgust
                 0,843    0,216    0,774      0,843    0,807      0,625    0,905     0,899     User_Happiness
                 0,659    0,116    0,673      0,659    0,666      0,547    0,881     0,750     User_Neutral
                 0,616    0,059    0,667      0,616    0,641      0,576    0,913     0,739     User_Sad
                 0,000    0,000    0,000      0,000    0,000      0,000    0,495     0,000     User_Surprised
Weighted Avg.    0,728    0,143    0,725      0,728    0,724      0,596    0,903     0,811

=== Confusion Matrix ===

    a    b    c    d    e    f   <-- classified as
  314    6  159  110   59    0 |    a = User_Anger
    6  524  148  123   34    0 |    b = User_Disgust
   75   57 5542  609  292    0 |    c = User_Happiness
   39   78  841 2454  311    0 |    d = User_Neutral
   33   17  468  351 1396    0 |    e = User_Sad
    0    0    1    2    0    0 |    f = User_Surprised
```

Figure A.9: Classifier Accuracy Details Subject 5

```
User: OpenFaceAndroid_56739597.arff - Attributes: pbAnger,pbHappiness,pbNeutral,app,class

Results
======

Correctly Classified Instances        2310               83.1833 %
Incorrectly Classified Instances       467               16.8167 %
Kappa statistic                          0.6131
K&B Relative Info Score             142835.3026 %
K&B Information Score                1442.7769 bits      0.5195 bits/instance
Class complexity | order 0           2795.8892 bits      1.0068 bits/instance
Class complexity | scheme           36922.7851 bits     13.2959 bits/instance
Complexity improvement     (Sf)     -34126.8959 bits    -12.2891 bits/instance
Mean absolute error                      0.1455
Root mean squared error                  0.289
Relative absolute error                 48.403  %
Root relative squared error             74.5564 %
Total Number of Instances            2777

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,348    0,005    0,533      0,348    0,421      0,423    0,814     0,422     User_Anger
                 0,698    0,087    0,783      0,698    0,738      0,631    0,896     0,831     User_Happiness
                 0,906    0,315    0,855      0,906    0,880      0,613    0,881     0,924     User_Neutral
Weighted Avg.    0,832    0,239    0,828      0,832    0,828      0,615    0,885     0,887

=== Confusion Matrix ===

    a    b    c   <-- classified as
   16    3   27 |    a = User_Anger
    2  602  259 |    b = User_Happiness
   12  164 1692 |    c = User_Neutral
```

Figure A.10: Classifier Accuracy Details Subject 6

```
User: OpenFaceAndroid_57573917.arff - Attributes: pbAnger,pbFear,pbHappiness,pbNeutral,pbSad,pbSurprised,app,class

Results
======

Correctly Classified Instances       18121               70.2991 %
Incorrectly Classified Instances      7656               29.7009 %
Kappa statistic                          0.5195
K&B Relative Info Score            1218956.495  %
K&B Information Score               21897.926  bits      0.8495 bits/instance
Class complexity | order 0          46293.0094 bits      1.7959 bits/instance
Class complexity | scheme          263991.3504 bits     10.2414 bits/instance
Complexity improvement     (Sf)    -217698.341  bits     -8.4454 bits/instance
Mean absolute error                      0.1345
Root mean squared error                  0.2601
Relative absolute error                 62.0342 %
Root relative squared error             78.9996 %
Total Number of Instances            25777

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,264    0,009    0,550      0,264    0,356      0,364    0,864     0,391     User_Anger
                 0,394    0,010    0,590      0,394    0,473      0,467    0,919     0,497     User_Fear
                 0,590    0,097    0,653      0,590    0,620      0,510    0,872     0,708     User_Happiness
                 0,845    0,337    0,722      0,845    0,779      0,518    0,852     0,854     User Neutral
                 0,611    0,047    0,736      0,611    0,668      0,608    0,899     0,749     User Sad
                 0,000    0,000    0,000      0,000    0,000      0,000    0,681     0,226     User Surprised
Weighted Avg.    0,703    0,204    0,696      0,703    0,693      0,523    0,868     0,769

=== Confusion Matrix ===

     a     b     c     d     e     f   <-- classified as
   276    12   164   534    61     0 |     a = User Anger
    11   360   110   336    96     0 |     b = User Fear
    40    56  3609  2122   294     0 |     c = User Happiness
   135   118  1232 11095   545     0 |     d = User Neutral
    40    64   407  1261  2781     0 |     e = User_Sad
     0     0     3    14     1     0 |     f = User_Surprised
```

Figure A.11: Classifier Accuracy Details Subject 7

```
User: OpenFaceAndroid_60334438.arff - Attributes: pbAnger,pbHappiness,pbNeutral,app,class

Results
======

Correctly Classified Instances        2512               88.513  %
Incorrectly Classified Instances       326               11.487  %
Kappa statistic                          0.4673
K&B Relative Info Score              91570.5257 %
K&B Information Score                  627.406  bits      0.2211 bits/instance
Class complexity | order 0            1936.1502 bits      0.6822 bits/instance
Class complexity | scheme            31070.7602 bits     10.9481 bits/instance
Complexity improvement     (Sf)     -29134.61   bits    -10.2659 bits/instance
Mean absolute error                      0.0887
Root mean squared error                  0.2299
Relative absolute error                 57.1472 %
Root relative squared error             82.5919 %
Total Number of Instances             2838

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,422    0,023    0,559      0,422    0,481      0,456    0,876     0,510     User_Anger
                 0,465    0,032    0,506      0,465    0,485      0,451    0,934     0,571     User_Happiness
                 0,950    0,499    0,928      0,950    0,939      0,487    0,916     0,984     User Neutral
Weighted Avg.    0,885    0,438    0,877      0,885    0,880      0,483    0,915     0,927

=== Confusion Matrix ===

    a    b    c   <-- classified as
   76   12   92 |    a = User_Anger
    9   86   90 |    b = User Happiness
   51   72 2350 |    c = User Neutral
```

Figure A.12: Classifier Accuracy Details Subject 8

```
User: OpenFaceAndroid_74060362.arff - Attributes: pbAnger,pbDisgust,pbFear,pbHappiness,pbNeutral,pbSurprised,app,class

Results
======

Correctly Classified Instances        2875               83.9416 %
Incorrectly Classified Instances       550               16.0584 %
Kappa statistic                          0.6263
K&B Relative Info Score             194846.278  %
K&B Information Score                 2668.1125 bits      0.779  bits/instance
Class complexity | order 0            4671.616  bits      1.364  bits/instance
Class complexity | scheme            33094.9487 bits      9.6628 bits/instance
Complexity improvement     (Sf)     -28423.3327 bits     -8.2988 bits/instance
Mean absolute error                      0.076
Root mean squared error                  0.1968
Relative absolute error                 49.4829 %
Root relative squared error             71.0723 %
Total Number of Instances             3425

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,664    0,059    0,680      0,664    0,672      0,611    0,925     0,739     User_Anger
                 0,329    0,004    0,692      0,329    0,446      0,469    0,911     0,489     User_Disgust
                 0,261    0,001    0,600      0,261    0,364      0,393    0,966     0,382     User_Fear
                 0,764    0,007    0,878      0,764    0,817      0,808    0,973     0,880     User_Happiness
                 0,934    0,323    0,878      0,934    0,905      0,646    0,925     0,965     User_Neutral
                 0,284    0,007    0,579      0,284    0,382      0,392    0,924     0,390     User_Surprised
Weighted Avg.    0,839    0,240    0,830      0,839    0,830      0,636    0,928     0,889

=== Confusion Matrix ===

    a    b    c    d    e    f   <-- classified as
  362    3    1    4  166    9 |    a = User_Anger
   15   27    0    0   40    0 |    b = User_Disgust
    1    1    6    2   13    0 |    c = User_Fear
    8    0    0  165   43    0 |    d = User_Happiness
  120    7    3   16 2282   15 |    e = User_Neutral
   26    1    0    1   55   33 |    f = User_Surprised
```

Figure A.13: Classifier Accuracy Details Subject 9

```
User: OpenFaceAndroid_82719134.arff - Attributes: pbAnger,pbDisgust,pbFear,pbHappiness,pbNeutral,pbSad,pbSurprised,app,class

Results
======

Correctly Classified Instances        4324               71.1301 %
Incorrectly Classified Instances      1755               28.8699 %
Kappa statistic                          0.5835
K&B Relative Info Score             336764.7228 %
K&B Information Score                 6789.5857 bits      1.1169 bits/instance
Class complexity | order 0           12236.0918 bits      2.0128 bits/instance
Class complexity | scheme            77824.7628 bits     12.8022 bits/instance
Complexity improvement     (Sf)     -65588.671  bits    -10.7894 bits/instance
Mean absolute error                      0.1114
Root mean squared error                  0.2379
Relative absolute error                 55.278  %
Root relative squared error             74.9439 %
Total Number of Instances             6079

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,499    0,022    0,628      0,499    0,556      0,531    0,925     0,611     User_Anger
                 0,440    0,003    0,564      0,440    0,494      0,495    0,945     0,453     User_Disgust
                 0,705    0,056    0,734      0,705    0,719      0,660    0,942     0,801     User_Fear
                 0,640    0,140    0,640      0,640    0,640      0,500    0,857     0,692     User_Happiness
                 0,819    0,192    0,757      0,819    0,787      0,622    0,900     0,863     User_Neutral
                 0,534    0,007    0,764      0,534    0,628      0,627    0,928     0,693     User_Sad
                 0,412    0,001    0,438      0,412    0,424      0,423    0,876     0,345     User_Surprised
Weighted Avg.    0,711    0,132    0,709      0,711    0,708      0,587    0,899     0,775

=== Confusion Matrix ===

    a    b    c    d    e    f    g   <-- classified as
  209    4   38   63   97    6    2 |    a = User_Anger
   10   22    2    8    8    0    0 |    b = User_Disgust
   16    2  770  204   96    1    3 |    c = User_Fear
   35    6  157 1088  408    4    1 |    d = User_Happiness
   50    5   72  310 2102   27    0 |    e = User_Neutral
    9    0    8   27   63  126    3 |    f = User_Sad
    4    0    2    1    2    1    7 |    g = User_Surprised
```

Figure A.14: Classifier Accuracy Details Subject 10

```
User: OpenFaceAndroid_90796292.arff - Attributes: pbFear,pbHappiness,pbNeutral,pbSad,app,class

Results
======

Correctly Classified Instances        2788               82.7791 %
Incorrectly Classified Instances       580               17.2209 %
Kappa statistic                          0.6915
K&B Relative Info Score             218254.4593 %
K&B Information Score                 2901.4004 bits      0.8615 bits/instance
Class complexity | order 0            4467.7351 bits      1.3265 bits/instance
Class complexity | scheme            25530.7676 bits      7.5804 bits/instance
Complexity improvement     (Sf)     -21063.0325 bits     -6.2539 bits/instance
Mean absolute error                      0.1141
Root mean squared error                  0.2484
Relative absolute error                 40.8055 %
Root relative squared error             66.4654 %
Total Number of Instances             3368

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,760    0,017    0,741      0,760    0,751      0,735    0,977     0,814     User_Fear
                 0,500    0,004    0,552      0,500    0,525      0,521    0,942     0,454     User_Happiness
                 0,855    0,170    0,840      0,855    0,847      0,685    0,922     0,922     User_Neutral
                 0,812    0,120    0,831      0,812    0,822      0,695    0,923     0,906     User Sad
Weighted Avg.    0,828    0,138    0,828      0,828    0,828      0,691    0,926     0,905

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
  152    0   28   20 |    a = User_Fear
    0   16   13    3 |    b = User_Happiness
   28   11 1469  211 |    c = User Neutral
   25    2  239 1151 |    d = User Sad
```

Figure A.15: Classifier Accuracy Details Subject 11

```
User: OpenFaceAndroid_93381394.arff - Attributes: pbAnger,pbHappiness,pbNeutral,pbSad,pbSurprised,app,class

Results
======

Correctly Classified Instances        4201               87.9791 %
Incorrectly Classified Instances       574               12.0209 %
Kappa statistic                          0.753
K&B Relative Info Score             340681.3056 %
K&B Information Score                 4839.6995 bits      1.0135 bits/instance
Class complexity | order 0            6772.3558 bits      1.4183 bits/instance
Class complexity | scheme            52409.1698 bits     10.9757 bits/instance
Complexity improvement     (Sf)     -45636.814  bits     -9.5574 bits/instance
Mean absolute error                      0.0658
Root mean squared error                  0.188
Relative absolute error                 32.8711 %
Root relative squared error             59.452  %
Total Number of Instances             4775

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,618    0,010    0,737      0,618    0,672      0,661    0,954     0,733     User_Anger
                 0,857    0,044    0,832      0,857    0,844      0,804    0,977     0,916     User_Happiness
                 0,943    0,178    0,916      0,943    0,930      0,779    0,962     0,978     User Neutral
                 0,550    0,011    0,721      0,550    0,624      0,613    0,933     0,649     User Sad
                 0,547    0,007    0,707      0,547    0,617      0,611    0,940     0,650     User Surprised
Weighted Avg.    0,880    0,130    0,875      0,880    0,876      0,765    0,963     0,928

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
  126   41   36    1    0 |    a = User Anger
   19  827  102    8    9 |    b = User Happiness
   25   93 3034   42   22 |    c = User Neutral
    1   21   83  132    3 |    d = User Sad
    0   12   56    0   82 |    e = User_Surprised
```

Figure A.16: Classifier Accuracy Details Subject 12

Acknowledgments

My special thanks go to Prof. Dr. Albrecht Schmidt for the opportunity to compose this work at his institute.

I most sincerely thank my both supervisors Mariam Hassib and Thomas Kosch for their continuous support throughout the past six months.

Furthermore, I'd like to express my gratitude to my family, friends and fellow students steadily encouraging me during my entire study.

Bibliography

[AP14] Y. Ayzenberg, R. W. Picard. "FEEL: A System for Frequent Event and Electrodermal Activity Labeling." In: *IEEE J. Biomedical and Health Informatics* 18.1 (2014), pp. 266–277. DOI: 10.1109/JBHI.2013.2278213. URL: https://doi.org/10.1109/JBHI.2013.2278213 (cit. on pp. 15, 24).

[Ban17a] C. Banzhaf. "Extracting Facial Data using Feature-based Image Processing and Correlating it with Alternative Biosensors Metrics." MA thesis. Institute for Visualization and Interactive Systems, Jan. 24, 2017 (cit. on pp. 15, 19–21, 27).

[Ban17b] C. Banzhaf. *GitHub - cyberjunk/OpenFace: OpenFace++ – an extended variant of OpenFace. Includes facial expression detection, Android support and more.* Oct. 13, 2017. URL: https://github.com/cyberjunk/OpenFace (cit. on p. 21).

[BC17] L. Breiman, A. Cutler. *Random forests - classification description*. Oct. 13, 2017. URL: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#features (cit. on p. 48).

[BLFM03] M. S. Bartlett, G. Littlewort, I. Fasel, J. R. Movellan. "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction." In: *2003 Conference on Computer Vision and Pattern Recognition Workshop*. Vol. 5. June 2003, pp. 53–53. DOI: 10.1109/CVPRW.2003.10057 (cit. on p. 15).

[Bre01] L. Breiman. "Random Forests." In: *Machine Learning* 45.1 (2001), pp. 5–32 (cit. on p. 48).

[BRM13] T. Baltrušaitis, P. Robinson, L.-P. Morency. "Constrained Local Neural Fields for robust facial landmark detection in the wild." In: *Conference on Computer Vision Workshops*. Jan. 1, 2013 (cit. on p. 19).

[BRM16] T. Baltrušaitis, P. Robinson, L. P. Morency. "OpenFace: An open source facial behavior analysis toolkit." In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Mar. 2016, pp. 1–10. DOI: 10.1109/WACV.2016.7477553 (cit. on p. 19).

[CC06] D. Cristinacce, T. F. Cootes. "Feature Detection and Tracking with Constrained Local Models." In: *Proceedings of the British Machine Vision Conference*. doi:10.5244/C.20.95. BMVA Press, 2006, pp. 95.1–95.10. ISBN: 1-901725-32-4 (cit. on p. 19).

[CL16] C.-C. Chang, C.-J. Lin. *LIBSVM: a library for support vector machines*. Dec. 22, 2016. URL: http://www.csie.ntu.edu.tw/~cjlin/libsvm (cit. on p. 48).

[Cor17] I. D. Corporation. *IDC: Smartphone OS Market Share*. Oct. 6, 2017. URL: https://www.idc.com/promo/smartphone-market-share/os (cit. on p. 29).

[Dar99] C. Darwin. *The Expression of the Emotions in Man and Animals*. 1899 (cit. on p. 15).

[DVD12] J. A. Datu, J. P. Valdez, N. Datu. "Does Facebooking make us sad? Hunting relationship between Facebook use and depression among Filipino adolescents." In: *International Journal of Research Studies in Educational Technology* (Oct. 1, 2012) (cit. on p. 16).

[EF70] P. Ekman, W. V. Friesen. "CONSTANTS ACROSS CULTURES IN THE FACE AND EMOTION." In: *Vol. 17, No. 2* (Jan. 29, 1970) (cit. on pp. 15, 19).

[EFE13] P. Ekman, W. V. Friesen, P. Ellsworth. *Emotion in the Human Face - Guidelines for Research and an Integration of Findings*. Amsterdam: Elsevier, 2013. ISBN: 978-1-483-14763-5 (cit. on pp. 15, 19).

[FZCX13] R. Fan, J. Zhao, Y. Chen, K. Xu. "Anger is More Influential Than Joy: Sentiment Correlation in Weibo." In: (Sept. 10, 2013) (cit. on p. 16).

[Goo17a] Google. *ActivityRecognitionClient | Google APIs for Android | Google Developers*. Oct. 6, 2017. URL: https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionClient (cit. on p. 34).

[Goo17b] Google. *Android – 8.0 Oreo*. Oct. 20, 2017. URL: https://www.android.com/versions/oreo-8-0/ (cit. on p. 69).

[Goo17c] Google. *Broadcasts | Android Developers*. Oct. 14, 2017. URL: https://developer.android.com/guide/components/broadcasts.html (cit. on p. 30).

[Goo17d] Google. *DetectedActivity | Google APIs for Android | Google Developers*. Oct. 6, 2017. URL: https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity (cit. on pp. 34, 35).

[Goo17e] Google. *ImageFormat | Android Developers*. Oct. 14, 2017. URL: https://developer.android.com/reference/android/graphics/ImageFormat.html (cit. on p. 34).

[Goo17f] Google. *Manifest.permission | Android Developers*. Oct. 14, 2017. URL: https://developer.android.com/reference/android/Manifest.permission.html (cit. on p. 30).

[Goo17g] Google. *Review app permissions thru Android 5.9 - Computer - Google Play Help*. Oct. 14, 2017. URL: https://support.google.com/googleplay/answer/6014972?hl=en (cit. on p. 33).

[Gro17] N. W. Group. *RFC 4180 - Common Format and MIME Type for Comma-Separated Values (CSV) Files*. Oct. 13, 2017. URL: https://tools.ietf.org/html/rfc4180 (cit. on p. 36).

[HP14] J. Hernandez, R. W. Picard. "SenseGlass: Using Google Glass to Sense Daily Emotions." In: *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST'14 Adjunct. Honolulu, Hawaii, USA: ACM, 2014, pp. 77–78. ISBN: 978-1-4503-3068-8. DOI: 10.1145/2658779.2658784. URL: http://doi.acm.org/10.1145/2658779.2658784 (cit. on pp. 23, 24).

[HTF09] T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning*. Springer, New York, 2009 (cit. on p. 48).

[IKK05] T. Iso, N. Kwasaki, S. Kurakake. "Personal Context Extractor with Multiple Sensors on a Cell Android-Smartphones." In: 2005 (cit. on p. 23).

[JY02] Q. Ji, X. Yang. "Real-time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance." In: *Real-Time Imaging* 8.5 (Oct. 2002), pp. 357–377. ISSN: 1077-2014. DOI: 10.1006/rtim.2002.0279. URL: http://dx.doi.org/10.1006/rtim.2002.0279 (cit. on p. 15).

[LCK+10] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews. "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression." In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. June 2010, pp. 94–101. DOI: 10.1109/CVPRW.2010.5543262 (cit. on pp. 20, 21).

[LF04] D. J. Litman, K. Forbes-Riley. "Predicting Student Emotions in Computer-human Tutoring Dialogues." In: *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*. ACL '04. Barcelona, Spain: Association for Computational Linguistics, 2004. DOI: 10.3115/1218955.1219000. URL: https://doi.org/10.3115/1218955.1219000 (cit. on p. 27).

[ME03] P. Michel, R. El Kaliouby. "Real Time Facial Expression Recognition in Video Using Support Vector Machines." In: *Proceedings of the 5th International Conference on Multimodal Interfaces*. ICMI '03. Vancouver, British Columbia, Canada: ACM, 2003, pp. 258–264. ISBN: 1-58113-621-8. DOI: 10.1145/958432.958479. URL: http://doi.acm.org/10.1145/958432.958479 (cit. on p. 24).

[MKK+12] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, M. Czerwinski. "AffectAura: An Intelligent System for Emotional Memory." In: Oct. 5, 2012 (cit. on pp. 15, 24, 25).

[MOBK16] C. Ma-Kellams, F. Or, J. H. Baek, I. Kawachi. "Rethinking Suicide Surveillance: Google Search Data and Self-Reported Suicidality Differentially Estimate Completed Suicide Risk." In: *Clinical Psychological Science* 4.3 (2016), pp. 480–484. DOI: 10.1177/2167702615593475. eprint: https://doi.org/10.1177/2167702615593475. URL: https://doi.org/10.1177/2167702615593475 (cit. on p. 66).

[MTHM17] A. Mehrotra, F. Tsapeli, R. Hendley, M. Musolesi. "MyTraces: Investigating Correlation and Causation Between Users&Rsquo; Emotional States and Mobile Phone Interaction." In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1.3 (Sept. 2017), 83:1–83:21. ISSN: 2474-9567. DOI: 10.1145/3130948. URL: http://doi.acm.org/10.1145/3130948 (cit. on p. 25).

[Net17] Netzwelt. *Android-Smartphones: Der große Update-Fahrplan - NETZWELT*. Ed. by J. Kluczniok. Oct. 6, 2017. URL: https://www.netzwelt.de/update-fahrplan/android-smartphones-grosse-update-fahrplan.html (cit. on p. 29).

[OPB00] N. Oliver, A. Pentland, F. Bérard. "LAFTER: a real-time face and lips tracker with facial expression recognition." In: *Pattern Recognition* 33.8 (2000), pp. 1369–1382. ISSN: 0031-3203. DOI: https://doi.org/10.1016/S0031-3203(99)00113-2. URL: http://www.sciencedirect.com/science/article/pii/S0031320399001132 (cit. on p. 15).

[PC97] R. Plutchik, H. R. Conte. *The circumplex as a general model of the structure of emotions and personality.* Jan. 1, 1997 (cit. on p. 19).

[Plu91] R. Plutchik. "The Nature of Emotions." In: *American Scientist, Volume 89* (Jan. 1, 1991) (cit. on p. 20).

[PLZP11] T. Pfister, X. Li, G. Zhao, M. Pietikäinen. "Recognising spontaneous facial micro-expressions." In: *2011 International Conference on Computer Vision*. Nov. 2011, pp. 1449–1456. DOI: 10.1109/ICCV.2011.6126401 (cit. on p. 68).

[Pou16] J. Poushter. *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies | Pew Research Center*. Feb. 22, 2016. URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ (visited on) (cit. on p. 15).

[Sta17a] Statista. *Number of smartphone users worldwide 2014-2020 | Statista*. (Accessed on 10/19/2017). Jan. 1, 2017. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ (cit. on p. 15).

[Sta17b] Statista. *Smartphone penetration worldwide 2014-2019 | Statista*. (Accessed on 10/19/2017). Jan. 1, 2017. URL: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/ (cit. on p. 15).

[SYLT14] S. Shojaeilangari, W.-Y. Yau, J. Li, E.-K. Teoh. "Multi-scale Analysis of Local Phase and Local Orientation for Dynamic Facial Expression Recognition." In: 1 (Jan. 2014) (cit. on p. 21).

[Uni17a] M. L. G. at the University of Waikato. *weka - ARFF*. Oct. 13, 2017. URL: http://weka.wikispaces.com/ARFF (cit. on p. 48).

[Uni17b] M. L. G. at the University of Waikato. *weka - ARFF (stable version)*. Oct. 13, 2017. URL: https://weka.wikispaces.com/ARFF+(stable+version) (cit. on p. 48).

[WBZ+15] E. Wood, T. Baltrusaitis, X. Zhang, Y. Sugano, P. Robinson, A. Bulling. "Rendering of Eyes for Eye-Shape Registration and Gaze Estimation." In: *CoRR* abs/1505.05916 (2015). arXiv: 1505.05916. URL: http://arxiv.org/abs/1505.05916 (cit. on p. 19).

[WH89] N. R. Whitesell, S. Harter. "Children's Reports of Conflict between Simultaneous Opposite-Valence Emotions." In: *Child Development* 60.3 (1989), pp. 673–682. ISSN: 00093920, 14678624. URL: http://www.jstor.org/stable/1130732 (cit. on p. 27).

All links were last followed on October 20, 2017.

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.

---

place, date, signature