EUSTRESS AND DISTRESS DETECTION FROM PHYSIOLOGICAL DATA USING SUPERVISED MACHINE LEARNING

Patrik Hagen

Master Thesis
Aufgabensteller: Prof. Dr. François Bry
Betreuer: Prof. Dr. François Bry, Yingding Wang
Abgabe am: 12. Januar 2017
Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst habe und keine anderen als die angegebenen Hilfsmittel verwendet habe.

München, den 12. Januar 2017

Patrik Hagen
Abstract

An especially important and relevant aspect in the context of effective learning is the stress level of students. Literature suggests that stress can be differentiated between negative stress (distress), which negatively influences the learning and positive stress (eustress), which is even beneficial for the learning success. To date researchers considered the eustress/distress concept mainly from a psychological perspective, where eustress is associated with positive psychological states and distress is associated with negative psychological states.

The main objective of this master thesis is to distinguish eustress from distress based on heart rate data. Therefore the terms computed eustress and computed distress are defined as two categories of stress computationally derived from physiological measures. A classification process is developed, where as eustress or as distress labeled vectors composed of eight heart rate variability features are used to train both a support vector machine and a decision tree classifier. The labels are determined by self-assessment and the features are derived from 10-minute-segments of approximately 60 heart rate values, measured with a fitness tracker. The classification process allows then to classify unlabeled 10-minute-segments of measurement into computed eustress or computed distress by applying the beforehand trained classifiers to the derived features of the segments. Both support vector machine and decision tree classifier have been widely used in the field of stress detection and showed good performances. The results of the stress classification and other interesting aggregated data are visualized in a web portal, in order to give users feedback about their emotional state.

A field study was conducted, where heart rate and survey data were collected from several participants, once during exams and once during the daily life. The performance of the classifiers in the aforementioned classification process was evaluated for various feature combinations with a ten-fold cross-validation. For detecting eustress and distress a maximum recognition accuracy of 76.7% was achieved with the data of all participants in the exam field study together. With the data of a subset of the participants in the daily life field study even a maximum recognition accuracy of 80.2% was achieved.
Acknowledgments

First I want to thank Prof. Dr. Bry, who gave me the opportunity to write this master thesis. He supported my work by providing valuable feedback and making new suggestions, which moved my work clearly forward.

I want to thank also M. Sc. Yingding Wang, who was the supervisor of my master thesis. He was the first contact person for any questions or problems during the project and he always encouraged me to achieve my goals. M. Sc. Yingding Wang was very interested in my work and provided much helpful feedback through our weekly meetings.

Furthermore I want to thank Marcel Heil, who provided the infrastructure for collecting data by developing the CStress app in his bachelor thesis. He also supported me by providing the data and findings of his pilot study. By exchanging information about our projects both of us could benefit from each other.

Lastly I also want to thank all the participants of the field study, who showed great motivation and performed their tasks conscientious.
CONTENTS

3.1.6 Daemons for the preprocessing and classification procedures ........................................ 59
3.1.7 Confirmation and refusal of classified segments ................................................................. 59
3.1.8 Further features .................................................................................................................... 60
3.1.9 Configuration parameters ..................................................................................................... 62
3.2 Front-end ................................................................................................................................. 64
  3.2.1 Implementation of the home screen ...................................................................................... 64
  3.2.2 Implementation of the detail view ....................................................................................... 65
  3.2.3 Implementation of the admin view ...................................................................................... 67
  3.2.4 Data input ........................................................................................................................... 69
  3.2.5 Used technologies ................................................................................................................ 70

4 Evaluation ................................................................................................................................... 71
  4.1 The field study ........................................................................................................................ 71
  4.2 Classification performance ....................................................................................................... 74
  4.3 Pattern analysis ........................................................................................................................ 77
  4.4 Correlation of stress intensity and stress type .......................................................................... 82
  4.5 Evaluation of the daily field study ........................................................................................... 83
  4.6 Summary of the field study findings ......................................................................................... 86

5 Conclusion .................................................................................................................................... 87
  5.1 Summary and discussion ........................................................................................................... 87
  5.2 Limitations and future work ...................................................................................................... 89

Bibliography ..................................................................................................................................... 99
CHAPTER 1

Introduction

1.1 Motivation

Background

Since the number of students at universities is quite high respectively the number of teachers is too low, it is normal that a lecturer teaches to more than one hundred students in a single lecture. This setup is disadvantageous for the teaching quality of mass universities since the lecturers may not be able to care for every single student. As a result, the students have to study a large amount of educational content on their own and take personal responsibility for their learning. Especially at the beginning of the study this usually new situation can overwhelm the students quiet fast and lead to stressful situations. Even in higher semesters too much and complex learning contents, tight schedules, a vast number of exams or also private issues can be the trigger for stress. Since the stress level of students is an especially important and relevant aspect in the context of effective learning, it would be great for lecturers to be able to take advantage of the knowledge about the stress level of students in order to receive a fast and direct feedback about the students emotional state during lectures and exams (or maybe also while learning). Lecturers could then use this feedback in order to adapt their teaching content or to adjust the examination duration accordingly. Such an emotional feedback could also be quite useful for the students themselves, so that they can adjust their learning behavior and find out the causes of their stress. Therefore the basic motivation of this work is to improve the teaching and learning at mass universities by taking the emotional state of students into account.

Previous work on the improvement of teaching and learning

The research project Backstage[1] of the Institute for Informatics at Ludwig-Maximilian University (LMU) in Munich is a first step to improve the teaching and learning at mass universities. As a digital backchannel for large class lectures, Backstage aims at facilitating social interactions and collaboration in a large audience, which is common at mass universities like the LMU. Backstage can be useful for both students and lecturers. It encourages

students, who would otherwise barely participate, to actively engage, even in large lecture classes. At the same time lecturers become aware of the students needs and can ascertain, which topics are difficult to understand for the students. To further improve this system and by association the teaching and learning at mass universities, the stress level and also the type of stress (positive or negative) that students experience during lectures and exams (or maybe also while learning) shall be taken into account in the future.

As a first step towards the consideration of stress in an educational environment as mentioned above, Marcel Heil studied in his bachelor thesis "Conception and Implementation of a Mobile Application with Fitness Trackers as Supportive Tools for Computed Stress Detection" [9] the detection of stress, based on physiological data (precisely the heart rate). In his work he used the collected heart rate measures to predict if someone is stressed or not. But in contrast to many other scientific studies about stress detection, which are mainly conducted in laboratory scenarios with professional equipment, Marcel Heil focused more on real-life scenarios. He collected the heart rate data with fitness trackers in conjunction with smartphones, since this technology is affordable for most people and hence good applicable in everyday life (for example in large lecture classes at mass universities). In addition to the bachelor thesis of Marcel Heil, the type of stress (positive or negative) that someone perceives shall now also be taken into account.

Consideration of the positive effects of stress

As mentioned above, the stress level of students is an especially important and relevant aspect in the context of effective learning. When talking about stress in everyday life, it is usually associated with negative effects. But literature suggests that stress is not inherently negative, but rather that stress can be differentiated between positive stress (called eustress) and negative stress (called distress). A more precise consideration of the eustress / distress concept is presented in section 1.2. So it is not only about whether someone is stressed or not, but also what kind of stress someone experiences. Parker and Ragsdale [24], who studied the influence of eustress and distress on changes in fatigue, stated that eustress can have positive impact on restoring energy, increasing self-efficacy and cognitive processing. With regard to effective learning, this means that whereas distress negatively influences the learning, eustress can be even beneficial for the learning success.

But the consideration of stress as not inherently negative is a rather new approach and did not get much attention in the past (68 search results for “eustress” on PubMed in contrast to 697656 search results for “stress”). This explains that whilst the simple detection of stress (if someone is stressed or not) with physiological measures is a well-studied subject, the literature lacks of studies about the relationship between physiological characteristics and the psychological type of stress (eustress or distress) someone perceives. Hence the motivation of this master thesis is to find physiological characteristics, which are able to separate eustress from distress and build a bridge between physiological stress measures and the psychological perception of eustress and distress. Such knowledge about the emotional state of students could then be used to improve the teaching and learning at mass universities.

1.2. EUSTRESS AND DISTRESS

Application scenarios for feedback about the emotional state

Feedback about the emotional state of students could help lecturers to adapt the teaching content accordingly. Also for the students themselves, a personalized feedback about their emotional state could help to adapt their learning behavior. There are several possible application scenarios of such emotional feedback for lecturers as well as for the students themselves. Lecturers could use the received feedback during class in order to adapt their teaching content accordingly. For example, much distress could be an indicator that the students experience difficulties to follow the lecture. However, eustress would mean that the students get along very well with the teaching subject. Also during exams feedback about the emotional state of students could be helpful. For example, if lots of students experience distress during an exam, it is an indication that the students find it hard to cope with the exam. As a consequence the exam duration could be extended for those, who are extremely under distress. Students could also use the feedback to adapt their learning environment. This means, if students experience distress at a loud and busy place they should think about learning at a calmer location. Another useful application of the emotional feedback could be to schedule study breaks according to the emotional state. For example, if a student is in a distress period, he should make a study break whereas if a student is in an eustress period he should increase the learning intensity. Of course, feedback about the emotional state of people is not limited to the above mentioned application scenarios. Such feedback could be used in every life situation where stress is an issue.

1.2 Eustress and distress

Since just very few stress researchers considered the concept of eustress, the phenomenon of eustress is insufficiently explored in literature. According to Kupriyanov and Zhdanov [15] the lack of research on eustress results from a lack of clear criteria for differentiating eustress from distress, what is due to an insufficient development of the conceptual basis of eustress. So the understanding of eustress varies considerable among scientists.

In the article “Stress without distress” [27] Hans Selye (who was a pioneer in the field of stress research) defined stress as “the nonspecific response of the body to any demand made upon it”. Where a demand (also called stressor) is the physical or psychological stimulus to which individuals respond with stress symptoms [21]. Selye also stated that “anything, pleasant or unpleasant, that speeds up the intensity of life, causes a temporary increase in stress”, so that “a painful blow and a passionate kiss can be equally stressful” [27]. This means that stress can be perceived as either positive or negative. In the course of Selye’s formulation of the general adaptation syndrome, where he emphasized the adaptive nature of the reaction to the stress, Selye introduced the terms “eustress” and “distress” as two different types of stress in order to distinguish the adaptive and non-adaptive effects to stress reactions [15]. Where eustress was identified as “healthy, positive, constructive results of stressful events and stress response” [15]. So according to Selye, eustress is considered as “the result of the body’s response to a stressor” [15].

Another view on the eustress / distress concept was proposed by Lazarus [17]. Lazarus considered eustress as a positive cognitive response to a stressor. In his stress model eustress is associated with positive emotions and a healthy physical state, whereas distress is associated with negative emotions and physical impairments. So whereas Selye considered stress from a physiological point of view (eustress as the result of the body’s response to a stressor), Lazarus considered stress from a more psychological point of view (eustress as a
positive cognitive response to a stressor).

Nelson and Simmons [21] who considered stress also from a psychological perspective believed that "eustress can best be conceptualized by identifying it as positive aspects of the stress response itself". They assumed demands or stressors to be inherently neutral and that only the response to demands or stressors has positive or negative valence dependent on the degree of attraction or aversion an individual experiences. In their stress model Nelson and Simmons [21] defined eustress as "a positive psychological response to a stressor, as indicated by the presence of positive psychological states” and distress as "a negative psychological response to a stressor, as indicated by the presence of negative psychological states”. As examples for positive psychological states they proposed inter alia hope, meaningfulness, manageability, satisfaction, commitment and positive affect and as examples for negative psychological states they mentioned inter alia anger, frustration, burnout, anxiety and negative affect. So they believe that "stressors can be perceived not only as threats but also as challenging or positive” [21].

Even though the three presented approaches towards eustress differ in their consideration perspective (Selye considers stress from a more physiological point of view whereas Lazarus and Nelson and Simmons focus on the psychological aspect of stress) they have in common that eustress is associated with positive emotions and distress is associated with negative emotions. This suggests that the mood of a person is responsible for perceiving a stressor as pleasant or unpleasant.

Another approach towards eustress is not to associate it with positive emotions, but rather to consider eustress as the ideal amount of stress. In this theory it is assumed that there is a nonlinear relationship between the stress intensity and the productivity of a person in such a way that stress is beneficial for the productivity until an ideal amount of stress is reached. After this peak the productivity declines with increasing stress intensity. For this theory researchers rely on the Yerkes-Dodson Law[3] which illustrates this relationship in the form of an inverted U-shaped diagram. [15] This theory suggests that eustress is associated with a rather moderate stress intensity whereas distress is associated with a high stress intensity.

Figure 1.1 shows the suggested placement of eustress in a circumplex model of emotions. The y-axis depicts the level of arousal (stress intensity) and the x-axis depicts the valence of the emotions, so that the center of the diagram represents neutral valence and medium arousal. In this model distress (red surrounded in figure 1.1) is located on the upper left, what means that distress has a high level of arousal and negative valence. Based on the aforementioned theories about eustress (in one aspect eustress associated with positive emotions, in another aspect eustress as the ideal amount of stress) it is hypothesized by the author of this work that eustress should be placed on the upper right of the model (green arrow in figure 1.1), but not as high as distress. So it is suggested that eustress has a rather moderate level of arousal (or at least not as high as distress) and positive valence.

Even though Nelson and Simmons [21] focused on the psychological aspect of stress they mention that the positive and negative stress responses can have psychological, behavioral and also physiological indicators. In conjunction with Selye’s definition that stress is the response of the body to a stressor [27], this suggests that it is possible to distinguish eustress from distress by means of physiological characteristics.

1.3 Measuring stress

Stress can be measured in various ways. The traditional way to measure stress is that humans rate the stress level on some scale. This can be done by interviewing people, letting them fill in questionnaires or by observing their behavior. The issue with these assessments is that they are all subjective and require significant human intervention.

In order to obtain objective information about the stress level of students, their stress symptoms have to be measured somehow. Stress symptoms can be hormonal imbalance and physiological and physical changes. For example, in stressed situations the amount of hormones in the body (e.g., the cortisol level) increases. In order to obtain measures for these hormones, invasive methods are necessary. This means that professional practitioners, e.g., have to take blood or urine samples and lengthy analyses by qualified scientists have to be done. On the other hand, physiological and physical measures (also called primary measures) can be obtained via non-invasive methods, which makes it much easier and faster to gather measurements and analyze them. Nevertheless, most of the methods to measure physical or physiological stress symptoms are still quite intrusive and not applicable in everyday life. Even though physical symptoms for stress are defined as properties that can be seen by humans without the need for special equipment, sophisticated tools using vision and audio technologies are needed to be able to measure physical stress features.
such as eye gaze, pupil diameter, voice characteristics or hand and finger movements. The measurement of physiological stress symptoms such as heart activity, brain activity, skin conductivity or skin temperature requires equipment that needs to be attached to the body. This equipment can be more or less intrusive.

In this work physiological measures (namely heart rate measures) are used to determine the stress level. The most precise and commonly used method to measure the heart rate is with an electrocardiogram (ECG), where the single heartbeats can be identified exactly. However, an ECG is quite intrusive and not applicable for a large amount of users in everyday life, since the equipment is quite complex and expensive. But with the upcoming trend of wearing fitness trackers, there is the possibility for everyone to measure the heart rate in everyday life with a non-intrusive (or at least less intrusive) method. The obtained data of fitness trackers are admittedly not as precise as the data of an ECG, but on the other hand such fitness trackers are much less expensive and complex. Heart rate measures of such fitness trackers are used as the data basis in this work, since the developed technology shall be accessible for a wide range of people.

![Example of an ECG equipment](image1.png) ![Fitbit Charge HR fitness tracker](image2.png)

Figure 1.2: Comparison of an ECG equipment and a fitness tracker.

1.4 Related work

There are a plenty of studies that validate the relationship between physiological characteristics and stress, where stress is commonly considered as inherently negative (distress). A commonly used method to quantify stress is the analysis of the heart rate variability (HRV) based on ECG measurements.

For example Taelman et al. analyzed the changes in heart rate and HRV for a group of 28 subjects at rest and with a mental stressor in a laboratory environment. The results of their study suggest that the heart rate and the HRV change with a mental task. Precisely they concluded, that the mean RR-interval and the pRR50 were both significantly lower for the mental task than in the rest condition, whereas the standard deviation of the RR-intervals did not differ significantly between the two conditions. In the frequency domain analysis

5https://www.fitbit.com/de/shop/chargehr
the LF/HF ratio increased for the mental task, but not significantly.

Curic et al. [4] examined if the HRV is lower in stress conditions than in rest conditions. In order to do this, the heart rate of 28 subjects was recorded during rest and stress conditions and four HRV features (namely the mean heart rate, the mean RR-interval, the RMSSD and the pRR50) were derived from the obtained heart rate measures. The results of the study confirmed their hypothesis. The mean RR-interval, the RMSSD and the pRR50 were all higher during the rest conditions than during the stress conditions whereas the mean HR was higher in the stress conditions and lower in the rest conditions. And they concluded that the HRV is a reliable indicator for stress.

Zhai and Barreto [36] used multiple physiological features to determine the stress level of subjects. In a laboratory study they tried to elicit emotional stress from subjects by applying a computer-based “Paced Stroop Test”, where the subjects have to select the font color of a word shown on the screen that names a potentially different color. 32 subjects were asked to complete congruent stages (matching color name and font color) to elicit normal states and incongruent stages (mismatching color name and font color) to elicit stress states. The physiological signals recorded during the experiment were the galvanic skin response (GSR), the blood volume pulse, the pupil diameter and the skin temperature. Several features derived from these signals were fed into three learning systems, in order to classify normal states and stress states. The used learning algorithms were a naive Bayes classifier, a decision tree classifier and a support vector machine (SVM). The evaluation of the classifiers was performed with a 20-fold cross-validation and yielded to an accuracy of 78.65% with the naive Bayes classifier, 88.02% with the decision tree classifier and 90.1% with the SVM.

Different from Zhai and Barreto, Healey and Picard [8] conducted their experiment in a real world scenario. They collected and analyzed physiological data (Electrocardiogram, electromyogram, skin conductance and respiration) from 24 subjects during real world driving tasks and tried to determine the driver’s stress level in three different settings (rest for low levels of stress, highway driving for medium levels of stress and city driving for high levels of stress). Features derived from 5-minute intervals of data were used to distinguish the three levels of stress. With the recognition algorithm an overall accuracy of 97.4% could be reached across multiple drivers and driving days. The results showed also that heart rate and skin conductivity measures provided the highest overall correlations with the stress level of the drivers.

Since the measuring of GSR data requires obtrusive equipment, Liu and Ulrich [19] presented a model to predict stress levels only with ECG data (precisely with time domain and frequency domain features of the HRV and with the spectral power components derived from the raw ECG signal). They used the data from the previous presented study by Healey and Picard [8] to perform a binary classification of stress and rest periods. For the binary classification they used a Gaussian naive Bayes classifier, a SVM with various kernels, a k-nearest neighbor classification and a random forest classifier. With only HRV features and a Gaussian naive Bayes classifier they achieved an accuracy of 78% on the most distinct samples. The best performance (98% on the most distinct samples and 85% over all samples) was achieved with ECG frequency features and HRV features together and while using a linear SVM.

The scientific community is not in agreement about if there exists a relation between different emotions and specific physiological characteristics and, if so, in what form they exist. On one extreme, there is the position that it is not possible to find unique and invariant
physiological characteristics for different kinds of emotions, but that conditions of threat and challenge can be distinguished, as well as positive versus negative affect. On the other side it is argued that why there should not be specific physiological patterns for emotions, if those have specific functions for human adaption. [14]

Rainville et al. [25] studied the profile of cardiorespiratory activity during four distinct emotions, namely fear, anger, sadness and happiness. For the study an [ECG] and the respiratory activity was recorded from 43 subjects during the recall of emotional autobiographical episodes. They concluded that the cardiorespiratory activity differs on several linear and spectral indices between the four emotions and a neutral control condition. With a stepwise discriminant analysis they could predict the emotions with a classification rate of 65.3% for all four emotions and of 72.0% - 83.3% for pair-wise differentiation.

Kop et al. [13] considered just two distinct emotions (happiness and anger) in order to study how positive and negative emotions influence the autonomic nervous system and particularly the [HRV]. They used 5-minute event recall tasks (happiness and anger recall) and a 5-minute Stroop Color Word Test (SCWT) to elicit positive and negative mood from 20 healthy subjects. During the emotion recall an [ECG] was recorded, in order to derive [HRV] frequency domain features (low frequency and high frequency) from 1-minute segments. Their study revealed that positive mood induction by the happiness recall task results in increased LF-HRV, increased LF/HF ratio and an increased heart rate, whereas the effect on HF-HRV was not significant. The negative mood induction by the anger recall had no significant effects on the [HRV] features. The negative mood induction by the SCWT task yielded to a decrease in HF-HRV and LF-HRV and an increased heart rate, whereas the LF/HF ratio did not show any significant changes. Even though they found no direct evidence for a generalized relationship between the autonomic nervous system activity and positive or negative mood induction, Kop et al. concluded that positive and negative mood induction result in different [HRV] responses.

All these presented studies considered either general stress recognition or the distinction of several emotions without regarding the stress level. To the best of the author’s knowledge the work of Li et al. [18] is the first, which is proposing a classification model towards eustress. They studied the feasibility of measuring eustress by means of [HRV] smartphone and computer usage data. To do this, they collected heart rate data (with a wearable heart rate sensor), data about the smartphone and computer usage and data from an hourly self-assessment about the stress level, mood and performance of seven subjects on five days during their waking hours. For the eustress recognition they defined eustress in twofold: Eustress as the combination of moderate stress with high performance and eustress as the combination of moderate stress with high mood. The classification was then performed with a random forest classifier, a multinomial logistic regression and a [SVM] and achieved a maximum accuracy of 71.33% for eustress as an urge for better performance and 57.34% for eustress as a state of better mood.

1.5 Objective

To date, the distinction of eustress and distress was primarily considered from a psychological point of view. To the best of the author’s knowledge, Li et al. [18] were the first who tried to measure eustress explicitly by means of physiological characteristics, namely the heart rate variability. This work is attended to further investigate the relationship between physiological measures and eustress and distress, in order to bridge from the physiology to
the psychological state (perception of eustress and distress).

But since there is no clear criterion for the distinction of eustress and distress, the first question that has to be tackled is, how to define eustress and distress, so that it is clearly distinguishable. According to the Yerkes-Dodson Law, Li et al. [18] assumed eustress to be under a moderate physiological stress level. However it is believed that this assumption does not have to be necessarily true, but rather it should be studied if the stress intensity influences the perception of eustress and distress. So for the purpose of this work eustress and distress are defined according to Nelson and Simmons, where eustress is “a positive psychological response to a stressor, as indicated by the presence of positive psychological states” and distress is “a negative psychological response to a stressor, as indicated by the presence of negative psychological states”. Since this definition was made from a psychological point of view, it should be transferred to a more physiological definition. So in this work eustress is assumed to be a self-reported pleasant mood along with the experience of physiological stress and distress is defined as a self-reported unpleasant mood along with the experience of physiological stress.

Sharma and Gedeon [28] defined the term computed stress as “the stress computationally derived from instantaneous measures of stress symptoms obtained by non-invasive methods”, where “a computational model of stress will take some combination of stress symptom measures as inputs to produce a computed stress measure as an instantaneous measure of stress at that point in time”.

According to this definition, the terms computed eustress and computed distress are defined as two distinct types of stress computationally derived from physiological measures, in order to delineate the computationally derived eustress and distress from the psychological definition of eustress and distress in section 1.2. **Computed eustress** is defined as the physiological response to a stressor that was classified as eustress by means of computed stress measures and **computed distress** is defined as the physiological response to a stressor that was classified as distress by means of computed stress measures.

So the objective of this master thesis is to distinguish eustress from distress based on physiological data (precisely the heart rate) in order to provide feedback about the emotional state of students.

With this end in mind, the further structure of this master thesis is as follows:
First a pre-study is conducted and the findings are presented to develop a model for the distinction of computed eustress and computed distress. Then the developed model and a web portal are implemented, in order to support the collection and analysis of further data and provide feedback about the emotional state of users. The implemented model and the findings of the pre-study are then evaluated in a field study. Finally a conclusion of the work is presented and further work for improvements is suggested.
2.1 Computed stress model (the heart rate variability)

2.1.1 The heart rate variability

The autonomic nervous system (ANS) is responsible for involuntary body activities and regulates the physiological activities of the body, for example the heart activity, the skin conductivity, the blood pressure, the skin temperature and the respiration [28]. The ANS is made up of two separate, mostly contrary working sub-systems, called the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS has an activating effect and is responsible for the fight-or-flight response of the body to a stressor, whereas the PNS is predominant active in periods of rest and allows primarily regeneration processes [4]. This means that in stress conditions the SNS activity increases and the PNS activity decreases (the SNS dominates), whereas in rest conditions the PNS activity increases and the SNS activity decreases (the PNS dominates) [28].

Since the ANS regulates inter alia the heart activity, measuring the heart rate is an ideal, non-invasive method to determine the state of the ANS [31]. Indeed, the heart rate is one of the main measures for stress used in literature [28]. Both the SNS and the PNS have impact on the heart activity. The low-frequency, slow impulses of the SNS lead to an increasing heart rate and a more regular heartbeat interval, whereas the high-frequency, fast impulses of the PNS lead to a decreasing heart rate and a more irregular heartbeat interval [4].

The heart rate variability (HRV) is a measurement parameter for the autonomic function of the heart and thus for the activity of the ANS. It is a physiological parameter that characterizes the adaptability of the human organism to inner and outer loading factors, such as stressors and describes the heart’s capability to change the heartbeat interval continuously according to the current load [4]. In more detail, the HRV is the variation between consecutive heartbeat intervals and describes the balance between sympathetic and parasympathetic activities [3]. So it can be adhered, that in relaxation and rest periods the heartbeat intervals of an organism are very irregular and thus the HRV is high, whereas in periods of exertion and stress the heartbeat intervals are more regular and thus the HRV is low [4].
In order to determine the heart rate variability several HRV features can be calculated. These HRV features are derived from time series of consecutive RR-intervals, where a RR-interval is the time between two consecutive heartbeats. Typically the RR-intervals are extracted from ECG recordings that are tracing the electrical activity generated by the heart. An ECG recording (illustrated in figure 2.1) consists of a P wave (representing atrial depolarization), a QRS complex (representing ventricular depolarization) and a T wave (representing the rapid repolarization of the ventricles). The time interval between two consecutive R peaks in the ECG signal is the RR-interval. In the literature RR-intervals are also called interbeat interval (IBI) or NN-interval (normal-to-normal-interval).

![Figure 2.1: Sample of an ECG signal with P wave, QRS complex and T wave.](image)

### CHAPTER 2. CONCEPTION

#### 2.1.2 Heart rate variability features

There are two methods that can be used to analyze RR-interval time series and thereby determine the HRV features. These methods are the time domain analysis and the frequency domain analysis, which are addressed in the following sections.

##### 2.1.2.1 HRV features for time domain analysis

In the time domain analysis the HRV features are directly derived from RR-interval values. This means that a time series of consecutive RR-intervals is used to calculate the various HRV features. According to Boonnithi some of the most promising HRV features in the time domain are the mean heart rate (meanHR), the standard deviation of the heart rate (SDHR), the mean RR-interval (meanRR), the standard deviation of the RR-intervals (SDRR), the coefficient of variance of the RR-intervals (CVRR), the root mean square successive difference (RMSSD) and the number of pairs of adjacent RR-intervals differing by more than 50 ms to all RR-intervals (pRR50). With the approximate entropy (ApEn) and the variability score (which is a self-made feature) two more time domain features are considered in this work. In the following, these HRV features and their relation to stress are described in more detail.
In contrast to other time domain features the meanHR and the SDHR are derived from time series of heart rate values and not from RR-intervals. The heart rate values can be easily calculated from the RR-intervals since the heart rate and the RR-intervals are reciprocals of each other [11]. So having an RR-interval in milliseconds, the corresponding heart rate value is calculated by \( HR_i = \frac{60000}{RR_i} \) and has units of beats per minute (bpm). The remaining time domain features are all derived directly from the RR-intervals, which were defined previously. With that said, the HRV features are derived from time series of \( N \) consecutive RR-intervals or heart rate values as follows.

**meanHR and meanRR (equation 2.1 and 2.2):**

The mean value of the consecutive heart rate values (in beats per minute) and the mean value of the consecutive RR-intervals (in milliseconds) [3]. The fact that the SNS (active under acute stress) increases the heart rate and the PNS (active in rest conditions) decreases the heart rate [4] implies that higher meanHR values indicate stress and lower meanHR values are an indicator for rest. This implication is in line with study results of [31], where the mean heart rate was a reliable feature to detect mental stress. Since the RR-intervals and the heart rate values are reciprocals of each other the statements about the mean heart rate can be easily transformed for the meanRR as follows. Under the influence of the SNS the RR-intervals decrease and under the influence of the PNS the RR-intervals increase, so that lower meanRR values indicate stress and higher meanRR values are an indicator for rest. The study results of Sun et al. [31] and Taelman et al. [32] also support this assertion and denote the meanRR as a reliable measure for mental stress detection.

\[
\text{meanHR} = \frac{\sum_{i=1}^{N} HR_i}{N} \tag{2.1}
\]
\[
\text{meanRR} = \frac{\sum_{i=1}^{N} RR_i}{N} \tag{2.2}
\]

**SDHR and SDRR (equation 2.3 and 2.4):**

The standard deviation to the mean value of the consecutive heart rate values (in beats per minute) and the standard deviation to the mean value of the consecutive RR-intervals (in milliseconds) [3]. Due to the fact that the standard deviation measures the variation of the values in the heart rate respectively RR-interval time series, it would be reasonable that both lower SDHR values and lower SDRR values indicate a low HRV (regular heartbeats) and thus stress, whereas higher values of the SDHR and SDRR indicate a high HRV (irregular heartbeats) and thus rest. This assumption is encouraged by the statement of Bong et al. [37] that higher SDRR values mean a higher HRV. But both the study results of Sun et al. [31] and Taelman et al. [32] did not present the standard deviation as a reliable feature to detect stress. This could be the case because a high standard deviation does not necessary mean that the values have a high variability, but rather only that the values have a high distance to the mean value.

\[
\text{SDHR} = \sqrt{\frac{\sum_{i=1}^{N} (HR_i - \text{meanHR})^2}{N - 1}} \tag{2.3}
\]
\[
\text{SDRR} = \sqrt{\frac{\sum_{i=1}^{N} (RR_i - \text{meanRR})^2}{N - 1}} \tag{2.4}
\]
CVRR (equation 2.5):

The coefficient of variance of the consecutive RR-intervals (normalization of the standard deviation) [3]. According to the study of Boonnithi [3], who investigated which HRV features perform best in stress detection, this feature does not belong to the effective ones. But this is plausible, since the CVRR is a normalization of the SDRR and it seems that the standard deviation is not a reliable feature to detect stress either [31, 32]. As a result, this feature is not commonly used in the literature to detect stress.

\[
CVRR = \frac{SDRR \times 100}{meanRR}
\]  

RMSSD (equation 2.6):

The root mean square successive difference of the consecutive RR-intervals (in milliseconds) [3]. In contrast to the standard deviation the RMSSD considers the diversity of the successive RR-intervals and thereby reflects the influence of the PNS on the heart rate. Higher RMSSD values imply a higher parasympathetic influence and thus indicate rest, whereas lower RMSSD values imply a lower parasympathetic influence and thus are an indicator for stress. Furthermore the RMSSD is a commonly used HRV feature in the field of stress detection. [4]

\[
RMSSD = \sqrt{\frac{\sum_{i=1}^{N-1} (RR_{i+1} - RR_i)^2}{N-1}}
\]  

pRR50 (equation 2.7):

The number of pairs of adjacent RR-intervals differing by more than 50 ms to all RR-intervals (in percent) [3]. The pRR50 reflects the influence of the PNS on the heart rate in the same manner as the RMSSD feature [4], so that higher pRR50 values indicate rest and lower pRR50 values indicate stress. The study results of Taelman et al. [32] also revealed that the pRR50 was significantly higher while rest conditions than under stress.

\[
pRR50 = \frac{\text{count}(|RR_{i+1} - RR_i| > 50\text{ms}) \times 100}{N-1}
\]  

Approximate entropy:

The approximate entropy (ApEn) is a measure for the complexity or irregularity of a time series of consecutive RR-intervals. Intuitively said, the ApEn analyzes if a given time series of consecutive RR-intervals has rather many or few repetitive patterns. Higher values of ApEn stand for a high irregularity in the time series (few repetitive patterns) and lower values of ApEn stand for a more regular time series (many repetitive patterns). [33, 20]

This implies that higher ApEn values indicate a high HRV and thus rest and lower ApEn values are an indicator for a low HRV and thus indicate stress. The study results of Valenza et al. [33] also support this implication that the ApEn decreases, when switching from rest to stress conditions. The calculation of the ApEn requires several computation steps and
works as follows (according to [33]).

First the RR-interval time series of length \( N \) is divided up into \( N - m + 1 \) vectors \( u_j \) of length \( m \), where \( j = 1, 2, ..., N - m + 1 \):

\[
    u_j = (RR_j, RR_{j+1}, ..., RR_{j+m-1})
\]  
(2.8)

Then for each vector \( u_j \) the relative number of vectors \( u_k \) with \( d(u_j, u_k) \leq r \) is determined (see equation 2.9), where \( r \) is the tolerance value for accepting matches and \( d(u_j, u_k) \) is the distance between two vectors.

\[
    C^m_j(r) = \frac{\text{nbr of } \{u_k|d(u_j, u_k) \leq r\}}{N-m+1} \quad \forall 1 \leq k \leq N-m+1
\]  
(2.9)

The distance \( d(u_j, u_k) \) is defined as the maximum absolute difference between the corresponding elements:

\[
    d(u_j, u_k) = \max_{n=0,1,...,m-1} \{|RR_{j+n} - RR_{k+n}|\}
\]  
(2.10)

For the next step it is important to notice that the value of \( C^m_j(r) \) is always lower than or equal to 1 (due to the normalization) and greater than 0 (because self matches are also included into the count). This fact guarantees that the natural logarithm is defined for each \( C^m_j(r) \) and so \( \Phi^m(r) \) can be calculated as the average over \( j \):

\[
    \Phi^m(r) = \frac{\sum_{j=1}^{N-m+1} \ln(C^m_j(r))}{N-m+1}
\]  
(2.11)

The ApEn can then be computed as:

\[
    \text{ApEn}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r)
\]  
(2.12)

Since the parameters \( m \) and \( r \) have a strong effect on the ApEn they should be chosen carefully. According to Valenza et al. [33], the tolerance \( r \) should be a fraction of the standard deviation of the RR-intervals and is commonly set to \( r = 0.2 \times \text{SDRR} \). This is also true for this work. The parameter \( m \) was set to 3 in this work. Although Valenza et al. [33] suggest to set \( m = 2 \), during this work it has emerged that setting \( m = 3 \) is more appropriate to distinguish eustress from distress. Also the length \( N \) of the RR-interval time series has an influence on the ApEn value, since the ApEn approaches its asymptotic value as \( N \) increases [33].

**Variability score (equation 2.13):**

The variability score \( (\text{VS}) \) is a self-made feature. It was constructed in order to get more insight into the pattern of RR-interval time series. The VS counts how often switch over the RR-intervals with respect to the mean value, where switch over is detected in case one of two successive RR-intervals is over the mean value and the other is under the mean value.

\[
    \text{VS} = \text{count}(\text{RR}_i | (\text{RR}_i > \text{meanRR} \land \text{RR}_{i+1} < \text{meanRR}) \lor (\text{RR}_i < \text{meanRR} \land \text{RR}_{i+1} > \text{meanRR}))
\]  
(2.13)

Figure 2.2 illustrates the calculation of the VS in a short example. The RR-interval time series in figure 2.2a also show the difference between a high and a low VS. Whereas a higher VS (figure 2.2a) is an indicator for an irregular heartbeat and therefore a high HRV, a lower VS (figure 2.2b) indicates a more regular heartbeat and hence a low HRV. This suggests that higher VS values are an evidence for rest and lower VS values are an evidence for stress.
2.1.2.2 HRV features for frequency domain analysis

In the frequency domain analysis the HRV features are derived from a power spectrum, which is obtained by transforming the consecutive RR-interval values from time domain to frequency domain [3]. This is done by decomposing the spectral components of the HRV in base frequencies using mathematical frequency analysis methods like the fast Fourier transform. As a result the relative share of the various base frequencies with regard to the total spectrum are obtained [4].

The following three features are the basic measures of the frequency domain analysis:

- **Very low frequency (VLF):** The power spectrum of very low frequency (from 0.003 to 0.04 Hz) [3]. According to the Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology [22] the VLF obtained from short-term recordings is an unreliable measure and should not be used when interpreting the frequency domain of short-term ECGs. For that reason this measure will not be further considered in this work.

- **Low frequency (LF):** The power spectrum of low frequency (from 0.04 to 0.15 Hz) [3]. The LF domain is predominantly controlled by the SNS, but contains also influences of the PNS [4]. But nevertheless the LF domain is commonly used as a measure of sympathetic activities [13].

- **High frequency (HF):** The power spectrum of high frequency (from 0.15 to 0.4 Hz) [3]. The HF domain is controlled by the PNS and reflects the parasympathetic activities relatively unbiased [13].

In addition there are several features that can be derived from these three basic measures, such as for example the normalized VLF, normalized LF and normalized HF. For more details about these and other features derived from the basic measures see [3]. But there is one more feature that has to be considered in more detail here, and this is the sympathovagal balance index (SVI).
2.2. PRE-STUDY AND FINDINGS

The SVI (equation 2.14) represents the ratio of low-frequency impacts (predominantly caused by the SNS) to high-frequency impacts (caused by the PNS). Since the SNS is dominant under stress conditions and the PNS is dominant while periods of rest, it is assumed that lower SVI values (high-frequency parasympathetic influences dominate) indicate rest and higher SVI values (low-frequency sympathetic influences dominate) indicate stress.

\[ SVI = \frac{LF}{HF} \]  
(2.14)

But in this work no frequency domain analysis shall be performed. The reason why no frequency domain analysis is used, is because the heart rate measurements (which are obtained from fitness trackers in this work) are not very precise (only about one measurement every 5 to 15 seconds). Thus a frequency domain analysis would not be very convincing. Furthermore the abandonment of a frequency domain analysis saves processing time, since no Fourier transform has to be performed. To be able to use the SVI feature even if there is no frequency domain analysis available, Wang and Huang [34] described a method to calculate a surrogate for the SVI using the time domain features SDRR and RMSSD (equation 2.15). This surrogate is used in the further course of this work to estimate the SVI and is named SVIsurrogate.

\[ SVIsurrogate = \frac{SDRR}{RMSSD} \]  
(2.15)

2.2 Pre-study and findings

First of all a pre-study was conducted, in order to determine if it is possible to distinguish eustress from distress with physiological data and which HRV features are best qualified for this purpose. In the following sections first the used data set, the data preprocessing and the analysis procedure are illustrated. Then the findings and the opportunities and limitations of the pre-study are presented.

2.2.1 The pre-study data set

The collected data of Marcel Heil’s pilot study [9] were used as data basis for this pre-study. In order to collect the heart rate data 5 participants were equipped continuously with a fitness tracker over a period of seven days. Whilst the study the participants were asked to document their activities in slots of one hour and if they felt stressed or not during these activities. The participants could state their stress level in three levels (not stressed, slightly stressed or very stressed). One of the participants (in the following called user 1) also documented the type of stress he experienced. This means that this participant additionally recorded if he experienced rather positive stress (eustress) or negative stress (distress). So at the end of the seven day long user study for every participant there is a list of the collected heart rate values along with the corresponding timestamps (see figure 2.3) and a list with the documented stress levels (see figure 2.4a). Additional, for user 1 there is a list where the eustress and distress is documented (see figure 2.4b). Because the intention of the pre-study is to find a way to distinguish eustress from distress, only the data of user 1 (who also documented the stress type) is of interest. Therefore only the data of user 1 is considered during the further procedure.
CHAPTER 2. CONCEPTION

Figure 2.3: A snippet of the heart rate list. The first column contains the heart rate value in bpm, the second column contains the timestamp of the record in milliseconds, the third column contains the date of the record and the fourth column contains the time of the record.

<table>
<thead>
<tr>
<th>heart rate (bpm)</th>
<th>timestamp (ms)</th>
<th>date</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>89</td>
<td>1457954678715</td>
<td>Mar 14, 2016</td>
<td>12:24:38 PM</td>
</tr>
<tr>
<td>88</td>
<td>1457954799796</td>
<td>Mar 14, 2016</td>
<td>12:26:39 PM</td>
</tr>
<tr>
<td>91</td>
<td>1457954924593</td>
<td>Mar 14, 2016</td>
<td>12:28:44 PM</td>
</tr>
<tr>
<td>79</td>
<td>1457955099125</td>
<td>Mar 14, 2016</td>
<td>12:31:39 PM</td>
</tr>
<tr>
<td>88</td>
<td>1457955222467</td>
<td>Mar 14, 2016</td>
<td>12:33:42 PM</td>
</tr>
<tr>
<td>89</td>
<td>1457955456494</td>
<td>Mar 14, 2016</td>
<td>12:37:36 PM</td>
</tr>
<tr>
<td>102</td>
<td>1457955585472</td>
<td>Mar 14, 2016</td>
<td>12:39:45 PM</td>
</tr>
<tr>
<td>84</td>
<td>1457955717455</td>
<td>Mar 14, 2016</td>
<td>12:41:57 PM</td>
</tr>
<tr>
<td>93</td>
<td>1457955918570</td>
<td>Mar 14, 2016</td>
<td>12:45:18 PM</td>
</tr>
<tr>
<td>82</td>
<td>1457956104471</td>
<td>Mar 14, 2016</td>
<td>12:48:24 PM</td>
</tr>
</tbody>
</table>

Figure 2.4: A snippet of the documented stress level and stress type per day. Every row represents one hour. (a) shows a snippet of the documented stress level per day, where -1 = not labeled, 0 = not stressed, 1 = slightly stressed and 2 = very stressed. (b) shows a snippet of the documented stress type per day, where -1 = not labeled, 0 = distress and 1 = eustress.
2.2. Data preprocessing

Before it is possible to analyze the collected data, they first have to be preprocessed. This involves linking the heart rate data with the corresponding stress level and stress type (labeling the data), segmenting the resultant data set into appropriate RR-interval time series and calculating the features for these segments.

2.2.2.1 Data transformation and labeling

As input for the first preprocessing step serve the three data sets introduced in section 2.2.1. This is the heart rate list (figure 2.3), the documented stress level (figure 2.4a) and the documented stress type (figure 2.4b). First the unit of the heart rate timestamps are converted from milliseconds to seconds and the heart rate list is ordered ascending according to the timestamps. The conversion can be done without the loss of information because the collected heart rate data have a sample rate of at least one second (usually the sample rate is even between 5 and 15 seconds). Then for every entry in the heart rate list the corresponding stress level and stress type is determined by looking up the hourly documented values for a given timestamp. At the end of this preprocessing step the three input data sets are linked, so that a stress level and a stress type is assigned to every heart rate value with timestamp (see figure 2.5).

<table>
<thead>
<tr>
<th>timestamp (sec)</th>
<th>heart rate (bpm)</th>
<th>labeled stress level</th>
<th>labeled stress type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1458428410</td>
<td>78</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428440</td>
<td>73</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428470</td>
<td>75</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428525</td>
<td>74</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428555</td>
<td>72</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428605</td>
<td>71</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428640</td>
<td>76</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428670</td>
<td>79</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428700</td>
<td>78</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1458428735</td>
<td>129</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2.5: A snippet of the output of the first preprocessing step. The first column contains the timestamp of the heart rate record in seconds, the second column contains the heart rate value in bpm, the third column contains the documented stress level at this time and the fourth column contains the documented stress type at this time.

2.2.2.2 Data segmentation

In the second preprocessing step the output data set of the previous preprocessing step is segmented into disjoint time series of 60 consecutive RR-intervals. The RR-intervals are obtained by taking the inverse of the corresponding heart rate value multiplied with 60 \((RR_i = \frac{60}{HR_i})\) and have units of seconds. For the extracted segments two important restrictions are applied:
1. Every segment has to be in a maximum time range of ten minutes. This means that if the next 60 RR-interval values are not within a time span of 10 minutes or less, no segment is created for this RR-interval time series. If this is the case, the first value of the considered time series is skipped and it is tried to create a segment with the next 60 RR-intervals. This restriction was implemented so that all segments are in a reasonable time span and thus are comparable with each other, what would not be the case if one segment would be 5 minutes long and another would comprise one hour.

2. The start point and end point of a segment have to be within the same hour. This restriction is necessary because the stress level and the stress type are documented per hour. So it is ensured that an explicit stress level and an explicit stress type can be assigned to every segment.

Once the data set is segmented, all segments that have a stress level lower than 1 (marked as not stressed) are filtered out, so that just segments that are marked as stressed (stress level greater than 0) are left over. Then the remaining segments are labeled with the corresponding stress type ("eu" for eustress and "di" for distress) that was documented by the user. At the end of this preprocessing step the data set is divided up into a set of segments, which contain disjoint time series of 60 consecutive RR-intervals and the corresponding stress type. Furthermore every segment has a unique start time (the timestamp of the first RR-interval) and a length (the time span between the first and the last RR-interval of the segment). Figure 2.6 illustrates the structure of the segments.

2.2.2.3 Feature extraction

After the data set was divided up into segments, now for each segment the several HRV features have to be calculated. For this pre-study 10 HRV features were calculated, namely the meanHR, the SDHR, the meanRR, the SDRR, the CVRR, the RMSSD, the pRR50, the ApEn, the VS and the SVIsurrogate. All of these features are measures from the time domain analysis and thus were derived directly from the RR-interval time series, without the need for a Fourier transform. This is also true for the SVIsurrogate since it is approximated through time domain features. For details about the calculation of the used features see section 2.1.2. At the end of this preprocessing step for each segment all features are calculated and the data set is ready for the analysis. The structure of the segments looks now as illustrated in figure 2.7.

---

**Figure 2.6: Structure of the labeled segments.**

<table>
<thead>
<tr>
<th>timestamp (sec)</th>
<th>length (sec)</th>
<th>rInterval_0</th>
<th>rInterval_1</th>
<th>...</th>
<th>rInterval_59</th>
<th>labeled stress type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1458212530</td>
<td>345</td>
<td>0.66</td>
<td>0.69</td>
<td>...</td>
<td>0.98</td>
<td>eu</td>
</tr>
<tr>
<td>1458212880</td>
<td>340</td>
<td>0.93</td>
<td>0.91</td>
<td>...</td>
<td>0.54</td>
<td>di</td>
</tr>
</tbody>
</table>

**Figure 2.7: Structure of the segments after the features were calculated.**

<table>
<thead>
<tr>
<th>timestamp (sec)</th>
<th>length</th>
<th>mHR</th>
<th>SDHR</th>
<th>mRR</th>
<th>SDRR</th>
<th>CVRR</th>
<th>RMSSD</th>
<th>pRR50</th>
<th>ApEn</th>
<th>VS</th>
<th>SVI</th>
<th>labeled stress type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1458212530</td>
<td>345</td>
<td>71.1</td>
<td>3.7</td>
<td>0.72</td>
<td>0.11</td>
<td>15.27</td>
<td>0.08</td>
<td>3.39</td>
<td>0.22</td>
<td>9</td>
<td>1.36</td>
<td>eu</td>
</tr>
<tr>
<td>1458212880</td>
<td>340</td>
<td>109.2</td>
<td>13.4</td>
<td>0.85</td>
<td>0.26</td>
<td>30.58</td>
<td>0.19</td>
<td>20.68</td>
<td>0.47</td>
<td>21</td>
<td>1.38</td>
<td>di</td>
</tr>
</tbody>
</table>
2.2. PRE-STUDY AND FINDINGS

2.2.3 Data analysis with WEKA

The data analysis of the pre-study was performed with the data mining software WEKA\footnote{http://www.cs.waikato.ac.nz/ml/weka/index.html} (Waikato Environment for Knowledge Analysis). WEKA is a unified workbench that allows researchers easy access to state-of-the-art machine learning techniques and enjoys widespread acceptance in both academia and business \cite{6}. Beside a number of other features, the software enables easy access to several preprocessing tools (e.g., data normalization), classification and regression algorithms (e.g., decision tree and SVM classifier) and data visualization with 2D and 3D scatter plots.

2.2.3.1 First analysis

When loading the data into WEKA it provides a first overview of the data set (see figure \ref{fig:overview}). The data set comprises 40 segments, each containing 10 HRV features and a label for the stress type. 23 segments are labeled as eustress and 17 segments are labeled as distress. Furthermore the overview displays for each feature the minimum and maximum value, the average value over all segments and the standard deviation.

![Figure 2.8: Overview of the analyzed data in WEKA.](figure2.8_overview.png)

Table \ref{tab:overview} comprises the average value and the standard deviation of each HRV feature separate for the eustress labeled segments, the distress labeled segments and the rest segments (segments that are marked as not stressed). The rest segments comprise all segments with a stress level of 0, which were filtered out in the segmentation step of the preprocessing (see section 2.2.2.2).

It can be seen that the meanRR and the pRR50 (both are reliable stress measures according to the literature) decrease from rest over eustress to distress, while the meanHR (also a reliable stress measure according to the literature) increases from rest over eustress to distress. As mentioned in section 2.1.2.1 high meanRR and pRR50 values and low meanHR values indicate a lower stress level (respectively rest), while low meanRR and pRR50 values and high meanHR values indicate a rather high stress level. So these findings support the assumption that eustress has a rather moderate stress level, while distress has a higher stress level. But since the RMSSD (considered to be the most relevant and accurate measure of the...
Table 2.1: Comparison of the average values of HRV features during rest, eustress and distress conditions. Furthermore the standard deviation is given in brackets.

<table>
<thead>
<tr>
<th>HRV features</th>
<th>Rest</th>
<th>Eustress</th>
<th>Distress</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. meanHR (bpm)</td>
<td>78.233 ± 11.610</td>
<td>83.094 ± 8.920</td>
<td>90.599 ± 10.863</td>
</tr>
<tr>
<td>avg. SDHR (bpm)</td>
<td>13.050 ± 5.473</td>
<td>12.198 ± 8.444</td>
<td>14.991 ± 6.113</td>
</tr>
<tr>
<td>avg. meanRR (seconds)</td>
<td>0.807 ± 0.110</td>
<td>0.749 ± 0.068</td>
<td>0.695 ± 0.076</td>
</tr>
<tr>
<td>avg. SDRR (seconds)</td>
<td>0.136 ± 0.061</td>
<td>0.105 ± 0.067</td>
<td>0.131 ± 0.058</td>
</tr>
<tr>
<td>avg. RMSSD (seconds)</td>
<td>0.087 ± 0.041</td>
<td>0.075 ± 0.047</td>
<td>0.075 ± 0.047</td>
</tr>
<tr>
<td>avg. pRR50 (%)</td>
<td>30.760 ± 11.862</td>
<td>22.402 ± 11.254</td>
<td>18.345 ± 8.156</td>
</tr>
<tr>
<td>avg. ApEn</td>
<td>0.227 ± 0.099</td>
<td>0.180 ± 0.094</td>
<td>0.283 ± 0.092</td>
</tr>
<tr>
<td>avg. SVI_surrogate</td>
<td>1.707 ± 0.683</td>
<td>1.426 ± 0.265</td>
<td>2.057 ± 0.837</td>
</tr>
</tbody>
</table>

Table 2.2 also comprises the average value and standard deviation of each HRV feature separate for the eustress labeled segments and the distress labeled segments. But now, not with the absolute values as in table 2.1 but with all features normalized to the interval [0,1]. So the various features can be better compared with each other. The normalization was performed with WEKA’s “Normalize” filter, which is an unsupervised attribute filter that scales all numeric values in the data set to lie within the interval [0,1] [35]. The last column of table 2.2 shows for each feature the difference between the average value of the eustress labeled segments and the average value of the distress labeled segments. Positive differences (marked black) mean that the average value of the eustress segments is higher than the average value of the distress segments and negative differences (marked red) mean that the average value of the eustress segments is lower than the average value of the distress segments.

Table 2.2: Comparison of the average values of HRV features (normalized to lie in the interval [0,1]) during eustress and distress conditions. Furthermore the standard deviation is given in brackets.

<table>
<thead>
<tr>
<th>HRV features</th>
<th>Eustress</th>
<th>Distress</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. meanHR</td>
<td>0.316 ± 0.234</td>
<td>0.512 ± 0.285</td>
<td>-0.196</td>
</tr>
<tr>
<td>avg. SDHR</td>
<td>0.325 ± 0.324</td>
<td>0.432 ± 0.234</td>
<td>-0.107</td>
</tr>
<tr>
<td>avg. meanRR</td>
<td>0.644 ± 0.228</td>
<td>0.461 ± 0.257</td>
<td>0.183</td>
</tr>
<tr>
<td>avg. SDRR</td>
<td>0.309 ± 0.294</td>
<td>0.422 ± 0.253</td>
<td>-0.113</td>
</tr>
<tr>
<td>avg. CVRR</td>
<td>0.295 ± 0.287</td>
<td>0.447 ± 0.261</td>
<td>-0.152</td>
</tr>
<tr>
<td>avg. RMSSD</td>
<td>0.301 ± 0.267</td>
<td>0.299 ± 0.268</td>
<td>0.002</td>
</tr>
<tr>
<td>avg. pRR50</td>
<td>0.467 ± 0.277</td>
<td>0.368 ± 0.200</td>
<td>0.099</td>
</tr>
<tr>
<td>avg. ApEn</td>
<td>0.344 ± 0.209</td>
<td>0.572 ± 0.204</td>
<td>-0.228</td>
</tr>
<tr>
<td>avg. VS</td>
<td>0.461 ± 0.210</td>
<td>0.235 ± 0.223</td>
<td>0.226</td>
</tr>
<tr>
<td>avg. SVI_surrogate</td>
<td>0.179 ± 0.093</td>
<td>0.401 ± 0.294</td>
<td>-0.222</td>
</tr>
</tbody>
</table>

The difference is used as a first indicator to assess which HRV features are good suited to distinguish eustress from distress. It is assumed that the higher the absolute value of the difference of a feature is, the better it is separable. Due to this criterion the ApEn, the VS and the SVI_surrogate are the most promising features, with an absolute difference greater
2.2. PRE-STUDY AND FINDINGS

than 0.2. With an absolute difference just slightly lower than 0.2 the meanHR and the meanRR seem also to be good criteria for the separation. Both the SDHR and the SDRR have significantly lower differences than the previously mentioned features and therefore are probably not such good estimators for eustress and distress. However the CVRR, which is a normalization of the SDRR seems to be a better estimator. The RMSSD and the pRR50 also do not separate the stress types very well (both with a difference lower than 0.1). Especially the RMSSD does not show any significant changes between eustress and distress conditions.

Additionally table 2.2 shows that there is no significant difference between the meanHR and the meanRR as well as between the SDHR and the SDRR, with respect to the separability of eustress and distress. For this reason and also because the heart rate and the RR-intervals are reciprocals of each other, the meanHR and the SDHR are not of further interest in the following investigations. So the number of considered features is reduced to eight (meanRR, SDRR, CVRR, RMSSD, pRR50, ApEn, VS and SVIsurrogate).

2.2.3.2 Visual analysis

In order to get more insight into the data set WEKA’s visualization capabilities are used to further analyze the data. Therefore data set is visualized in 2D scatter plots, where every data point represents a labeled segment (eustress or distress). For the visualization the normalized data is used to better compare the various features with each other. First the three most promising features of the previous analysis (ApEn, VS and SVIsurrogate) and their correlations are observed with respect to the separation of eustress and distress segments.

Figure 2.9: Scatter plot visualization of the relationship between SVIsurrogate and VS.

Figure 2.9 shows the correlation between the SVIsurrogate and the VS for eustress and distress segments. It is apparent that the two features are negatively correlated (the higher the SVIsurrogate the lower is the VS and the higher the VS, the lower is the SVIsurrogate). For lower SVIsurrogate values and a higher VS (top left on the scatter plot) the eustress segments dominate whereas for higher SVIsurrogate values and a lower VS (bottom right on the scatter plot) the distress segments are in the majority. So the SVIsurrogate and the VS separate the eustress and distress segments quite well (see the black line in figure 2.9).
Next the interaction of the ApEn and the VS is considered (figure 2.10). The scatter plot shows that except for three outliers these two features are also negatively correlated (the higher the ApEn, the lower is the VS and the higher the VS, the lower is the ApEn). For lower ApEn values and a higher VS (top left on the scatter plot) the eustress segments dominate whereas for higher ApEn values and a lower VS (bottom right on the scatter plot) the distress segments are in the majority. It can be seen that the ApEn and the VS also separate the eustress and distress segments quite well along the black line in figure 2.10.

![Figure 2.10: Scatter plot visualization of the relationship between ApEn and VS.](image1)

At least the correlation of the ApEn and the SVIsurrogate is illustrated in figure 2.11. In contrast to the two previous feature combinations the correlation between the ApEn and the SVIsurrogate is not so obvious. But it can be seen, that with a higher ApEn the number of data points with a higher SVIsurrogate value increases (positive correlation). For lower SVIsurrogate values and a lower ApEn (bottom left on the scatter plot) the eustress segments...
2.2. PRE-STUDY AND FINDINGS

dominate whereas for higher $\text{SVI}_{\text{surrogate}}$ values and a higher ApEn (top right on the scatter plot) the distress segments are in the majority. The black line in figure 2.11 illustrates the ability of the ApEn and the $\text{SVI}_{\text{surrogate}}$ to separate the eustress and distress segments quite well.

The findings of this visual analysis, which considered the correlation between the ApEn, the VS and the $\text{SVI}_{\text{surrogate}}$, suggest that the combination of these three features can separate eustress from distress quite well. Furthermore, these findings support the assertion that the ApEn, the VS and the $\text{SVI}_{\text{surrogate}}$ are good indicators for the distinction of eustress and distress.

The next best feature according to the analysis of table 2.2 was the meanRR. Figure 2.12 shows the relationship between the meanRR and the VS with regard to the separation of eustress and distress segments. As also seen in figure 2.9 and 2.10, the VS separates the segments so that higher values represent eustress and lower values represent distress. Additionally, here it can be seen that the meanRR also separates the segments in a way that for higher values predominantly eustress segments exist and for lower values exist predominantly distress segments. Taken together, this means that for lower meanRR values and a lower VS (bottom left on the scatter plot) the distress segments dominate, whereas for higher meanRR values and a higher VS (top right on the scatter plot) the eustress segments are in the majority. As a result, the eustress and distress segments can be separated quite well by the meanRR and the VS along the black line in figure 2.12. These findings confirm the assumption that the meanRR is a good criterion for eustress and distress distinction.

Figure 2.12: Scatter plot visualization of the relationship between meanRR and VS.

Another interesting fact can be seen when regarding the correlation between the SDRR or CVRR and the pRR50, with respect to the separation of eustress and distress segments (see figure 2.13 respectively 2.14). Although these features did not perform very well in the analysis of table 2.2, the visual analysis reveals a good separability of eustress and distress segments when using SDRR and pRR50 or CVRR and pRR50 as feature combinations. The first noticeable thing is that the scatter plots show for both of the two feature combinations that the considered features are positively correlated (the higher the SDRR or CVRR, the higher is the pRR50). And although every feature considered on its own does not separate the segments very well, the combination of the features does. It can be seen that the higher
the SDRR or CVRR, the higher is the pRR50 boundary value that separates the eustress and distress segments. Whereby the eustress segments are located above the boundary and the distress segments are located below the boundary. The black line in figure 2.13 and 2.14 represents this boundary. So the feature combinations composed of the SDRR or CVRR and the pRR50 separate the eustress and distress segments quite well along the boundary line. The only difference between the two feature combinations is that with the CVRR instead of the SDRR the data points are more spread out, what might be useful when trying to separate the segments with machine learning algorithms. This fact also explains why in table 2.2 the difference of the eustress and distress segments is higher for the CVRR than for the SDRR.

Figure 2.13: Scatter plot visualization of the relationship between SDRR and pRR50.

Figure 2.14: Scatter plot visualization of the relationship between CVRR and pRR50.
2.2. PRE-STUDY AND FINDINGS

2.2.3.3 Classification performance

Based on the results of the previous analysis, now several feature combinations are used to classify the data set. As classification algorithms a decision tree classifier (WEKA's implementation of the C4.5 algorithm [35]) and a support vector machine (SVM) with a linear kernel (WEKA's LibSVM implementation) are used. The objective is to verify, if it is possible to classify eustress and distress segments with an appropriate accuracy and which feature combinations achieve the best results.

In order to evaluate the classification a 10-fold cross-validation is performed. This means that the data set is divided up into 10 parts with approximately equal proportions of the classes (eustress and distress). Then 9 parts are used to train a classification model and one part is used as test set to evaluate this model. This is done 10 times, so that every part was once used as test set. The results of each of the 10 classifications are then aggregated, so that the classification performance can be evaluated for the whole data set. [7] This method is preferred in contrast to a simple percentage split (where just about 30% of the data set is tested), because the here used data set is very small and thus a cross-validation is more expressive.

In the following the classification performance of both decision tree and SVM is evaluated for six different feature combinations and checked against each other. As performance measures the accuracy, the $F_1$ score (also known as F-score or F-measure) and a confusion matrix are used. The Accuracy is the number of correctly classified segments relative to the total number of segments. The $F_1$ score is the harmonic mean of precision and recall (see equation 2.16) and is here calculated as the weighted average of the $F_1$ score for eustress and the $F_1$ score for distress. The confusion matrix presents the distribution of the classified data. Top left is the number of correctly classified eustress segments (true positive), top right is the number of incorrectly classified eustress segments (false negative), bottom right is the number of correctly classified distress segments (true negative) and bottom left is the number of incorrectly classified distress segments (false positive).

$$ F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} $$

(2.16)

The following listing shows the six most promising feature combinations, used for the classification of the data set:

- Feature combination 1: meanRR, SDRR, CVRR, RMSSD, pRR50, ApEn, VS, SVIsurrogate
- Feature combination 2: ApEn, VS, SVIsurrogate
- Feature combination 3: meanRR, SDRR, CVRR, RMSSD, pRR50
- Feature combination 4: SDRR, pRR50
- Feature combination 5: meanRR, SDRR, pRR50, ApEn, VS, SVIsurrogate
- Feature combination 6: meanRR, SDRR, RMSSD, ApEn, VS, SVIsurrogate

Table 2.3: Comparison of the classification performance of the decision tree (DT) classifier and the SVM for the six different feature combinations.

<table>
<thead>
<tr>
<th>Feature combination</th>
<th>Accuracy (DT)</th>
<th>$F_1$ score (DT)</th>
<th>Accuracy (SVM)</th>
<th>$F_1$ score (SVM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70.0 %</td>
<td>0.697</td>
<td>77.5 %</td>
<td>0.767</td>
</tr>
<tr>
<td>2</td>
<td>70.0 %</td>
<td>0.700</td>
<td>75.0 %</td>
<td>0.747</td>
</tr>
<tr>
<td>3</td>
<td>82.5 %</td>
<td>0.822</td>
<td>67.5 %</td>
<td>0.669</td>
</tr>
<tr>
<td>4</td>
<td>85.0 %</td>
<td>0.848</td>
<td>50.0 %</td>
<td>0.440</td>
</tr>
<tr>
<td>5</td>
<td>67.5 %</td>
<td>0.663</td>
<td>80.0 %</td>
<td>0.800</td>
</tr>
<tr>
<td>6</td>
<td>75.0 %</td>
<td>0.747</td>
<td>75.0 %</td>
<td>0.747</td>
</tr>
</tbody>
</table>

Table 2.4: Confusion matrix for the two classification schemes (decision tree and SVM) for the six different feature combinations. The rows represent the labeled segments and the columns represent the predicted segments.

<table>
<thead>
<tr>
<th>Classification schemes</th>
<th>Decision tree</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eustress</td>
<td>Distress</td>
</tr>
<tr>
<td>Feature combination 1</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Feature combination 2</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Feature combination 3</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Feature combination 4</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Feature combination 5</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Feature combination 6</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

First all eight features (feature combination 1) were used for the classification. With an accuracy of 77.5% the SVM performed quite well, whereas the decision tree classifier performed little worse (but still 70% accuracy). But the confusion matrices for feature combination 1 show that the good accuracy results mainly from the correct classification of eustress segments.

Then just the three most promising features according to the analysis results of table 2.2 were used for the classification (feature combination 2). Although the accuracy of both classification schemes did not change significantly, it should be noticed that the correct classification of the eustress and distress segments becomes more balanced.

When applying the remaining five features (feature combination 3) to the classification algorithms, it can be seen that whereas the accuracy of the decision tree classifier increases up to 82.5%, the accuracy of the SVM decreases significantly (down to 67.5%). Most of all, the classification of the distress segments gets worse for the SVM (almost half of the distress segments are misclassified).

Next, only the SDRR and pRR50 feature (feature combination 4) were used for the classifica-
tion, since they showed a good separability in figure 2.13 of the visual analysis. Although in the visual analysis it was mentioned that the conjunction of CVRR and pRR50 may be better suited for the classification, this assumption could not be approved. This setting further reinforced the tendency of feature combination 3, by increasing the accuracy of the decision tree classifier up to a maximum of 85% and decreasing the accuracy of the SVM to a minimum of only 50% and a even worse F1 score of 0.44. Regarding the SVM compared to the previous feature combination just the classification of the distress segments worsened, but at a rate so that nearly all distress segments were classified wrong.

The evaluation results of feature combination 3 and 4 suggest that especially the ApEn, the VS and the SVIsurrogate are important features for the SVM classification, whereas they are rather obstructive for the decision tree classifier.

For the next classification all features were used, except for the CVRR and the RMSSD (feature combination 5). The CVRR was excluded because it had no benefits for the classification compared to the SDRR and the RMSSD was excluded because it did not reveal any differences between eustress and distress, neither in the analysis of table 2.2 nor in the visual analysis. This time the accuracy of the decision tree classifier decreased to a minimum of 67.5%. Although, compared to feature combination 1 and 2 the accuracy did not change significantly, the classification gets more imbalanced (more than half of the distress segments were classified wrong). In contrast, the SVM achieved its maximum accuracy with 80% and has a quite good balanced classification result.

Finally the previous feature combination 5 was modified by replacing the pRR50 with the RMSSD, in order to examine the impact of the RMSSD feature (feature combination 6). Whereas the accuracy of the decision tree classifier increased (resulting from a better classification of the distress segments), the accuracy of the SVM decreased, due to a worse classification of the distress segments. Both classification schemes achieved an accuracy of 75% and have the same confusion matrix.

So the RMSSD seems to have a rather positive impact on the performance of the decision tree classifier and a rather negative impact on the performance of the SVM.

2.2.4 Findings of the pre-study

The results of the pre-study suggest that it is possible to separate eustress from distress by means of HRV features with an appropriate accuracy. But the pre-study has also shown that the choice of a proper feature combination is a big issue for the accuracy of a classifier. Whereas the feature combinations without ApEn, VS and SVIsurrogate yielded to the best performances of the decision tree classifier, they were the worst for the SVM. When taking the ApEn, VS and SVIsurrogate into account the SVM achieved significant better accuracies, whereas the decision tree classifier lost in accuracy. Since no dominant feature combination could be found in the pre-study, their suitability has to be further investigated.

Furthermore the results of the pre-study (especially regarding table 2.1 and table 2.2) allow the following assumptions:

- The meanRR decreased linearly from rest over eustress to distress. Thus it is assumed that rest has a higher meanRR, eustress has a moderate meanRR and distress has a lower meanRR.
CHAPTER 2. CONCEPTION

- The SDNN showed no significant difference between rest and distress, but was lower for eustress. Thus it is assumed that rest and distress have a higher standard deviation (greater distance to the mean value), whereas eustress has a lower standard deviation (smaller distance to the mean value).

- The RMSSD showed no difference between eustress and distress, but was higher for rest. Thus it is assumed that rest has a higher RMSSD (greater distance between consecutive RR-intervals), whereas eustress and distress have a lower RMSSD (smaller distance between consecutive RR-intervals).

- The pRR50 decreased from rest over eustress to distress. However the difference between rest and eustress was significant higher than between eustress and distress. Thus it is assumed that rest has a higher pRR50 (more consecutive RR-intervals with a great difference), eustress has a rather moderate pRR50 and distress has a lower pRR50 (fewer consecutive RR-intervals with a great difference).

- The ApEn increased linearly from eustress over rest to distress. Thus it is assumed that rest has a moderate ApEn, whereas eustress has a lower ApEn (more repetitive patterns) and distress has a higher ApEn (fewer repetitive patterns).

- The VS was significantly higher for eustress than for rest and distress, where distress had the lowest value. Thus it is assumed that eustress has a higher VS (more changes around the mean value), whereas rest and distress have a lower VS (fewer changes around the mean value).

- The SVIsurrogate increased linearly from eustress over rest to distress. Thus it is assumed that eustress has a lower SVIsurrogate (much high-frequency influences), rest has a moderate SVIsurrogate (still more high-frequency influences) and distress has a higher SVIsurrogate (more low-frequency influences).

Based on these assumptions and so resulting from the findings of the pre-study, now three distinct heart rate patterns for rest, eustress and distress are proposed.

Rest:
Irregular heart rate with irregular changes among the consecutive RR-intervals (figure 2.15).
In other words, there are frequent changes of the heart rate respectively of the RR-intervals (irregular heart rate) while there is an irregularity in the size of the changes (irregular changes among the consecutive RR-intervals). This means that the heart rate time series for rest contain few repetitive patterns.

Figure 2.15: Heart rate pattern for rest.
2.2. PRE-STUDY AND FINDINGS

Eustress:
Irregular heart rate with regular changes among the consecutive RR-intervals (figure 2.16). In other words, there are frequent changes of the heart rate respectively of the RR-intervals (irregular heart rate) while there is a regularity in the size of the changes (regular changes among the consecutive RR-intervals). This means that the heart rate time series for eustress contain much repetitive patterns.

Figure 2.16: Heart rate pattern for eustress.

Distress:
Regular heart rate with regular changes among the consecutive RR-intervals (figure 2.17). In other words, there are just few changes of the heart rate respectively of the RR-intervals (regular heart rate) and there is a regularity in the size of the changes (regular changes among the consecutive RR-intervals). This means that the heart rate time series for distress contain much repetitive patterns.

Figure 2.17: Heart rate pattern for distress.

The following listing summarizes the assumed feature characteristics for eustress and distress:

- The meanRR is higher for eustress than for distress (eustress > distress)
- The SDRR and CVRR are lower for eustress than for distress (eustress < distress)
- The RMSSD does not differ significantly between eustress and distress, but with a tendency of higher values for eustress than for distress (eustress ≥ distress)
CHAPTER 2. CONCEPTION

- The pRR50 is higher for eustress than for distress (eustress > distress)
- The ApEn is lower for eustress than for distress (eustress < distress)
- The VS is higher for eustress than for distress (eustress > distress)
- The SVis surrogate is lower for eustress than for distress (eustress < distress)

2.2.5 Opportunities and limitations of the pre-study

Although the data for the pre-study were collected over a period of seven days, with just 40 labeled segments the data set was very small. This was the case because for one thing, only one study participant recorded eustress and distress experiences and for another thing, just a few labeled segments had an acceptable sample rate so that they could be used for the data analysis. The mainly bad sampling rate results from the usage of Xiaomi’s Mi Band 1s[^1] that does not provide continuous heart rate monitoring itself, but needs a third-party application that runs on a nearby connected smartphone[^9]. Problems with the connection can then lead to inappropriate sampling rates.

The fact that the data set contains only few measurements from only one participant means that the pre-study results are not universally valid, but rather that they only allow first assumptions. In order to verify the pre-study assumptions a larger field study with multiple participants has to be conducted.

Another critical issue of the pre-study is the accuracy of the data labels (eustress and distress estimation). Different from a laboratory study where the participants can be confronted with designated situations to elicit e.g. eustress and distress, the data labels in the conducted pre-study are based on the self-assessment of the participants. So the accuracy of the data labels is highly dependent on a conscientious estimation of the participants.

However, there are not only limitations that come along with the pre-study, but also some benefits. Since the analyzed data is derived from just one participant the data is not biased by physical differences between various participants (e.g. physical fitness, age, sex, etc.). Furthermore the data is also barely biased by physical activities, since all the data were collected while periods without physical effort.

Despite the mentioned limitations the pre-study also showed the opportunity to detect eustress and distress by means of various HRV features, even with simple technologies suitable for everyday use.

2.3 A model for the classification of eustress and distress

In order to separate eustress from distress in a data set of labeled heart rate patterns that are described through a set of HRV features, two classification algorithms are used, namely a decision tree classifier and a support vector machine (SVM).

2.3.1 Supervised learning

Classification is basically a synonym for supervised learning and is one of the classic problems in machine learning. Inter alia, machine learning is concerned with computer programs that can automatically learn to recognize complex patterns and make intelligent

2.3. A MODEL FOR THE CLASSIFICATION OF EUSTRESS AND DISTRESS

decisions based on data. The classification of eustress and distress in this work is a typical machine learning problem, since it is intended to automatically recognize heart rate patterns after learning from a set of examples. [7]

Both machine learning algorithms that are utilized in this work (decision tree and SVM) are referred to the field of supervised learning. Supervised learning means that the examples in the training data set (which is used to train the classification model) are labeled with a class. So the classification is a process of finding a model that describes and distinguishes data classes, where the model is created based on the analysis of a set of training data (data objects for which the class labels are known). The model is then used to predict the class label of objects for which the class label is unknown. For the eustress and distress classification in this work, this means that a set of heart rate patterns and their corresponding class label (eustress or distress) are used as the training examples, which supervise the learning of the classification model. With the resulting model, the class label of all heart rate patterns without class label can be predicted. [7]

In contrast to the supervised learning, which is used in this work, unsupervised learning would mean that the input examples are not class labeled. Typically unsupervised learning, which is essentially a synonym for clustering, is used to discover classes within the data. Using the example of heart rate patterns, this would mean that if a set of heart rate patterns without eustress or distress labels is taken as input, then a clustering algorithm may find two or more clusters of data. However, since the training data is not labeled, the learned model could not tell the semantic meaning of the clusters found. So there would be no prediction possible, which heart rate patterns are eustress and which are distress. But since exactly this is the objective, in this work supervised learning methods are used. [7]

The decision to choose decision tree and SVM as classification methods in this work was made, because both of them are commonly used machine learning techniques in the field of stress detection and achieved good performances in the literature. In [28] SVM was ranked as the top machine learning technique for modeling stress, with an reported accuracy of 90.1%. With an reported accuracy of 88.02%, the decision tree classifier performed just slightly worse. In the study of Sun et al. [31] the decision tree classifier (accuracy of 92.4%) even outperformed the SVM (accuracy of 85%).

2.3.1.1 Decision tree

The decision tree classifier is based on a divide-and-conquer approach, where the basic idea is to break up a complex decision into several simpler decisions, with the aim that the final solution obtained this way would resemble the intended desired solution. [35]

In detail, a decision tree has a flowchart-like tree structure, consisting of:

- internal nodes (all non-leaf nodes), which represent test conditions for attributes in order to divide the input space
- branches, which represent the outcome of a test
- leaf nodes (also called terminal nodes), which hold class labels representing a target class.

The classification in a decision tree works then as follows. Given a data object (consisting of several attributes) with an unknown class label, the attribute values of the data object are tested against the test conditions of the internal nodes beginning at the topmost node in the
CHAPTER 2. CONCEPTION

tree (the root) and moving along the branches to a leaf node. The reached leaf node then holds the class prediction for the input data object. For more details about the construction of decision trees refer to [7].

Figure 2.18 shows a suppositious decision tree for the classification of eustress and distress. Assuming a heart rate pattern that is described through two attributes ($meanRR = 0.3$ and $SDRR = 0.07$) and for which the class label is unknown. Then, by starting at the root node, first the meanRR attribute is checked against the condition of the root node ($meanRR > 0.6$). Since the check is negative ($0.3 < 0.6$), the right branch (“no”-branch) is taken to the next internal node. Now, the SDRR attribute is checked against the condition in this internal node ($SDRR > 0.02$). Since now the check is positive ($0.07 > 0.02$), the left branch (“yes”-branch) is taken and a leaf node is reached. Since the reached leaf node represents the target class “eustress”, the class label of the input heart rate pattern is predicted as eustress.

Figure 2.18: A suppositious decision tree for the classification of eustress and distress.

The popularity of decision trees results from various qualities [7]:

- Neither domain knowledge nor parameter settings are necessary for the construction of decision tree classifiers.
- Decision trees can handle multidimensional data.
- Decision trees have an intuitive and readily understandable knowledge representation.
- The learning and classification steps are simple and fast.
- In general, the accuracy of decision trees is quite good.

But still there are also some problems with decision trees. Although decision tree classifiers perform well on the training data set, they are usually overfitted to the training data and do not generalize well to independent test sets [35]. Another potential problem using decision trees to model stress are the crisp splits for prediction [28].

2.3.1.2 Support vector machine

The SVM is a method for the classification of both linear and nonlinear data and works as follows. When classifying a two-class problem where the classes are linearly separable, a linear optimal separating hyperplane is determined within the search space of labeled
2.3. A MODEL FOR THE CLASSIFICATION OF EUSTRESS AND DISTRESS

training tuples. In other words, a decision boundary is searched that separates the data objects of one class from another. The SVM uses so called support vectors (“essential” training tuples) and margins (defined by the support vectors), in order to find the maximum marginal hyperplane (the hyperplane with the largest margin to the closest training tuple of either class) that separates the two classes best. The support vectors are the training tuples of either class that have the shortest distance to the maximum marginal hyperplane. The margin is then just this shortest distance, where the shortest distance from the hyperplane to one side of its margin is equal to the shortest distance from the hyperplane to the other side of its margin. For a more detailed description of SVMs refer to [7].

Figure 2.19 shows a suppositious separation of labeled training data (classes eustress and distress) by means of a maximum marginal hyperplane (MMH). Again, given a heart rate pattern (described through the two attributes \( \text{meanRR} = 0.3 \) and \( \text{SDRR} = 0.07 \)) for that the class label is unknown, the SVM checks on which side of the hyperplane the input heart rate pattern falls. In this case the heart rate pattern falls above the maximum marginal hyperplane (representing the target class “eustress”) and thus the class label of the input heart rate pattern is predicted as eustress.

![Figure 2.19: A suppositious separation of eustress and distress training data by means of a maximum marginal hyperplane (MMH).](image)

In the case that the classes are not linear separable (no straight line can be found that would separate the classes), a linear SVM would not be able to find a feasible solution. In order to find nonlinear decision boundaries the linear SVM has to be extended to create a nonlinear SVM. This is done by first transforming the original training data into a higher dimensional space by applying a nonlinear mapping. Then a linear optimal separating hyperplane is searched within this new dimension. Solving this optimization problem can be done in the same manner as for the linear SVM. The maximal marginal hyperplane found in the new space corresponds to a nonlinear separating hypersurface in the original space (see figure 2.20). For the transformation of the training data into a higher dimension different types of kernel functions can be applied, so that data from two classes can always be separated by a
hyperplane, when using an appropriate nonlinear mapping to a sufficiently high dimension. [7]

![Figure 2.20: A suppositious separation of eustress and distress training data by means of a nonlinear separating hypersurface.](image)

Although a shortcoming of SVMs is that the training time can be extremely slow, this classification method has also some crucial benefits [7]:

- Due to their ability to model complex nonlinear decision boundaries, SVMs are highly accurate.
- Compared to other classification methods (for example decision tree), SVMs are much less prone to overfitting.
- The support vectors found provide a compact description of the learned model.
- SVMs can be used for numeric prediction as well as for classification.

### 2.3.2 Classification process

For the classification with both decision tree and SVM labeled vectors are required as training data in order to create the classification model. Therefore several preprocessing steps are necessary to get from the raw heart rate data to labeled vectors.

The starting point for the classification process are heart rate measures and activities, collected with the CStress app developed by Marcel Heil [9]. The heart rate measures are provided in a list, where they are ordered ascending by their corresponding timestamp (see figure 2.21). The activities are reported from the users of the CStress app and comprise the start timestamp and end timestamp of the activity, the name of the activity and the feeling during the activity (see figure 2.22). The reported feeling can be either neutral, positive or negative and is determined by self-assessment.
2.3. A MODEL FOR THE CLASSIFICATION OF EUSTRESS AND DISTRESS

Figure 2.21: A snippet of the heart rate list. The first column contains the timestamp of the record in seconds, the second column contains the heart rate value in beats per minute.

<table>
<thead>
<tr>
<th>timestamp (seconds)</th>
<th>heart rate (bpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1457996400</td>
<td>77</td>
</tr>
<tr>
<td>1457996415</td>
<td>77</td>
</tr>
<tr>
<td>1457996420</td>
<td>79</td>
</tr>
<tr>
<td>1457996425</td>
<td>79</td>
</tr>
<tr>
<td>1457996430</td>
<td>80</td>
</tr>
<tr>
<td>1457996435</td>
<td>82</td>
</tr>
<tr>
<td>1457996440</td>
<td>83</td>
</tr>
<tr>
<td>1457996445</td>
<td>85</td>
</tr>
</tbody>
</table>

Figure 2.22: Structure of the reported activities. The first column contains the start timestamp of the activity, the second column contains the end timestamp of the activity, the third column contains the name of the activity and the fourth column contains the feeling during the activity.

<table>
<thead>
<tr>
<th>start timestamp (seconds)</th>
<th>end timestamp (seconds)</th>
<th>name</th>
<th>feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>1467700200</td>
<td>1467702000</td>
<td>studying</td>
<td>negative</td>
</tr>
<tr>
<td>1467747000</td>
<td>1467754200</td>
<td>watching tv</td>
<td>positive</td>
</tr>
<tr>
<td>1467784800</td>
<td>1467788400</td>
<td>eating</td>
<td>neutral</td>
</tr>
</tbody>
</table>

Figure 2.23 shows the layout of the classification process, which is described in the following.

![Classification process diagram]

Figure 2.23: Layout of the classification process, consisting of six essential steps. The segmentation of the raw heart rate data, the labeling of the resulting segments, the extraction of the features, the generation of the classification model, the prediction of class labels for unlabeled segments and the confirmation or refusal of the predicted class labels.
2.3.2.1 Segmentation

In a first step the heart rate measures are divided up into disjoint segments with a length of 10 minutes. In contrast to the pre-study, every segment has the same length and the number of heart rate measures per segment may vary. For the extracted segments apply the following restrictions:

1. The segments always begin in steps of 10 minutes. This means that for example the first segment begins at 08:00 a.m., the next at 08:10 a.m., the next at 08:20 a.m., and so on. So it is ensured that all segments are disjunct.

2. Each segment has to contain a minimum number of $n_{min}$ heart rate values, otherwise the segment is discarded and the heart rate values in this time interval are no longer considered. In this work $n_{min}$ is set to 60, so that there is a sufficient amount of heart rate measures per segment.

Since time series of RR-intervals are required to calculate HRV features, for each segment the heart rate values are converted into RR-intervals with units of seconds, by taking the inverse of the corresponding heart rate value multiplied with 60 ($RR_i = \frac{60}{HR_i}$). So now each segment contains a time series of $n$ consecutive RR-intervals.

2.3.2.2 Determining the stress level

In order to ascertain during which of the just created segments a user was stressed, the stress level of each segment has to be determined. For this purpose, the computed stress model defined by Marcel Heil [9] is used. The computed stress is calculated as:

$$\text{Stress}_{\text{computed}} = 110 - HRV_{\text{score}}$$

(2.17)

where $HRV_{\text{score}} = \ln(RMSSD) \times 20$ (for more details about the calculation refer to [9]). So a lower computed stress value indicates a lower stress level and a higher computed stress value indicates a higher stress level.

The computed stress is also provided by the CStress app and since it is calculated for every ten minutes the computed stress values can be easily assigned to the segments. In order to use the computed stress value to determine whether a segment is stressed or not, a stress threshold has to be defined. According to [9] a good threshold to determine whether a user is stressed or not is a computed stress value of 50. So if the computed stress value of a segment is lower than 50, the segment is marked as not stressed (respectively calm) and if the computed stress value is greater or equal to 50, the segment is marked as stressed.

2.3.2.3 Determining the stress type

In order to determine the stress type (eustress or distress) of the segments now the aforementioned activities reported by the users are considered. For each stressed segment (not stressed segments are not considered because they can not be eustress or distress) the corresponding activity is identified. The assessed feeling of this activity is then used to determine the class label. If there is no activity reported for the time range of the segment or the feeling of the activity is neutral, the segment remains unlabeled (the class label is unknown). If the feeling of the activity is positive the segment is labeled as eustress and if the feeling of the activity is negative the segment is labeled as distress.
So after these preprocessing steps there is a set of 10-minute segments. Each contains a time series of \( n \geq n_{\text{min}} \) consecutive RR-intervals, a binary stress level (stressed or not stressed) and if stressed a binary stress type, which is either eustress or distress (class label is known) or the stress type is undefined (class label is unknown).

### 2.3.2.4 Train the classification model

In order to train the decision tree classifier or the SVM, a set of training data is required. Since the training data has to be labeled with a class, just segments that are labeled with eustress or distress (segments that are marked as stressed and for which the class label is known) can be used as training data. For these segments the HRV features are calculated as described in section 2.1.2. This is done in the feature extraction step of figure 2.23. The used features are the meanRR, the SDRR, the CVRR, the RMSSD, the pRR50, the ApEn, the VS and the SVIsurrogate. For the SVM the extracted features are normalized to lie in the interval \([0,1]\) and standardized to have unit variance and zero mean. Then for each segment a vector is created with a n-tuple comprising the extracted features and the vector is labeled with the stress type of the corresponding segment as class label.

The just created labeled vectors are then used as training data set to train the classification model (decision tree classifier or SVM) as described in section 2.3.1.

### 2.3.2.5 Classify the unlabeled segments

With the trained classification model, the remaining unlabeled segments (segments that are marked as stressed and for which the class label is unknown) can be classified. For these segments, again the features are extracted in the same way as described above and composed as n-tuple in a vector. But this time the created vectors are not labeled, since the class label is unknown and has to be predicted now. So the just created unlabeled vectors are used as input for the classification model (decision tree classifier or SVM), which predicts the class label (eustress or distress) for each input vector and thus for each unlabeled segment.

### 2.3.2.6 Enhance the training data

The last step of the classification process is the confirmation or refusal of the predicted class labels. This means that the predicted class labels of the previously classified segments can be either confirmed or refused, so that these segments become labeled permanently. This is done in order to further enhance the training data. Figure 2.24 shows in which way the confirmed or refused segments become labeled.

<table>
<thead>
<tr>
<th>predicted as</th>
<th>confirm</th>
<th>refuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>eustress</td>
<td>eustress</td>
<td>distress</td>
</tr>
<tr>
<td>distress</td>
<td>distress</td>
<td>eustress</td>
</tr>
</tbody>
</table>

Figure 2.24: Labeling of confirmed and refused segments. If confirmed, as eustress predicted segments become labeled as eustress and as distress predicted segments become labeled as distress. If refused, as eustress predicted segments become labeled as distress and as distress predicted segments become labeled as eustress.
In this way, newly emerged labeled segments are then added to the set of labeled segments that are used as training data for the generation of the classification model. So when the classifier is trained again, a larger set of training data is available and the classification model becomes more precise, due to the property of the high variance machine learning algorithms.

2.4 System design

In order to implement the classification process of section 2.3.2, a system is designed, which covers all the tasks from the input of the raw data through to the visualization of the classification results. The system consists of five main components:

- **The data input component**: Responsible for the data synchronization with the CStress app and the input of data via a web portal.

- **The database**: Stores all the data of the application at a central storage location. Section 2.5 provides a detailed description of the database concept.

- **The visualization Component**: Displays the data from the database (inter alia heart rate measures, activities and segments) in a web portal. Section 2.6 describes the visualization concept in more detail.

- **The preprocessor**: Covers the first two steps of the classification process (segmentation and labeling).

- **The classification component**: Covers the next three steps of the classification process (feature extraction, the training of the classifier and the prediction of class labels).

Figure 2.25 shows the components of the developed system and the data flow in this system.

The data flow of the system looks as follows:

1. The data input component receives the collected data from the CStress app (raw heart rate data, computed stress and activities) and saves them in the database.
2.5. DATABASE CONCEPT

2. When the preprocessor is started, it gets the raw heart rate data, the computed stress values and the activities from the database and creates segments, as described in section 2.3.2.

3. The resulting segments are then saved in the database. The set of segments comprises eustress and distress labeled segments as well as segments, which are not labeled or not stressed.

4. The classification component gets all eustress and distress labeled segments from the database to train the classification model and all unlabeled segments to predict their class label.

5. The classified segments are then updated in the database with their predicted class label.

6. The visualization component gets all its required data from the database, performs in some circumstances some data aggregation and displays the data. The displayed data may be activities, the raw heart rate measures, the computed stress, the created segments or thereout aggregated data.

Chapter 3 covers the single components of the system in more detail.

2.5 Database concept

The database holds all the data that have to be stored permanently in the application. Figure 2.26 outlines the database structure.

2.5.1 The users collection

The users collection has one entry for every user who is registered in the system. Each user has a unique id that serves as primary key for this collection. It is used as identifier for the whole application as well as for the data synchronization with the CStress app. The name attribute can be freely chosen by the user, since it has not to be unique. Furthermore a unique consecutive number is assigned to each user in the database (the number attribute). This number is used to create the alias for the users by concatenating the term ”User” with the user’s number. So the alias is also unique and is used to make the visualization anonymous. Lastly, each user gets a role assigned that determines what the user is allowed to see and to do in the application. Especially, the role regulates the access to certain parts of the web portal. The default value of this attribute is the role with the lowest access rights.

The authorization concept

The authorization concept is implemented by means of roles that can be assigned to users. The user roles are organized in a hierarchical structure, where each superior role has all access rights of the subordinated roles. In the following the roles are listed in ascending order:

1. Student (1): The role with the lowest access rights and the default value if nothing else is specified. Allows just read access to the own data.

2. Student Plus (2): This role allows additionally write access to the own data (adding new activities via the web portal).
3. **Lecturer (3):** This role allows additionally read access to aggregated data of all users.

4. **Admin (4):** The role with the highest access rights, allows additionally read access to detailed data of all users.

![Database Structure Diagram](image)

**Figure 2.26:** Outline of the database structure. The first column contains the name of the attributes and the second column contains the data type. Primary keys are bold and underlined.

### 2.5.2 User specific collections

Each user then has separate collections to store the raw heart rate data, the computed stress, the reported activities and the calculated segments. In order to identify the collections of each user they are named with a prefix before the actual collection name, where the prefix is the id of the respective user.

#### 2.5.2.1 The heartrate collection

The *heartrate* collection comprises all the collected heart rate measures that were synchronized with the CStress app. Each entry in this collection has a unique *timestamp* as primary key, which is the moment of measurement and is stored as UTC timestamp in seconds. The *heartrate* attribute contains the measured heart rate value in beats per minute. Furthermore, each entry has a tag called *processed* that indicates if the respective heart rate value has been already processed (*processed = 1*) or not (*processed = 0*). A heart rate value is tagged as processed (*processed = 1*) if it was utilized in the preprocessor, whether or not it is finally part of a segment. The default value of the tag is unprocessed (*processed = 0*).
2.5. DATABASE CONCEPT

2.5.2.2 The hrvdata collection
The hrvdata collection contains the computed stress values calculated by the CStress app, which are used to decide if someone is stressed or not. Each entry in this collection has a unique timestamp as primary key (stored as UTC timestamp in seconds), which is the start time of the ten minute long time interval for that the computed stress value was calculated. The HRV score (hrvScore attribute) is the initial calculated value that is used to derive the computed stress value (computedStress attribute).

2.5.2.3 The activities collection
The activities collection comprises all the reported activities of the user (either reported via the CStress app or via the web portal). Each entry in the activities collection contains a start time (startTimestamp attribute) and an end time (endTimestamp attribute). Both timestamp attributes are stored as UTC timestamps in seconds and form together the primary key. So just the time interval consisting of the startTimestamp and the endTimestamp has to be unique, but not the single timestamps. This implies that there can not be two activities at the exact same time, but the activities can overlap. Furthermore, each activity has a name (activityName attribute) that can be freely chosen by the user and a self-assessed feeling. The self-assessed feeling can be either "neutral", "positive" or "negative". Like the entries in the heartrate collection, each activity entry has a processed tag that indicates if the respective activity has been already processed (processed = 1) or not (processed = 0). For the tagging of activities apply the same rules as for heart rate entries (the default value is unprocessed and tagging as processed if the activity was utilized in the preprocessor).

2.5.2.4 The segments collection
The segments collection contains all the segments that were created in the segmentation step of the classification process (rest segments, eustress/distress labeled segments and unlabeled segments). Each entry in this collection contains a unique timestamp as primary key (stored as UTC timestamp in seconds), which is the start time of the segment and the length of the segment in seconds (the same for each segment). The number of heart rate measures respectively the number of RR-intervals that are covered by the segment (which may vary among the segments) are stored in the measures attribute. The computed stress value (computedStress attribute) and the therefrom derived stress level (computedStresslevel) are also stored for each segment. The computed stress level can be either not stressed (computedStresslevel = 0), stressed (computedStresslevel = 1) or non-available (computedStresslevel = -1) if no computed stress value exists for the time interval of the segment. The labeledStressType of a segment is derived from the feeling attribute of the corresponding activity and can therefore be either neutral (labeledStressType = 0), positive (labeledStressType = 1), negative (labeledStressType = 2) or unlabeled (labeledStressType = -1) if no corresponding activity exists. The computedStressType combines the computedStresslevel and the labeledStressType and can be either rest (if computedStresslevel = 0 -> computedStressType = 0), eustress (if computedStresslevel = 1 and labeledStressType = 1 -> computedStressType = 1), distress (if computedStresslevel = 1 and labeledStressType = 2 -> computedStressType = 2) or unclassified (else -> computedStressType = -1). Figure 2.27 shows the mapping from the combination of computed stress level and labeled stress type to computed stress type compact in a table. Lastly, each segment contains a list of RR-intervals, which are stored in an array named rrList and, if already calculated, the eight HRV features derived from this list.
2.5.3 The statistics collection

In this collection statistics are collected from the confirmation/refusal step of the classification process (see section 2.3.2.6). Like the users collection the statistics collection has one entry for every user who is registered in the system. So each entry has an unique id (the user ID) that serves as primary key for the collection and which identifies the corresponding user of the statistics entry. The following statistics are collected for each user:

- **confirmedSegments**: The number of confirmed segments.
- **refusedSegments**: The number of refused segments.
- **truePositive**: The number of correctly classified eustress segments. These are all as eustress classified segments that were confirmed.
- **falsePositive**: The number of incorrectly classified distress segments. These are all as eustress classified segments that were refused.
- **trueNegative**: The number of correctly classified distress segments. These are all as distress classified segments that were confirmed.
- **falseNegative**: The number of incorrectly classified distress segments. These are all as distress classified segments that were refused.

2.6 Visualization concept

In order to display all the emerged information a visualization concept has to be elaborated. This shall be done in this chapter. So the main questions that have to be tackled are:
2.6. VISUALIZATION CONCEPT

• **What information shall be displayed?**
  Since there is a wealth of information it is important to just show relevant data that provides an added value for the user and not to overstrain the user with extraneous data.

• **Shall the information be displayed unwrought or aggregated?**
  Some information may be more meaningful when it is aggregated, whereas for the meaningfulness of other information it is important to display it unwrought.

• **How shall the information be displayed?**
  This is often a question of aesthetics and individual preferences, but it is also important to consider the usability of the visualization components.

• **Who shall be able to see which information?**
  Not all users are allowed to read and modify all of the information.

In the following sections the conception of the three main components of the visualization are presented. Section 3.2 then covers the visualization capabilities in more detail.

2.6.1 Home screen

After logging in into the web portal the users shall be directed to a general view of their individual data. This view is called home screen and is only accessible by the respective user. Figure 2.28 shows a sketch of the home screen. The home screen provides a calendar, where all the reported activities are listed (component 1). By selecting a particular activity or a whole day, the corresponding raw heart rate measures and the computed stress values are displayed (component 3 and 4). Lastly, the home screen provides an aggregated view of the activities, where the share of the distinct feelings is shown (component 2).

![Figure 2.28: Sketch of the home screen.](image-url)
2.6.2 Detail view

In addition a detail view of the individual data of the users shall be provided. Beside the respective user this view is also accessible by administrators (but just anonymized). Figure 2.29 shows a sketch of the detail view. The detail view provides a timeline, where the reported activities and the calculated segments (representing rest, eustress and distress) are displayed (component 1). So this component provides the feedback, when a user had eustress and when distress. As in the home screen, the raw heart rate measures and the computed stress values are displayed for the specified time range (component 2 and 3), but this time in a bigger diagram. A time range can be specified by selecting the corresponding period of time in the timeline. Lastly, in component 4 and 5 the characteristics of the various HRV features are displayed by means of eustress and distress patterns for a specified time range.

![Figure 2.29: Sketch of the detail view.](image)

2.6.3 Admin view

Furthermore an admin view shall be provided, where the data of multiple users can be displayed. In this view the users are anonymized and it is just accessible with a higher permission (lecturers or administrators). Figure 2.30 shows a sketch of the admin view. In the admin view a set of users can be selected, for whom the data is then displayed (component 2). In component 1 the admin view provides a timeline, where the calculated segments (representing rest, eustress and distress) are displayed for every user, who is selected in component 2. Again a time range can be specified by selecting the corresponding period of time in the timeline. The raw heart rate data and the computed stress values are then displayed for every selected user in the specified time range (component 3 and 4). In component 5 the number of users is shown, who experienced eustress or distress at a
specific moment. Lastly, the share of rest, eustress and distress in the specified time range is displayed for every selected user (component 6).

Figure 2.30: Sketch of the admin view.
In this chapter the implementation of the classification process (see section 2.3.2) and the implementation of the visualization concept (see section 2.6) are described. The system is implemented after the client-server model, where the server (back-end) is responsible for the storage and the processing of the data and the client (front-end) is responsible for the visualization of the data. The communication between the clients and the server is realized by a REST API.

3.1 Back-end

3.1.1 Used technologies

3.1.1.1 Play Framework

The Play Framework\(^1\) is used to build the web application. It is a stateless, asynchronous and non-blocking web development framework, which is built on the open source toolkit Akka\(^2\). Play is open source and provides both Scala and Java APIs (the latter is used in this work).

3.1.1.2 MongoDB

As database MongoDB\(^3\) is used, one of the leading NoSQL databases. MongoDB is an open source document-oriented database with dynamic schemas that uses a JSON data model. In MongoDB database entries are BSON documents, which are a binary representation of JSON documents. A BSON document is a data structure composed of field-value pairs, where the value of a field can be a BSON data type\(^4\), an other document, an array or an array of documents. The BSON documents are then stored in collections (groupings of documents), which are analogous to tables in relational databases.

---

\(^1\) [https://www.playframework.com](https://www.playframework.com)
\(^2\) [http://akka.io](http://akka.io)
\(^3\) [https://www.mongodb.com](https://www.mongodb.com)
\(^4\) [https://docs.mongodb.com/manual/reference/bson-types/](https://docs.mongodb.com/manual/reference/bson-types/)
CHAPTER 3. IMPLEMENTATION

Furthermore each document stored in a collection requires a unique id field (named \_id) that acts as primary key, where the \_id field is always the first field in the documents and may contain values of any BSON data type, other than an array. For example in the documents of the users collection the \_id field would be the user's id. In other collections where a timestamp represents the primary key (for example in the heartrate collection) a Timestamp object is used as value for the \_id field (see figure 3.1).

```json
{
   "\_id": {"timestamp": NumberLong(1467410590)},
   "heartrate": 78,
   "processed": 0
}
```

Figure 3.1: A MongoDB document of the heartrate collection, with exemplary data.

In order to work with the MongoDB database from Java (the used programming language for the back-end) MongoDB provides a Java Driver for the interaction with MongoDB. By using JONGO (a wrapper for the MongoDB Java Driver) it is then possible to make database queries in Java as in the MongoDB shell.

3.1.1.3 Apache Spark MLlib

Apache Spark is an open source large-scale data processing engine, which is intended to run on a cluster. Apache Spark's machine learning library MLlib (which is built upon the Apache Spark Core) is used for the classification procedure (see section 3.1.5). The MLlib contains many machine learning algorithms and utilities including a decision tree algorithm and a linear SVM, which are used in the classification procedure. In this work the RDD-based API of the MLlib library is used.

The so called resilient distributed dataset (RDD) is the main abstraction that Spark provides. A RDD is a collection of elements that are partitioned across the nodes of the cluster and that can be operated on in parallel. RDDs can be created either with a file in the Hadoop file system or by transforming an existing collection in the driver program. The latter is done in this work, by fetching the desired data from the database (for example the segments that shall be classified) and distributing it to form an RDD. So all following operations in the classification procedure are then performed on RDDs.

In order to run Spark a so called JavaSparkContext object has to be created, which is the main entry point for Spark functionality. The JavaSparkContext represents the connection to a Spark cluster and is inter alia used to create RDDs. The JavaSparkContext is configured via a SparkConf object that provides information about the Spark application. The most important configuration that has to be specified is the master URL, which determines where to run Spark. There are several possibilities to run Spark:

- Run Spark locally with a specified number of worker threads
- Run Spark on a Spark standalone cluster

[http://jongo.org](http://jongo.org)
[http://spark.apache.org](http://spark.apache.org)
3.1. BACK-END

- Run Spark on a Mesos cluster\textsuperscript{11}
- Run Spark on a YARN cluster\textsuperscript{12}

In this work, by default Spark runs locally with as many worker threads as logical cores are available on the machine. But via a configuration variable called \textit{sparkMaster} (see table \ref{tab:spark-config}) the master URL can be changed.

3.1.2 Overview

Figure \ref{fig:back-end-diagram} shows a class diagram of the back-end. The following listing describes briefly some of the principal components.

- The \texttt{InputDataController} corresponds to the data input component in the system design (section \ref{sec:system-design}) and is thus responsible for the data synchronization with the CStress app (section \ref{sec:dataset}), the upload of activities via the user interface (section \ref{sec:activity-upload}) and the confirmation or refusal of classified segments (section \ref{sec:classification}).

- The \texttt{OutputDataController} provides all the data that is required for the visualization component (section \ref{sec:visualization}), by responding to HTTP requests from the client.

- The \texttt{UserController} provides user-related actions like the login of users into the client, the access control of users for distinct functionalities and the adaptation of access rights.

- The \texttt{DownloadController} handles the download of heart rate data via the user interface (section \ref{sec:download}) and the download of data for further analysis (section \ref{sec:data-analysis}).

- Via the \texttt{SparkController} the evaluation procedure for the decision tree classifier and for the SVM can be executed (section \ref{sec:evaluation}).

- The \texttt{Preprocessor} implements the segmentation and labeling of the data (section \ref{sec:segmentation}) and the \texttt{Classificator} implements the creation of the classification model and the class prediction (section \ref{sec:classification}.

- The \texttt{DaemonService} comprises an instance of the \texttt{PreprocessorDaemon} and of the \texttt{ClassificationDaemon} in a queue and is thus responsible for their execution (section \ref{sec:daemon-service}).

- The classes \texttt{User}, \texttt{Statistics}, \texttt{Heartrate}, \texttt{HRVData}, \texttt{Activity} and \texttt{Segment} represent the in section \ref{sec:database} introduced database collections.

- The \texttt{DBService} implements an interface between the application and the database. Every database query is handled by this class.

- The \texttt{Authentication} class is responsible for the authentication of HTTP requests to the server. It is important to notice that all requests to the server are secured with a password.

\textsuperscript{11} \url{http://spark.apache.org/docs/latest/running-on-mesos.html}
\textsuperscript{12} \url{http://spark.apache.org/docs/latest/running-on-yarn.html}
Figure 3.2: Class diagram of the back-end.
3.1. BACK-END

3.1.3 Data synchronization

For each user, the required data for the classification process (see section 2.3.2) is sent to
the server as a JSON data object via a HTTP POST request from the CStress app. Figure 3.3
shows the structure of such a JSON object with exemplary data.

```json
"secret": "dummy",
"data": {
"profile": [
{"name": "Joe Doe",
"matricula": "12345678"
},
"heart_rate": [
{"timestamp": 1466979015,
"heart_rate": 155
},
...],
"hrv_data": [
{"timestamp": 1466979000,
"hrv_score": 55.06634561053727,
"computed_stress": 64.93365438946273
},
...],
"activities": [
{"start_timestamp": 1466978400,
"end_timestamp": 1466982000,
"activity_name": "LEARNING",
"feeling": "POSITIVE"
},
...]
}
```

Figure 3.3: Structure of the JSON data object received by the server, with exemplary data.

The first attribute of the JSON object is the secret, which is a password that is used by
the server to check the authorization of the JSON object. The authorization is required to ensure
that the received data comes from the CStress app and not from any other unauthorized
source. If the authorization of the received JSON object was successful, then the transferred
data can be stored in the database, otherwise the server responds with a 401 Unauthorized
result to the HTTP POST request.

The second part of the JSON object contains the actual data collected by the CStress app
in the data attribute. This comprises the personal data of the user (profile attribute), a list
of heart rate measures (heart_rate attribute), a list composed of the calculated HRV score
and the therefrom derived computed stress (hrv_data attribute) and a list of the reported
activities by the user (activities attribute). The profile of a user comprises his or her name and
student number (here called matricula), where the matricula corresponds to the id of a user in
the database.
Figure 3.4 shows the data synchronization procedure in a simplified sequence diagram, after the received JSON object was authorized. First the matricula is used to search the users collection of the database for the corresponding user. If no such user exists, a new user is created and inserted into the users collection of the database (with the matricula as id and the specified name). Then all entries of the heart_rate list are saved in the heartrate collection of the corresponding user, all entries of the hrv_data list are saved in the hrvdata collection of the corresponding user and all entries of the activities list are saved in the activities collection of the corresponding user. Notice that the thereby inserted heart rate measures and activities are all tagged as unprocessed.

If the data synchronization is finished successfully the server responds with a 200 OK result to the HTTP request and runs the data preprocessing and the classification of the segments for the just synchronized user directly in an asynchronous task. So first the preprocessing of the data is started by calling the calculateSegments-method for the corresponding user and once the preprocessing is finished the classification of the created segments is started by calling the classifySegments-method (see figure 3.5).
3.1 BACK-END

Figure 3.5: Sequence diagram of the method invocations for the data preprocessing and classification.

3.1.4 Data preprocessing

The data preprocessing procedure (calculateSegments-method) corresponds to the first two steps of the classification process in section 2.3.2 (data segmentation and labeling) and is performed for one specified user. This means that only the data of this user is considered (the collections in the database with the user’s prefix). Figure 3.6 shows the preprocessing procedure in a simplified sequence diagram.

Figure 3.6: Sequence diagram of the preprocessing procedure.

In order to divide the heart rate measures into disjunct 10-minute segments the preprocessing procedure considers the data of consecutive 10-minute intervals (where the startTime defines the beginning of the respective intervals) until a defined stopTime is reached. To avoid redundant work just unprocessed data is considered in the preprocessing procedure. This means that the preprocessing procedure starts with the timestamp (rounded off to
the last full hour) of the earliest unprocessed heart rate entry of the `heartrate` collection or with the timestamp of the earliest unprocessed activity of the `activities` collection. And the preprocessing procedure ends when the latest unprocessed heart rate entry of the `heartrate` collection is reached or when the latest unprocessed activity of the `activities` collection is reached.

For each considered 10-minute time interval in the preprocessing procedure the following steps are conducted. First all heart rate measures that lie inside the considered 10-minute interval are retrieved from the database and converted into a list of RR-interval values. For doing this the minimum time between two consecutive RR-intervals can be specified in a configuration variable called `minRRIntervalGap` (see table 3.1). The default value of the `minRRIntervalGap` is 5 seconds. This value was chosen because the heart rate is measured normally every 5 to 15 seconds, leading to approximately 60 measures per segment. So if there are some segments with measures every second, this would lead to 600 measures per segment without a `minRRIntervalGap`. But if the measures per segment varies this much the segments would not be very good comparable.

Before proceeding with the next steps it is checked if enough RR-interval values exist for the considered time interval. For doing this the minimum number of RR-interval values can be specified in a configuration variable called `minSegmentValues` (see table 3.1). If there are enough values the procedure proceeds with the next steps for the considered time interval and a segment will be created, otherwise the currently considered 10-minute interval is discarded and the procedure continues with the next 10-minute time interval.

In the next steps first the computed stress value for the 10-minute interval is determined by retrieving the corresponding `hrvdata` entry from the database. Then the `labeledStressType` is determined by retrieving all activities from the database, which were reported in the considered time interval. For doing this in a configuration variable called `allowMultipleActivitiesPerSegment` (see table 3.1) it can be specified if multiple activities per segment are allowed or not. If true the activity with the greatest coverage of the considered time interval is used for determining the `labeledStressType`, else the segment is marked as unlabeled if multiple activities exist for the considered time interval. Furthermore the minimum amount of the time interval that has to be covered by the activity is specified in a configuration variable called `minActivityAmount` (see table 3.1). This means that an activity is just allowed if it covers the considered time interval with the specified amount. If an activity could be found that corresponds all conditions the `feeling` attribute of this activity is used to determine the `labeledStressType`.

Now the segment for the considered 10-minute time interval can be created as described in section 2.5.2.4. Additionally all `HRV` features are calculated and stored directly after the segment was created, in order to avoid that the features have to be calculated each time they are required. Lastly, the just created segment is saved in the `segments` collection of the database and the next 10-minute time interval is considered (if the `stopTime` is not yet reached).

Once all data is processed (`startTime ≥ stopTime`) all heart rate entries and all activities of the just preprocessed user are marked as processed in the database. Since the preprocessing procedure always considers all unprocessed data, after the preprocessing is finished all data can be marked as processed.
3.1.5 Classification

The classification procedure (classifySegments-method) corresponds to the next three steps of the classification process in section 2.3.2 (feature extraction, train classifier and predict class label) and is also performed for one specified user. The classification is conducted using Apache Spark’s machine learning library MLlib (see section 3.1.1.3). It is possible to choose either a decision tree classifier or a linear SVM for the classification procedure. This is done by specifying a configuration variable called classificationAlgorithm (see table 3.2).

3.1.5.1 Classification procedure with the decision tree classifier

Figure 3.7 shows the classification procedure with the decision tree classifier in a simplified sequence diagram.

First all segments that are labeled as eustress or distress have to be loaded, in order to utilize them to train the classifier. Therefore all non-neutral labeled stress segments (all segments with a class label of either eustress or distress) are retrieved from the database, either only from the considered user or from all users (explanation follows). These are all segments, which are marked as stressed (computedStresslevel = 1) and labeled either positive or negative (labeledStressType > 0). The decision if the segments are used from all users or only from the considered user is dependent on the available number of labeled segments from the considered user. The primary goal is to train the classifier only with the data of the considered user, but if there are not enough labeled segments available from the considered user to create an appropriate classification model, the segments of all users are used. What is “enough” can be specified in a configuration variable called minTrainingData (see table 3.2).

In order to train the classifier, each segment has to be converted into a LabeledPoint (a class from the MLlib API that represents the features in a vector and the binary label of a data point as 0 or 1). To convert a segment into a LabeledPoint the HRV features that
shall be applied in the classifier are extracted from the segment and composed in the vector of the `LabeledPoint` and the `computedStresstype` of the segment, which is either eustress or distress (just as eustress or distress labeled segments are considered here) determines the label of the `LabeledPoint` (0 for eustress and 1 for distress). With these labeled points as training set the decision tree classifier is trained and a classification model is created.

To classify the unlabeled segments, all unlabeled stress segments (all segments without a class label) from the considered user are retrieved from the database. These are all segments, which are marked as stressed (`computedStresslevel` = 1) but are not labeled as positive or negative (`labeledStresstype` < 1). Then for each unlabeled segment the HRV features that are applied in the classifier are extracted from the segment and composed in a vector. This vector is then used by the previously created classification model to predict the class label of the corresponding segment (either 0 for eustress or 1 for distress). The unlabeled segments in the database are then updated according to their prediction. This means that if the predicted class label is 0 (for eustress), the `computedStresstype` of the segment is set to 1 (eustress) and if the predicted class label is 1 (for distress), the `computedStresstype` of the segment is set to 2 (distress).

So how can be determined if the stress type of a segment with a `computedStresstype` > 0 was labeled or computed? This can be easily determined by looking at the `labeledStresstype`. If the `labeledStresstype` is greater than 0, the stress type of the segment was labeled and if the `labeledStresstype` is less or equal 0, the stress type of the segment was computed.

### 3.1.5.2 Classification procedure with the linear SVM classifier

Figure 3.8 shows the classification procedure with the SVM classifier in a simplified sequence diagram.

![Figure 3.8: Sequence diagram of the SVM classification procedure.](image)

The classification procedure with the SVM classifier works analogous to the classification procedure with the decision tree classifier except for two points. Of course, now a SVM
classification model is trained and then used for the prediction and not a decision tree model. And in contrast to the decision tree classifier, here all extracted \textit{HRV} features are normalized to lie in the interval [0,1] and standardized to have unit variance and zero mean.

The normalization is done by applying a min-max normalization. Therefore, the minimum value (\textit{min}) and the maximum value (\textit{max}) of each feature is determined (considering both the labeled and unlabeled segments of the classification process). Then equation \ref{eq:norm} is applied to all \textit{n} features (feature\textsubscript{1}, feature\textsubscript{2},..., feature\textsubscript{n}) of each segment.

\begin{equation}
\text{normalized feature}_i = \frac{\text{feature}_i - \text{min}_i}{\text{max}_i - \text{min}_i}
\end{equation}

The standardization of the \textit{HRV} features is performed with the \texttt{StandardScaler} of Apache Spark’s MLlib after the features were normalized.

### 3.1.6 Daemons for the preprocessing and classification procedures

Beside the preprocessing and classification of the data directly after new data was synchronized with the server and stored in the database, there are also two daemons, which start the preprocessing and classification procedure periodically. These daemons were implemented in order to react to failures in the preprocessing and classification procedures. This means that if for any reason the execution of the preprocessing or classification procedure fails, it is retried periodically. Both daemons are extended from an abstract super class called \texttt{TemplateDaemon}, which defines the basic structure of a daemon. This abstract class can be also the starting point for the implementation of additional daemons, which could accomplish further tasks if necessary.

The \texttt{PreprocessorDaemon} starts the preprocessing procedure (see section \ref{sec:preprocess}) periodically every 12 hours for each user. But since the preprocessing procedure is just executed if there is unprocessed data in the database no overhead arises by the \texttt{PreprocessorDaemon}. The \texttt{ClassificationDaemon} starts the classification procedure (see section \ref{sec:classify}) also periodically every 12 hours for each user. This means that every 12 hours a new classification model is created and all unlabeled segments are classified again.

The time interval for the execution of the preprocessing and the classification procedure (which is 12 hours by default) can be specified in a configuration variable called \texttt{executionInterval} (see table \ref{tab:config}), where every daemon has its own configuration variable so that the execution intervals can be specified independently. Furthermore, each daemon has a configuration variable called \texttt{delay} (see table \ref{tab:config}), which defines the time from the system startup till the first execution of the respective daemon. By varying the delay of the daemons it can be ensured that they do not work at the same time.

### 3.1.7 Confirmation and refusal of classified segments

Via the detail view of the implemented user interface (see section \ref{sec:ui}) the users can confirm or refuse before classified segments (see section \ref{sec:confirm}). The server then receives a JSON data object via a HTTP POST request from the web client of a user. This JSON object contains the \textit{id} of the user, a password for the authentication, the start time and end time of the desired time range in that the segments shall be confirmed or refused and a Boolean variable called

---

\footnote{http://spark.apache.org/docs/latest/mllib-feature-extraction.html}
CHAPTER 3. IMPLEMENTATION

confirm that indicates if the segments shall be confirmed (confirm = true) or refused (confirm = false). Figure 3.9 shows a simplified sequence diagram of the procedure for the confirmation and refusal of classified Segments after the received JSON object was authorized.

Figure 3.9: Sequence diagram of the confirmation and refusal of classified segments.

Since statistics from the confirmation and refusal of classified segments shall be saved in the database (see section 2.5.3), first the user ID is used to fetch the corresponding statistics entry from the statistics collection of the database. Next all classified segments in the specified time range are fetched from the database. These are all unlabeled segments (labeledStresstype \( \leq 0 \)) that were classified either as eustress or as distress (computedStresstype > 0).

Then all of these segments are labeled according to the value of the confirm variable (see figure 2.24) by updating the labeledStresstype to positive (labeledStresstype = 1) for eustress or to negative (labeledStresstype = 2) for distress. Additionally new activities for the labeled segments are created and saved in the database, so that the labels are stored persistently even if for any reason the segments are deleted and calculated again (for example because of changed parameter settings). Lastly, for every confirmed or refused segment the statistics entry of the user is updated consequently. For detailed information about the statistics see section 2.5.3 of the database concept.

3.1.8 Further features

This section provides a list of further features, which are described briefly. All these features are just available for system administrators and therefore protected with an admin password.

- Manually starting the preprocessing and classification procedure:
  Via two separate URLs it is possible to start the preprocessing and the classification procedure for all users manually.
3.1. BACK-END

- **Mark data as unprocessed:**
  Via an URL it is possible to mark all heart rate entries and activities in the database as unprocessed. In this way it is possible to preprocess the data again after configuration parameters were changed.

- **Download of data for analysis:**
  Via an URL it is possible to download data from the server for further analysis. This includes the following data:

  1. The labeled segments of all users in WEKA’s Attribute-Relation File Format (ARFF) [14](https://weka.wikispaces.com/ARFF), so that the segments can be further analyzed with WEKA. For each user one file is created and one file with the data of all users together.

  2. A table containing for each user the average value, the standard deviation, the minimum value and the maximum value of each HRV feature separate for the rest (not stressed) segments, the eustress labeled segments and the distress labeled segments. Furthermore the table contains for each user the difference between the average value of the rest segments and the average value of the eustress segments, the difference between the average value of the rest segments and the average value of the distress segments and the difference between the average value of the eustress segments and the average value of the distress segments. By default all features are normalized to lie in the interval [0,1].

- **Evaluation of the decision tree and SVM classifier:**
  Via separate URLs it is possible to evaluate the decision tree and the SVM classifier either with the data of all users together or only with the data of a single user. For the evaluation just labeled segments are considered. The evaluation is performed with a percentage split and with a k-fold cross-validation:

  1. For the percentage split evaluation the data set is randomly split into disjunct sets of training data and test data. The percentage of training data and test data can be specified in the configuration variable `percentageSplit` (see table 3.2). Then the training data is used to train the classifier and the test data is used to test the classifier.

  2. For the k-fold cross-validation the data set is randomly divided up into k parts with approximately equal proportions of the classes. The parameter k can be specified in the configuration variable `kFold` (see table 3.2). Then k-1 parts are used to train the classifier and the remaining part is used as test data to test the classifier. This is done k times, so that every part is once used as test data. The results of each of the k classifications are then aggregated, so that the classification performance can be evaluated for the whole data set.

For both evaluation methods several performance measures are calculated (inter alia a confusion matrix, the accuracy and the $F_1$ score). Both evaluation methods are performed for multiple iterations, where the number of iterations can be specified in a configuration variable called `evaluationIterations` (see table 3.2). After all iterations were performed the minimum, maximum and average value of the aforementioned performance measures are calculated.
3.1.9 Configuration parameters

Configuration parameters for the preprocessing procedure

Table 3.1: Configuration parameters for the preprocessing procedure.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Domain</th>
<th>Default value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrvdataLength</td>
<td>time in seconds</td>
<td>600</td>
<td>The time interval of the HRV data from the CStress app.</td>
</tr>
<tr>
<td>segmentLength</td>
<td>time in seconds</td>
<td>600</td>
<td>Specifies the length of the segments. Should be the same as the hrvdataLength. If the values differ the computed stress has to be calculated extra (not yet implemented).</td>
</tr>
<tr>
<td>calculateAllSegments</td>
<td>boolean</td>
<td>true</td>
<td>If false, segments are created only for time intervals containing an activity. If true, segments are created for all time intervals.</td>
</tr>
<tr>
<td>minSegmentValues</td>
<td>positive integer</td>
<td>60</td>
<td>The minimum number of RR-intervals in a segment.</td>
</tr>
<tr>
<td>minRRIntervalGap</td>
<td>time in seconds</td>
<td>5</td>
<td>The minimum time between two consecutive RR-intervals.</td>
</tr>
<tr>
<td>allowMultiple ActivitiesPerSegment</td>
<td>boolean</td>
<td>false</td>
<td>If false, segments are marked as unlabeled if multiple activities exist for the considered time interval. If true, the activity with the greatest coverage of the considered time interval is used to label the segments.</td>
</tr>
<tr>
<td>minActivityAmount</td>
<td>floating-point number in the interval [0,1]</td>
<td>0.8</td>
<td>The minimum amount of the considered time interval that has to be covered by the activity.</td>
</tr>
<tr>
<td>stressThreshold</td>
<td>positive floating-point number</td>
<td>50.0</td>
<td>Every segment with a computed stress value smaller than the threshold is considered as not stressed and every segment with a computed stress value bigger than the threshold is considered as stressed.</td>
</tr>
</tbody>
</table>
### Configuration parameters for the classification procedure.

Table 3.2: Configuration parameters for the classification procedure.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Domain</th>
<th>Default value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sparkMaster</td>
<td>String with the master URL</td>
<td>&quot;local[*]&quot;</td>
<td>The master URL to connect to, which determines where Spark runs. For example &quot;local&quot; to run locally with one worker thread, &quot;local[4]&quot; to run locally with 4 worker threads, &quot;local[*]&quot; to run locally with as many worker threads as logical cores on the machine or a cluster URL.</td>
</tr>
<tr>
<td>classificationAlgorithm</td>
<td>&quot;DT&quot; or &quot;SVM&quot;</td>
<td>&quot;DT&quot;</td>
<td>Determines the applied classification algorithm. &quot;DT&quot; = decision tree; &quot;SVM&quot; = support vector machine</td>
</tr>
<tr>
<td>minTrainingData</td>
<td>positive integer</td>
<td>50</td>
<td>If a user has less labeled segments the classifier is trained with the segments of all users.</td>
</tr>
<tr>
<td>evaluationIterations</td>
<td>positive integer</td>
<td>10</td>
<td>The number of performed iterations for the evaluation of a classifier.</td>
</tr>
<tr>
<td>percentageSplit</td>
<td>floating-point number in the interval [0,1]</td>
<td>0.7</td>
<td>Determines the amount of training data used to train the classifier in an evaluation with percentage split. The amount of test data to evaluate the classifier is $1 - \text{percentageSplit}$</td>
</tr>
<tr>
<td>kFold</td>
<td>positive integer greater than 1</td>
<td>10</td>
<td>Determines the number of folds for the evaluation with k-fold cross-validation.</td>
</tr>
</tbody>
</table>

### Configuration parameters for the daemon service.

Table 3.3: Configuration parameters for the daemon service.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Domain</th>
<th>Default value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>preprocessorDaemon delay</td>
<td>time in seconds</td>
<td>10</td>
<td>The initial delay till the first execution of the PreprocessorDaemon.</td>
</tr>
<tr>
<td>preprocessorDaemon executionInterval</td>
<td>time in seconds</td>
<td>43200</td>
<td>The time interval for the periodically execution of the PreprocessorDaemon.</td>
</tr>
<tr>
<td>classificationDaemon delay</td>
<td>time in seconds</td>
<td>21610</td>
<td>The initial delay till the first execution of the ClassificationDaemon.</td>
</tr>
<tr>
<td>classificationDaemon executionInterval</td>
<td>time in seconds</td>
<td>43200</td>
<td>The time interval for the periodically execution of the ClassificationDaemon.</td>
</tr>
</tbody>
</table>
3.2 Front-end

This section describes the user interface (UI) of the implemented web application (called "Stila Portal"). It implements inter alia the three visualization components presented in section 2.6 and additionally a data input component and a component for configurations. Via a menu bar, which is located at the top of the page, the user can navigate through the UI (see figure 3.10). The home screen, the detail view, the admin view and the data input component are covered in the following sections. The configuration component just gives the administrator the ability to configure the access rights of the users.

![Figure 3.10: The menu bar.](image)

### 3.2.1 Implementation of the home screen

After logging in into the web portal, the user is directed to a general view called home screen that is accessible for all users and displays the user’s individual data (see figure 3.11). In this view the visualization conception of section 2.6.1 is implemented. It comprises a calendar with activities (component 1), a chart for the heart rate data (component 3) and the computed stress (component 4) and an aggregated view of the activities (component 2 respectively figure 3.12).

![Figure 3.11: Extract of the home screen.](image)

The calendar (component 1) displays all the reported activities of the user, color-coded with black for neutral activities, green for positive activities and red for negative activities. It can be switched between a monthly view, a weekly view and a daily view. By selecting a
whole day or just an activity in the calendar the corresponding heart rate data measures and computed stress values are displayed in line charts (component 3 and 4). The charts are colored in the background according to the activities, furthermore in the computed stress chart the stress threshold is depicted as dashed yellow line. If not all values fit into the chart they are automatically aggregated. Via the download-button the raw heart rate measures of the selected day can be downloaded.

The aggregated view of the activities (figure 3.12) is shown by clicking on the corresponding button (component 2 of the home screen). This includes a bar chart, which displays for each activity type the amount of time (either in hours or in percent) that was reported as neutral, positive or negative feeling and a pie chart, which displays the overall amount of time that was reported as neutral, positive or negative feeling. The aggregated activities are always shown for the time range, selected in the calendar.

Figure 3.12: Aggregated activities in the home screen.

3.2.2 Implementation of the detail view

From the home screen the user can navigate to a detailed view of his or her data. In this view the visualization conception of section 2.6.2 is implemented. The detail view (pictured in figure 3.13) primarily serves to display the computed eustress and computed distress. Therefore a timeline was implemented, which displays again the reported activities of the user and additionally the calculated segments (component 1). The segments are color-coded with blue for rest segments (classified as not stressed), green for eustress labeled segments and red for distress labeled segments. Segments that are not labeled and also not yet classified are colored in transparent grey. The timeline can be also displayed in a monthly view, a weekly view or a daily view.

Once the unlabeled segments (displayed as transparent grey in figure 3.13) were classified, their predicted class is illustrated with transparent green for computed eustress and transparent red for computed distress (see figure 3.14). Now the user has the possibility to confirm or to refuse the predicted class of the classified segments. This is done by selecting the segments in the timeline that shall be confirmed or refused and then submit the decision and send it to the server by clicking on the corresponding button. As a result, server-sided the selected segments are labeled according to the desired action and activities are created,
which represent these new labeled segments (see section 3.1.7).

So after classified segments were confirmed or refused by the user, the resulting activities (all named RATED) are displayed in the timeline and the corresponding segments are adjusted. In this example the as distress classified segments from 7:40 pm till 9:30 pm were confirmed and thus are now labeled as distress and the as distress classified segments from 9:40 pm till 10:50 pm were refused and thus are now labeled as eustress (see figure 3.15).
3.2. FRONT-END

according to the corresponding segments. Furthermore, with the date pickers (component 2) the time range can be selected in which the charts are displayed.

Another important part of the detail view are the spider charts (see figure 3.16). Multiple spider charts for different time ranges can be added to the detail view by specifying the desired time range with the aforementioned date pickers (component 2 of figure 3.13). In the spider charts the values of all in the classification process used HRV features are displayed separate for rest segments, eustress segments and distress segments, where both labeled and classified segments are considered. All features are normalized to lie in the interval [0,1] so that they can be better compared with each other. The spider charts can be displayed either for separate segments (chart 1) or aggregated for all segments in the specified time range (chart 2). The goal of this depiction of eustress and distress segments is to be able to better analyze the combination of features for eustress and distress, with the intention to find distinctive patterns.

![Figure 3.16: Spider charts of the detail view.](image)

3.2.3 Implementation of the admin view

Via the menu bar (see figure 3.10) it can be navigated to the admin view, which is just accessible with lecturer and admin access rights. In this view the visualization conception of section 2.6.3 is implemented. Since the admin view is intended for displaying the data of multiple users, via a drop-down menu (see figure 3.17) multiple users can be selected, whose data shall then be displayed.

![Figure 3.17: Selection of users in the admin view.](image)
The selected users are then listed in a timeline (component 1 of figure 3.18), where for each user the segments are displayed as previously described for the detail view. With admin access rights it can be navigated to the detail views of the listed users by selecting a user and clicking on the corresponding navigation button. Via two date pickers (component 2 of figure 3.18) a time range can be specified in which the following charts are displayed. Component 3 and 4 of figure 3.18 display the raw heart rate measures and the computed stress values of all selected users for the specified time range.

![Figure 3.18: First extract of the admin view.](image)

Component 1 and 2 of figure 3.19 display spider charts, where one chart depicts for each user the aggregated feature values of all eustress segments in the specified time range (component 1) and the other chart depicts for each user the aggregated feature values of all distress segments in the specified time range (component 2). Like in the spider charts of the detail view both labeled and classified segments are considered and all features are normalized to lie in the interval [0,1].

Component 3 of figure 3.19 displays an area chart, which illustrates how many users have distress, how many users have eustress and how many users are not stressed throughout the specified time range. This chart could be especially helpful for lecturers to see at a glance, if at a given time a class is rather positively stressed, rather negatively stressed or even not stressed at all. Lastly, component 4 of figure 3.19 illustrates for each user the amount of eustress, distress and rest during the specified time range.
3.2. FRONT-END

3.2.4 Data input

For users with the access right student plus it is also possible to report new activities via the web portal (see figure 3.20). In order to report an activity, the start time, the end time, the activity name and the feeling during the activity has to be specified. The activity name can be chosen freely and the feeling can be selected in a drop-down menu that is either positive, negative or neutral. For the input of the time it is important to notice that it can be switched between local time (the default setting) and Greenwich Mean Time (GMT). GMT is the time zone that has no offset from the Coordinated Universal Time (UTC), which is the time that is used for the timestamps in the database. This means that if “GMT” is selected for the time input the entered time is saved in the database just as specified. If “local time” is selected for the time input the entered time is converted into a UTC timestamp before it is saved in the database. Nevertheless, the local time is what you normally want.

Figure 3.20: Data input via the web portal.
When saving the activity, it is just cached on the client-side and displayed in a list ("Saved Activities" in figure 3.20). There the user can check if all entered data is correct. If not, the user has the option to delete the respective activity. If everything is correct, the user can upload the entered activities to the server, where they are saved in the database.

### 3.2.5 Used technologies

For the front-end implementation Angular 2\(^\text{15}\) was used, an open source JavaScript framework for the building of web applications. It is the successor of AngularJS 1 and is not just a version upgrade but rather an incompatible rewrite\(^\text{16}\). An issue during the implementation phase was that the final version of Angular 2 was not yet released, which caused minor code adjustments from time to time.

For the implementation of the charts (except for the spider charts) the open source charting library Highcharts\(^\text{17}\) was used. It is written in pure JavaScript and supports numerous chart types like the here used line, bar, area and pie charts. Beside the large choice of chart types the main reason to choose this charting library was that it is very good in displaying timeline charts (required for the heart rate and computed stress charts). In order to use the Highcharts library with Angular 2 a wrapper of the Highcharts library is required. Therefore the angular2-highcharts\(^\text{18}\) module was used, which provides Angular 2 charting components based on Highcharts. Because of this wrapper the functionality of the Highcharts was a little bit limited, but still adequate for this work.

For the spider charts and also for most of the other UI elements (buttons, the date picker, the timeline and the calendar) the open source library PrimeNG\(^\text{19}\) was used, which is a collection of rich UI components for Angular 2. An especially important component was the PrimeNG Schedule that is based on FullCalendar\(^\text{20}\) (used for the timelines in the detail view and in the admin view as well as for the calendar of the home screen).

\(^1\) [https://angular.io](https://angular.io)
\(^2\) [https://en.wikipedia.org/wiki/AngularJS](https://en.wikipedia.org/wiki/AngularJS)
\(^3\) [http://www.highcharts.com](http://www.highcharts.com)
\(^4\) [https://www.npmjs.com/package/angular2-highcharts](https://www.npmjs.com/package/angular2-highcharts)
\(^5\) [http://www.primefaces.org/primeng](http://www.primefaces.org/primeng)
\(^6\) [https://fullcalendar.io](https://fullcalendar.io)
4.1 The field study

In order to evaluate the developed classification process and to verify the proposed hypotheses of the pre-study (see section 2.2.4) a field study was conducted. The intention of the study was to record the stress history of students during exams. This setting was chosen because it is assumed that especially during exams students are under acute stress (either positive or negative).

Ten students of the LMU in Munich (all studying informatics or media informatics) participated in the field study. Five of them were female and five were male. The age of the participants ranged from 20 to 30. To collect the required data all ten study participants were equipped with a fitness tracker (Fitbit Charge HR) and became the instructions to use the Android application CStress (developed by Marcel Heil in his bachelor thesis [9]) to record their heart rate and the resultant computed stress during their exams. The study was conducted during the exam period of the summer semester in 2016.

Before and after each exam the participants were asked to answer various questions about how good they were prepared for the exam and about their performance and feelings during the exam. The intention of these questions was to assess, if the exam was a rather positive experience or a negative experience for the students. The following questions were asked in the survey:

1. Was this exam a pleasant experience for you?
2. Have you answered all exam questions?
3. Could you complete your work in due time?
4. Do you think you performed well during this exam?
5. Were any exam questions difficult for you to answer?
6. Were in your opinion the exam questions fair?

[https://www.fitbit.com/de/chargehr]
7. Do you think you will pass this exam?

8. Do you think you will pass this exam with a good grade?

9. Have you prepared yourself sufficiently well for this exam?

The results of the survey questions were then used to label the reported exams either as positive experience or as negative experience. Thereby each exam can be understood as an activity (that is used to label the segments in the classification process), where the usually self-assessed feeling is now determined by the result of the survey. First the exams were labeled as positive or negative according to the first question of the survey, whether the exam was a pleasant experience or not (self-assessment of the mood). If the question was answered with “yes” (the exam was a pleasant experience), the corresponding exam was labeled as positive and if the question was answered with “no” (the exam was an unpleasant experience), the corresponding exam was labeled as negative.

Since the self-assessment of the mood during the exams may not always be accurate and in order to approve the self-assessment an assessment matrix was created, where a score is determined based on the remaining survey questions. The score is increased or decreased depending on how the questions were answered. Figure 4.1 shows the calculation of the score in a table. So after considering all questions, the score can have a minimum value of -7 and a maximum value of 7 (the score lies in the interval \([-7,7]\)). A negative score is then assumed to be an indicator for an unpleasant mood and a positive score is assumed to be an indicator for a pleasant mood.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you answered all exam questions?</td>
<td>score + 1 score - 1</td>
</tr>
<tr>
<td>Could you complete your work in due time?</td>
<td>score + 1 score - 1</td>
</tr>
<tr>
<td>Do you think you performed well during this exam?</td>
<td>score + 1 score - 1</td>
</tr>
<tr>
<td>Were any exam questions difficult for you to answer?</td>
<td>score - 1 score + 1</td>
</tr>
<tr>
<td>Were in your opinion the exam questions fair?</td>
<td>score + 1 score - 1</td>
</tr>
<tr>
<td>Do you think you will pass this exam?</td>
<td>score + 0.5 score - 0.5</td>
</tr>
<tr>
<td>Do you think you will pass this exam with a good grade?</td>
<td>score + 0.5 score - 0.5</td>
</tr>
<tr>
<td>Have you prepared yourself sufficiently well for this exam?</td>
<td>score + 1 score - 1</td>
</tr>
</tbody>
</table>

Figure 4.1: Calculation of the score for the assessment matrix.

Figure 4.2 shows the assessment matrix and in it, how the exams are labeled according to the calculated score. If the score corresponds with the initial self-assessed label (initial positive label and score > 0 or initial negative label and score < 0) the self-assessment is confirmed. If the score does not correspond with the initial self-assessed label but is quite low (the score lies between the interval \([-2,2]\)), then the initial label is retained even the self-assessment is not confirmed. If the score is greater than the absolute value of 2 and does not correspond with the initial self-assessed label (initial positive label and score < -2 or initial negative label and score > 2) the self-assessment is refuted and the exam is labeled newly.
4.1. THE FIELD STUDY

After the application of the assessment matrix three exams were labeled newly (one of them with a score of just -1 because of a crucial comment of the participant) and three exams had a conflicting score but with a value lower or equal to the absolute value of 2. All other exams were confirmed by the assessment matrix. Finally, 33 exams were reported, 15 of them were labeled as positive experience and 18 were labeled as negative experience.

The heart rate data collected during the exams were then processed as described in the classification process in section 2.3.2. This means that the heart rate measures were divided up into 10-minute segments of RR-intervals and the segments were labeled according to the computed stress values and the labels of the exams, where the exam labels correspond to the feelings of the activities in the classification process. This resulted in 263 segments, where 100 segments were classified as not stressed and 163 segments were classified as stressed. So 62% of the time the participants of the field study were stressed according to the stress model of the bachelor thesis of Marcel Heil [9]. Out of these 163 stressed segments 82 segments were labeled as eustress and 81 segments were labeled as distress.

In addition to the data collection during the exams, the participants of the aforementioned field study were asked to wear the fitness tracker during their daily life and report activities via the CStress app as often as possible, in order to collect as much additional data as possible. As a result the data of 4 participants of the exam field study could be collected during a period of approximately one month. Furthermore, two other persons recorded their heart rate and reported activities via the CStress app. By combining these data a second field study emerged with 6 participants, who collected data during their daily life. In order to differentiate the two studies in the further work, the first study is called the “exam field study” and the second study is called the “daily field study”.

In contrast to the exam field study, in the daily field study the participants reported also activities with physical loading (e.g. doing sport). But since physical activities have a strong impact on the physiology and consequently on the heart rate, such activities were filtered out to the end that the data is relatively unbiased from physical loading. After filtering out the physical biased activities 112 activities were reported, where 69 were labeled as positive experience and 43 were labeled as negative experience. By segmenting the collected heart rate data and labeling the created segments (as described in the classification process in section 2.3.2), a set of 1264 segments emerged (804 segments classified as not stressed and 460 segments labeled as eustress or distress). From the 460 labeled segments 184 were labeled as eustress and 276 were labeled as distress.

<table>
<thead>
<tr>
<th>Assessment score</th>
<th>Initial label</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ≥ score ≥ -2</td>
<td>positive</td>
</tr>
<tr>
<td>score &gt; 2</td>
<td>positive</td>
</tr>
<tr>
<td>score &lt; -2</td>
<td>negative</td>
</tr>
</tbody>
</table>

Figure 4.2: Assessment matrix for the labeling of exams.
CHAPTER 4. EVALUATION

4.2 Classification performance

First the data of the exam field study (82 eustress labeled segments and 81 distress labeled segments) was used to evaluate the classification performance. But the problem with the data set of the exam field study is, that there is an insufficient amount of data per participant to create separate models for each participant (the number of segments per participant that are labeled with eustress or distress range from 2 to 35). Furthermore, the distribution of the eustress and distress segments is mostly not balanced. This means that for each participant there are either almost only eustress segments or almost only distress segments. This is especially true for participants with a greater number of segments. For this reason the classification performance was evaluated with the data of all participants together.

Using the same approach as in section 2.2.3.3 of the pre-study, a decision tree classifier and a SVM are used as classification algorithms. But this time the decision tree classifier and the SVM of the Apache Spark machine learning library MLlib are used (see section 3.1.1.3), which are implemented in the developed web application (see section 3.1.5). Since the Apache Spark MLlib provides only a linear SVM for the purpose of evaluating the classification performance with a nonlinear SVM WEKAs LibSVM implementation with a nonlinear kernel is also used.

In total, for the evaluation of the classification performance three algorithms are used:

- The decision tree classifier of the Apache Spark MLlib
- The linear SVM of the Apache Spark MLlib
- The nonlinear SVM of WEKAs LibSVM implementation with a radial basis function as kernel (rbf-kernel)

For the evaluation of the classification performance again a 10-fold cross-validation is performed for several feature combinations. The accuracy, the $F_1$ score and a confusion matrix are used as performance measures (as described in section 2.2.3.3 of the pre-study). Moreover for both SVM classifiers all feature values are scaled to lie within the interval [0,1] and the features are standardized to have zero mean and unit variance.

The following listing shows the six applied feature combinations for the evaluation of the classification performance:

- Feature combination 1: meanRR, SDRR, CVRR, RMSSD, pRR50, ApEn, VS, SVI_surrogate
- Feature combination 2: ApEn, VS, SVI_surrogate
- Feature combination 3: meanRR, SDRR, CVRR, RMSSD, pRR50
- Feature combination 4: SDRR, pRR50
- Feature combination 5: meanRR, SDRR, pRR50, ApEn, VS, SVI_surrogate
- Feature combination 6: meanRR, SDRR, RMSSD, ApEn, VS, SVI_surrogate

- \[ \text{http://spark.apache.org/docs/latest/mllib-decision-tree.html} \]
- \[ \text{http://spark.apache.org/docs/latest/mllib-linear-methods.html} \]
- \[ \text{http://www.csie.ntu.edu.tw/~cjlin/libsvm/} \]
4.2. CLASSIFICATION PERFORMANCE

Table 4.1: Comparison of the classification performance of the decision tree classifier, the linear SVM and the nonlinear SVM for the six different feature combinations.

<table>
<thead>
<tr>
<th>Feature combination</th>
<th>Decision tree</th>
<th>Linear SVM</th>
<th>Nonlinear SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>F₁ score</td>
<td>Accuracy</td>
</tr>
<tr>
<td>1</td>
<td>71.0 %</td>
<td>0.709</td>
<td>65.4 %</td>
</tr>
<tr>
<td>2</td>
<td>55.6 %</td>
<td>0.554</td>
<td>46.9 %</td>
</tr>
<tr>
<td>3</td>
<td>71.0 %</td>
<td>0.710</td>
<td>67.3 %</td>
</tr>
<tr>
<td>4</td>
<td>58.0 %</td>
<td>0.580</td>
<td>58.0 %</td>
</tr>
<tr>
<td>5</td>
<td>63.0 %</td>
<td>0.630</td>
<td>56.8 %</td>
</tr>
<tr>
<td>6</td>
<td>67.9 %</td>
<td>0.679</td>
<td>66.0 %</td>
</tr>
</tbody>
</table>

Table 4.2: Confusion matrix for the three classification schemes (decision tree, linear SVM and nonlinear SVM) for the six different feature combinations (FC 1 to FC 6). The rows represent the labeled segments and the columns represent the predicted segments.

<table>
<thead>
<tr>
<th>Classification schemes</th>
<th>Decision tree</th>
<th>Linear SVM</th>
<th>Nonlinear SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eustress</td>
<td>Distress</td>
<td>Eustress</td>
</tr>
<tr>
<td>FC 1</td>
<td>62</td>
<td>19</td>
<td>53</td>
</tr>
<tr>
<td>FC 2</td>
<td>41</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>FC 3</td>
<td>60</td>
<td>21</td>
<td>47</td>
</tr>
<tr>
<td>FC 4</td>
<td>48</td>
<td>33</td>
<td>57</td>
</tr>
<tr>
<td>FC 5</td>
<td>51</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>FC 6</td>
<td>57</td>
<td>24</td>
<td>57</td>
</tr>
</tbody>
</table>

Table 4.1 summarizes the accuracy and the F₁ score of the classification with the decision tree classifier, the linear SVM and the nonlinear SVM for all six applied feature combinations. It can be seen that the nonlinear SVM achieved the best classification performance for all feature combinations. With an accuracy of 76.7% and F₁ score of 0.767 the best classification performance was obtained from using the nonlinear SVM in conjunction with feature combination 6 (meanRR, SDRR, RMSSD, ApEn, VS and SVIsurrogate). The high F₁ score and the confusion matrix of this classification (see table 4.2) reveal also that the eustress and distress segments were approximately identically good classified, with 65 true positives (correctly classified eustress segments) and 60 true negatives (correctly classified distress segments).

However, with 74.8% accuracy and a F₁ score of 0.748 respectively with 75.5% accuracy and a F₁ score of 0.755 feature combination 1 (all eight features) and feature combination 3 (meanRR, SDRR, CVRR, RMSSD and pRR50) achieved just a marginal lower classification performance. As it can be seen in the confusion matrices the distribution of the correctly classified segments is also quite good for both feature combinations. The distribution of the correctly classified segments for feature combination 3 is even better than for feature combination 1 and 6 (see table 4.2). The classification accuracy of the other three applied
feature combinations (feature combination 2, 4 and 5) was significantly worse (more than 10%).

The decision tree classifier (also a nonlinear classification method), also achieved the best classification performances for feature combination 1, 3 and 6. Nevertheless, with an accuracy of 71% just feature combination 1 and 3 reach roughly the best performances of the nonlinear SVM (but still about 4% below). With 67.9% accuracy feature combination 6 has actually already a significant worse classification performance compared to the nonlinear SVM.

Obviously the linear SVM performed worst for all feature combinations, compared to the decision tree as well as compared to the nonlinear SVM. When considering the three best performing feature combinations, the classification accuracy of the linear SVM (with a maximum of 67.3% for feature combination 3) is nearly 10% worse compared to the nonlinear SVM (indeed for feature combination 6 it is over 10%). Nevertheless, again feature combination 1, 3 and 6 achieved the significantly best classification performances.

So the feature combinations 1, 3 and 6 seem to be the most promising ones (for all three classification algorithms). Figure 4.3 shows the classification performance of these three most promising feature combinations for the decision tree classifier, the linear SVM and the nonlinear SVM graphically in a bar chart.

![Figure 4.3: Accuracy of the three most promising feature combinations for the decision tree classifier, the linear SVM and the nonlinear SVM.](image)

Findings of the evaluation of the classification performance

The first and perhaps most important finding of the evaluation is that eustress and distress can be distinguished with an accuracy of over 75% for the classification with multiple participants, using a nonlinear SVM. Also the distribution of correctly classified eustress and correctly classified distress segments was quite good balanced. This means that not only all eustress or only all distress segments were classified correctly and the others were classified wrong, but that both eustress segments and distress segments were recognized equally good. Whether the achieved accuracy is sufficiently high for the purpose of this
Another finding is that the nonlinear classification methods (decision tree and nonlinear SVM) separate the eustress and distress segments considerably better compared to the linear SVM (about 10% higher accuracy for nonlinear SVM and about 4% higher accuracy for decision tree). Consequently, nonlinear classification methods are required to achieve the best classification performances. This finding suggests that eustress and distress are not reasonable linear separable (at least not with data from multiple persons together).

The poor performance of feature combination 2 and 4 (where just three respectively two features were used) reveal that multiple different features are required to achieve the best classification performances. However, a consistent ranking of the feature combinations among all three classification methods could not be found. But for all three classification methods feature combination 1, 3 and 6 achieved the best performance applying 10-fold cross-validation.

It is important to notice that these statements are admittedly only true for the classification of data from multiple persons together. Classification results with data from individuals may differ from the findings presented in this evaluation.

### 4.3 Pattern analysis

In section 2.2.4 two distinct heart rate patterns for eustress and distress were proposed as a hypothesis on how the characteristics of the HRV features differ among these two states. In this section it is intended to confirm or disprove these heart rate patterns based on the data of the exam field study. In order to analyze the characteristics of the used HRV features, the average values of the single features are compared with each other separately for eustress and distress segments (as in section 2.2.3.1 of the pre-study).

In table 4.3 the average values and standard deviations of the HRV features are shown separately for the eustress and distress labeled segments, where the features are normalized to lie in the interval [0,1]. The last column of the table shows for each feature the difference between the average value of the eustress labeled segments and the average value of the distress labeled segments, where a positive value means that the average value of eustress is larger than that of distress and a negative value means that the average value of eustress is smaller than that of distress.

Table 4.3: Comparison of the average values of HRV features (normalized to the interval [0,1]) during eustress and distress conditions. Furthermore the standard deviation is given in brackets.

<table>
<thead>
<tr>
<th>HRV Features</th>
<th>Eustress</th>
<th>Distress</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. meanRR</td>
<td>0.333 (±0.238)</td>
<td>0.406 (±0.163)</td>
<td>-0.073</td>
</tr>
<tr>
<td>avg. SDRR</td>
<td>0.269 (±0.294)</td>
<td>0.231 (±0.203)</td>
<td>0.038</td>
</tr>
<tr>
<td>avg. CVRR</td>
<td>0.274 (±0.183)</td>
<td>0.220 (±0.200)</td>
<td>0.054</td>
</tr>
<tr>
<td>avg. RMSSD</td>
<td>0.242 (±0.175)</td>
<td>0.249 (±0.147)</td>
<td>-0.007</td>
</tr>
<tr>
<td>avg. pRR50</td>
<td>0.197 (±0.220)</td>
<td>0.252 (±0.224)</td>
<td>-0.055</td>
</tr>
<tr>
<td>avg. ApEn</td>
<td>0.349 (±0.197)</td>
<td>0.291 (±0.208)</td>
<td>0.058</td>
</tr>
<tr>
<td>avg. VS</td>
<td>0.503 (±0.250)</td>
<td>0.497 (±0.215)</td>
<td>0.006</td>
</tr>
<tr>
<td>avg. SVs surrogate</td>
<td>0.224 (±0.149)</td>
<td>0.174 (±0.146)</td>
<td>0.050</td>
</tr>
</tbody>
</table>
What can be seen at a first glance is that the absolute value of the differences between eustress and distress are significantly smaller than in the pre-study. The most distinct difference exists for the meanRR with an absolute value of 0.073, whereas in the pre-study there was a maximum difference of 0.228 for the ApEn and a bunch of features with an difference larger than 0.1. This means that in the exam field study the difference of single features between eustress and distress segments was not clearly.

When comparing the feature characteristics of the exam field study with the assumed feature characteristics of the pre-study (see section 2.2.4) it can be seen that all features (except for the variability score) behave adverse. This means that if in the pre-study it was suggested that for a certain feature the eustress values are higher or lower than the distress values, then in the exam field study it is just the other way around. The variability score is the only feature that seems to confirm the assumed characteristic of the pre-study (in both studies the average eustress value is higher than the average distress value), but in the exam field study the difference is insignificant. So unfortunately, the results of table 4.3 do not approve the assumed heart rate patterns of the pre-study (see section 2.2.4).

One reason for these results might be that in the pre-study the data came just from one person, whereas in the exam field study the evaluation was made with all ten participants together. Thus the results are likely to be biased by the different physical conditions of the participant. To overcome this bias, a further analysis based on data of single participants might be more reasonable.

Analysis of individual data

For the following analysis of the data of single persons just six participants out of ten are regarded because the data of the four other participants of the exam field study compose either only eustress segments or only distress segments. Figure 4.4 shows for each feature the difference between the average value of the eustress labeled segments and the average value of the distress labeled segments per participant (all the features are normalized to lie in the interval [0,1]). A positive value means that the respective feature has tendentially higher values for eustress labeled segments than for distress labeled segments and a negative value means that the respective feature has tendentially lower values for eustress labeled segments than for distress labeled segments. The positive, negative and equal signs in brackets represent the assumed tendencies of the feature characteristics in the pre-study (see section 2.2.4).

It can be seen that none of the users in the exam field study corresponds exactly to the assumed feature characteristics of the pre-study. Only user 4 has a certain similarity, where just the SDRR feature differs from the assumed feature characteristics. But this would be tolerably compatible with the proposed heart rate patterns of the pre-study. Thus user 4 would confirm the proposed heart rate patterns of the pre-study, but what about the others?

One major statement of the proposed heart rate patterns in section 2.2.4 was that the eustress pattern has a higher variability among the heart beats (higher variability score) and is very predictable (many recurring patterns and thus lower ApEn). This is only true for three of the six users, for the others it is just the other way around. And except for the SDRR, the other features also vary to some extent among the study participants. Although the SDRR is the only feature that has the same tendency among all six users (higher SDRR for eustress segments), its values are still the other way around than in the pre-study suggested.
4.3. PATTERN ANALYSIS

Figure 4.4: Comparison of the differences between the average value of the eustress labeled segments and the average value of the distress labeled segments separately for each feature and for each participant of the exam field study.

In total there are no consistent heart rate patterns for eustress and distress identifiable among the users and therefore the proposed heart rate patterns for eustress and distress from the pre-study do not generalize to the participants in the exam field study. In fact it seems that there are no universal heart rate patterns for eustress and distress, but rather the findings suggest that different persons respond with diverse feature characteristics and thus with different heart rate patterns to eustress and distress. This would imply that the characteristic of a single feature does not compulsory indicate if someone has eustress or distress, but rather that the question if someone has eustress or distress is dependent on the combination of the feature characteristics.

Graphical analysis with spider charts

With the previous findings in mind, in a next step the eustress and distress patterns shall be analyzed graphically. The goal is to find distinctive patterns for eustress and distress by means of HRV features. Namely it is intended to consider the combination of the features separately for each eustress and distress segment.

In order to do that the spider charts, which were implemented in the web application (see section 3.2.2) are used to identify feature patterns for eustress and distress. Figure 4.5 displays an example of such a spider chart, which shows the values of the HRV features for one segment. All eight features, used in the classification process are displayed in the spider chart. For a reasonable depiction of the patterns in the spider charts, the features are normalized to lie in the interval [0,1] and the features are ordered ascending according to their average value, beginning at the top of the chart and going clockwise around the chart. So at the top is the RMSSD, then the CVRR, the SDRR, the SVI surrogate, the pRR50, the ApEn, the meanRR and lastly the VS.
CHAPTER 4. EVALUATION

Figure 4.5: Example of a spider chart with normalized features that are ordered ascending according to their average value in a clockwise direction beginning with the RMSSD.

Then all 163 segments (all eustress and distress labeled segments of the exam field study) were examined with the purpose to find distinctive patterns. Thereby 11 distinctive patterns could be found (listed in figure 4.6). For every discovered pattern it was counted how often it occurs for eustress and how often it occurs for distress (column two and three of figure 4.6). In this way a total number of 112 segments (57 eustress segments and 55 distress segments) could be assigned to the 11 patterns. The next step was, that if there exists a clear majority of eustress or distress segments for a pattern, then it was assigned to the respective stress type (last column of figure 4.6).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Count eustress</th>
<th>Count distress</th>
<th>Majority</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:</td>
<td>6</td>
<td>17</td>
<td>Distress</td>
</tr>
<tr>
<td>2:</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>3:</td>
<td>12</td>
<td>11</td>
<td>—</td>
</tr>
<tr>
<td>4:</td>
<td>1</td>
<td>5</td>
<td>Distress</td>
</tr>
<tr>
<td>5:</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>6:</td>
<td>12</td>
<td>4</td>
<td>Eustress</td>
</tr>
<tr>
<td>7:</td>
<td>12</td>
<td>1</td>
<td>Eustress</td>
</tr>
<tr>
<td>8:</td>
<td>7</td>
<td>1</td>
<td>Eustress</td>
</tr>
<tr>
<td>9:</td>
<td>0</td>
<td>3</td>
<td>Distress</td>
</tr>
<tr>
<td>10:</td>
<td>0</td>
<td>8</td>
<td>Distress</td>
</tr>
<tr>
<td>11:</td>
<td>3</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Total:</td>
<td>57</td>
<td>55</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.6: Number of identified patterns for eustress and distress. The first column lists the discovered patterns, the second and third column shows the number of eustress or distress segments that were assigned to the respective patterns and the last column shows to which stress type the majority of the respective pattern were assigned. The last row shows the total number of assigned eustress and distress segments.
As a result of this pattern analysis, 3 patterns for eustress and 4 patterns for distress could be identified. Figure 4.7 shows the identified eustress patterns and figure 4.8 shows the identified distress patterns, both in descending order according to their relevance (where relevance is defined by how obviously a pattern could be assigned to a stress type). Figure 4.9 shows the patterns, which could not be clearly assigned to a stress type.

When comparing the patterns for eustress and distress it can be seen that it is hardly possible to determine the stress type by means of the characteristics of single features, but rather that the combination of the features is crucial. Furthermore, it can be seen that there are multiple different patterns for either stress types, which all differ from each other to some degree. But there are also some similarities among the patterns that can be recognized.

For example a distinctive shape among the eustress patterns is that they have an inbound kink at the top and at the bottom of the pattern (see pattern 6 and 8). This means that for the top kink the variability score and the CVRR are higher, whereas the RMSSD is lower and that for the bottom kink the ApEn and the SVI_surrogate are higher, whereas the pRR50 is
lower. For the distress patterns a distinctive shape is that they have an outward kink at the bottom of the pattern (see pattern 1, 4 and 9). In these distress patterns the pRR50 is higher, whereas the ApEn and the SVIsurrogate are lower. But for both stress types also exceptions exist, which do not correspond to these aforementioned pattern interpretations (see pattern 7 for eustress and pattern 10 for distress). Nevertheless, the aforementioned patterns could be a basis for further investigations, which is out of the scope of this work.

As a conclusive finding of this section it can be said that the proposed heart rate patterns for eustress and distress (which were derived from the pre-study results in section 2.2.4) do not generalize to the participants of the exam field study and that the combination of the features is more decisive than the characteristics of single features.

### 4.4 Correlation of stress intensity and stress type

Another assumption of the pre-study (and also supported by the literature [18]) was that eustress occurs at a lower stress level, whereas distress occurs at a higher stress level.

In figure 4.10 two HRV features (namely the meanRR and the RMSSD) and their correlation with eustress and distress are depicted (data of all participants of the exam field study together). According to the literature both features are reliable indicators for the stress level, where lower meanRR and RMSSD values indicate a higher stress level and higher meanRR and RMSSD values indicate a lower stress level [4, 31, 32].

![Figure 4.10: Scatter plot visualization of meanRR and RMSSD and their correlation with eustress and distress for all users of the exam field study.](image)

For the RMSSD feature there is no difference between eustress and distress segments identifiable. Also for the meanRR there is no clear separation of eustress and distress segments identifiable. In fact, there is a tendency that some eustress segments have a lower meanRR than the distress segments, what would mean that the stress level is even higher for eustress than for distress. So the assumption that eustress has a lower stress level than
4.5. Evaluation of the daily field study

The in section 4.1 introduced second field study (called daily field study) was conducted in order to collect more data and thus be able to further investigate the feature characteristics of eustress and distress. With 460 labeled segments almost three times as much data is available compared to the exam field study. But still, just for two out of six participants of the daily field study there is enough and balanced data available to create a separate classification model. For the other four participants there are just either too few or too unbalanced data available.

As in figure 4.4 of section 4.3 first the characteristics of the used HRV features are analyzed separately for each participant. Therefore, figure 4.11 shows for each feature and for each participant of the daily field study the difference between the average value of the eustress labeled segments and the average value of the distress labeled segments, where all features are normalized to lie in the interval [0,1].

<table>
<thead>
<tr>
<th>HRV features</th>
<th>Difference eu - dl User 1</th>
<th>Difference eu - dl User 2</th>
<th>Difference eu - dl User 3</th>
<th>Difference eu - dl User 4</th>
<th>Difference eu - dl User 5</th>
<th>Difference eu - dl User 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. meanRR (+)</td>
<td>0.043</td>
<td>-0.021</td>
<td>0.015</td>
<td>-0.117</td>
<td>-0.107</td>
<td>-0.052</td>
</tr>
<tr>
<td>avg. SDRR (-)</td>
<td>0.097</td>
<td>0.011</td>
<td>0.036</td>
<td>0.039</td>
<td>0.009</td>
<td>0.029</td>
</tr>
<tr>
<td>avg. CVRR (-)</td>
<td>0.098</td>
<td>0.016</td>
<td>0.071</td>
<td>0.054</td>
<td>0.036</td>
<td>0.051</td>
</tr>
<tr>
<td>avg. RMSSD (+=)</td>
<td>0.084</td>
<td>-0.001</td>
<td>0.080</td>
<td>-0.028</td>
<td>-0.074</td>
<td>-0.035</td>
</tr>
<tr>
<td>avg. pRR50 (+)</td>
<td>0.030</td>
<td>0.008</td>
<td>0.011</td>
<td>-0.021</td>
<td>-0.066</td>
<td>-0.048</td>
</tr>
<tr>
<td>avg. ApEn (-)</td>
<td>0.018</td>
<td>0.077</td>
<td>0.047</td>
<td>0.051</td>
<td>0.142</td>
<td>-0.010</td>
</tr>
<tr>
<td>avg. VS (+)</td>
<td>-0.019</td>
<td>0.053</td>
<td>-0.113</td>
<td>-0.017</td>
<td>-0.181</td>
<td>-0.077</td>
</tr>
<tr>
<td>avg. SVI (-)</td>
<td>0.044</td>
<td>0.013</td>
<td>0.029</td>
<td>0.033</td>
<td>0.067</td>
<td>0.045</td>
</tr>
</tbody>
</table>

Figure 4.11: Comparison of the differences between the average value of the eustress labeled segments and the average value of the distress labeled segments separately for each feature and for each participant of the daily field study.
It can be seen that, as well as in the exam field study, none of the users in the daily field study corresponds exactly to the assumed feature characteristics of the pre-study and again there are no consistent heart rate patterns for eustress and distress identifiable among the users. This supports the findings of the exam field study that there are no general heart rate patterns for eustress and distress, but rather that different persons respond with different heart rate patterns to eustress and distress. But what is noticeable here is that there are two groups of users identifiable where the users have similar feature characteristics. One group is composed of user 1 and user 3 (user group 1) and the other group is composed of user 4, 5 and 6 (user group 2).

So now these two groups were used to further analyze the data set of the daily field study. In order to further analyze the correlation of stress intensity and stress type figure 4.12 and figure 4.13 depict the HRV features meanRR and RMSSD and their correlation with eustress and distress for the two user groups.

When considering user group 2 (see figure 4.12) it can be seen that eustress occurs rather at low meanRR and RMSSD values, whereas distress occurs primarily at higher meanRR and RMSSD values. This would suggest that eustress occurs at a higher stress level and distress occurs at a lower stress level (recall: lower meanRR and RMSSD values indicate a higher stress level and higher meanRR and RMSSD values indicate a lower stress level). So the assumption that eustress occurs at a lower stress level, whereas distress occurs at a higher stress level would be totally disproved.

But taking a look at the scatter plot of user group 1 (figure 4.13), for a start it can be seen that the separation of eustress and distress segments is not as clear as for user group 2 and furthermore there is a tendency identifiable that distress occurs primarily at lower meanRR and RMSSD values and eustress occurs primarily at higher meanRR and RMSSD values. This in turn would suggest that distress occurs at a higher stress level and eustress occurs at a lower stress level. So the assumption that eustress occurs at a lower stress level, whereas distress occurs at a higher stress level would be confirmed.

Figure 4.12: Scatter plot visualization of meanRR and RMSSD and their correlation with eustress and distress for user 4, 5 and 6 of the daily field study (user group 2).
4.5. EVALUATION OF THE DAILY FIELD STUDY

Figure 4.13: Scatter plot visualization of meanRR and RMSSD and their correlation with eustress and distress for user 1 and 3 of the daily field study (user group 1).

Due to these contradictory results of the analysis of the two user groups, it can be concluded that both eustress and distress can occur at a lower as well as at a higher stress level. So also with the data from the daily field study the assumption that eustress has in general a lower stress level than distress could not be confirmed. In fact, it seems that for some people eustress occurs at a lower stress level than distress and for others eustress occurs at a higher stress level than distress. It is also noteworthy that for user 2 (who is not represented in one of the aforementioned user groups) no correlation between the stress intensity and the stress type could be identified. So the results of the analysis of the daily field study support to some extend the findings of the exam field study that there is no general correlation between the stress intensity (indicated by the HRV features meanRR and RMSSD) and the stress type (eustress and distress), which is true for all people.

Lastly, the classification performance of the daily field study was evaluated, once with all users and once with the two aforementioned user groups. This time just the nonlinear SVM with the rbf-kernel is used for the classification. Again the evaluation of the classification performance was made with a 10-fold cross-validation for the three most promising feature combinations (see section 4.2). Figure 4.14 shows the achieved classification accuracy for all users, user group 1 and user group 2 of the daily field study.

With a maximum accuracy of 68.3% the classification with all users of the daily field study performed worst. The classification of the two user groups achieved significant better results (73.9% accuracy for user group 1 and 80.2% accuracy for user group 2). These results suggest that it seems to be more appropriate to create a separate classification model for each user or at least for certain groups of users with similar physiological characteristics and not to create one general model for all users.
4.6 Summary of the field study findings

The following listing summarizes the findings of the field study:

- It is possible to distinguish eustress from distress with an accuracy better than 75%
  - 76.7% accuracy for all participants of the exam field study with a nonlinear SVM and feature combination 6
  - 80.2% accuracy for user group 2 of the daily field study with a nonlinear SVM and feature combination 3

- Nonlinear classification methods are required to achieve the best performances (⇒ eustress and distress seem to be not reasonable linear separable)

- Ranking of the classification algorithms:
  1. Nonlinear SVM
  2. Decision tree
  3. Linear SVM

- The most promising feature combinations are feature combination 1, 3 and 6 (the classification results do not allow a general ranking of the feature combinations)

- The proposed heart rate patterns of the pre-study do not generalize to the participants of the field study (neither for the exam field study nor for the daily field study)

- The combination of the features is more decisive than the characteristics of single features

- There is no general correlation between the stress intensity (indicated by the HRV features meanRR and RMSSD) and the stress type (eustress and distress)

- Separate classification models for each person seem to be more appropriate than a general model for all persons
Conclusion

5.1 Summary and discussion

The intention of this work was to distinguish eustress from distress based on physiological data in order to provide feedback about the emotional state of students. Prior studies about stress mainly considered stress as inherently negative and thus just studied the recognition of general stress. This work is one of the first that tries to separate eustress from distress by means of physiological data. More precisely, heart rate measures collected with a fitness tracker (Fitbit’s Charge HR) are used as data basis for the recognition of eustress and distress. The fact that such simple measurement methods are used and not any clinical devices such as ECGs is another aspect that sets this work apart from most other studies about stress detection.

In the conducted pre-study several HRV features and their suitability for the separation of eustress and distress were investigated. Although the data collected in the pre-study were not very precise the results suggested that it is possible to distinguish eustress from distress by means of multiple HRV features. Furthermore, based on the pre-study results two distinct heart rate patterns for eustress and distress were hypothesized, which should be verified in the later conducted field study.

Then a classification process was developed that utilizes a decision tree classifier and a SVM in order to classify before calculated segments by means of several HRV features, where the label of the segments (which is required to train the classification model) is partly computed and partly estimated. A computed stress model is used to determine the stress level (this means that if someone is stressed or not is computed by means of physiological characteristics), whereas the valence (if someone has positive or negative mood) is determined by self-assessment. The classification process then allows to classify unlabeled segments, which are presented in a web portal as feedback about the emotional state. For the web portal a visualization concept was developed that brings together all the collected and calculated data of the classification process. The developed classification process in conjunction with the implemented UI is then not only useful to provide feedback about the emotional state of students (or also other users), but rather the whole implemented infrastructure is a starting point to collect much more data and to analyze them.
In order to evaluate the developed classification process and the assumptions of the pre-study a field study was conducted, where 10 students of the LMU recorded their heart rate data during exams and answered a survey in order to label the data. Although it was not possible to create a separate classification model for each participant (due to an insufficient amount of data per participant) the classification with the data of all participants together resulted in a maximum recognition accuracy of over 75%. In a second field study even a maximum recognition accuracy of over 80% was achieved for a subset of the participants. But is this an appropriate accuracy for the purpose of this work? At a first glance one could argue that compared to other studies about stress detection 75% or 80% accuracy seems to be not very high (other studies achieved up to over 90% accuracy [19, 36, 8]). But there are two major differences between these studies and this work:

1. In other studies often more precise measurement methods are used. Either at least a full ECG signal is used or even multiple different physiological measurement parameters are used, such as the GSR, the skin temperature or the pupil diameter. But in order to obtain such measures much more intrusive equipment is required, which is not really applicable in everyday life. When looking at other studies, which also apply only simple HRV features and not a full ECG signal it can be seen that the accuracy also decreases significantly (for example Liu and Ulrich [19] achieved only 78.5% accuracy when using just HRV features).

2. In other studies often stress is considered as inherently negative and thus it is just distinguished between stress and rest and not between eustress and distress. Li et al. [18], who also considered the recognition of eustress and distress during the daily life achieved just about 71% accuracy for eustress as an urge for better performance and even just about 57% accuracy for eustress as a state of better mood.

So when considering these two points, the achieved classification performance in this work seems to be not so bad. In fact it is dependent on the intention of the work whether the recognition accuracy is good enough. Since the intention of this work is to provide a fast feedback about the emotional state of students during everyday life and therefore by means of non-invasive and even non-intrusive methods (at least as less intrusive as possible) a recognition accuracy of over 75% for the separation of eustress and distress seems to be acceptable. But nevertheless it would be reasonable to check in the future whether in practice the achieved accuracy is good enough or whether the users of the system feel not comfortable with the feedback given.

During the further analysis the proposed heart rate patterns of the pre-study and the correlation between the stress intensity and the stress type were investigated. Unfortunately the heart rate patterns proposed in the pre-study could not be confirmed in the field study. In fact several different patterns for both eustress and distress could be found. This result suggests that more complex hypotheses have to be developed in order to be able to describe eustress and distress more accurately. Also the assumption that eustress occurs at a rather moderate stress intensity, while distress occurs at a higher stress intensity could not be confirmed. It rather looks like there is no general correlation between the stress intensity and the stress type.
5.2 Limitations and future work

Finding clear criteria for differentiating eustress from distress

There is a significant discrepancy among scientists in the understanding of eustress. On the one hand some scientists consider eustress to be the ideal amount of stress that improves the performance but on the other hand there exists also the position that eustress is associated with positive feelings [18]. Consequently there is no general definition of eustress in the literature, which makes it difficult to label data precisely. In this work the latter definition was used, by labeling the data according to the self-assessed feeling of persons, which led to quite good classification results. But nevertheless there is still a lot of work to do.

Conducting further field studies

A limitation of the work reported about in this thesis is the small number of labeled data that was available in the pre-study and in the field study. As a consequence it was not possible to create classification models for single participants, which would be probably more appropriate. Thus, further field studies should be conducted in order to collect more labeled data. Especially the number of data per participant should be increased. On the one hand this is important to be able to better analyze the classification process and the feature characteristics of single participants and on the other hand the assumptions and findings of the field study could be generalized better.

Collecting more precise data

Besides to simply collect a larger quantity of data to investigate the physiological characteristics of eustress and distress another approach could be to increase the precision of the collected data (both the labels and the heart rate measures).

One problem in this work was that the participants often were not stressed during the field study and that the labels are based on self-assessments, which are not always absolutely accurate. And although in the exam field study it was tried to control the self-assessment with various further questions it could be argued if these questions were all answered carefully and if the interpretation of the answers was right. In order to get more precise labels a possibility could be to induce eustress and distress explicitly in a controlled environment (for example in a laboratory study). A possible scenario could be to stress the participants explicitly and once let them do something that gives them pleasure and then something that discontents them.

Another issue in this work was the inaccuracy of the heart rate measures. The used fitness trackers measure the blood volume pulse and derive therefrom an approximation of the heart rate about every 5 to 15 seconds. Furthermore the study of Marcel Heil [9] showed that the accuracy of the measurements is highly dependent on how users have been wearing their wristband. By using ECG records the heart rate measures could be obtained more precisely. Furthermore the whole ECG signal (which contains the full spectrum of the heart’s electrical activity and thus provides data that is richer than the timing of the heartbeats [19]) could be used to get a better insight into the physiological characteristics of eustress and distress.

Till now the drawback of this method was that recording the ECG is highly intrusive and
not applicable in everyday life. But there is a new technology called SenceBand\footnote{https://www.kickstarter.com/projects/455414429/sence%2Dthe%2Devolution%2Dof%2Dmindfulness%2Dand%2Dproductiv} that is said to have just the size of a fitness tracker but can record full ECG signals (figure 5.1 shows such a SenceBand). It is intended to come on the market late in 2017. If this technology is really as precise as advertised will have to be examined in due time.

![Figure 5.1: SenceBand.](image)

Using additional frequency domain analysis

When having more precise heart rate measures or even a full ECG signal further HRV features could be considered in the classification process. Since in this work only approximated heart rate measures from fitness trackers are available, only HRV features from the time domain were applied (features that are derived directly from RR-intervals). Therefore just a surrogate of the SVI was used, although it is normally a feature of the frequency domain. With more precise heart rate measures it would be possible to apply also HRV features from the frequency domain of the ECG signal. These features could then better reflect the low frequency and high frequency impacts of the ANS on the heart rate.

Improving the computed stress model

A crucial step in the classification process is the labeling of the data and when doing this the assessment of the stress level, which is used to determine if someone is stressed or not. So the more precise the computed stress level is, the more precise is the final label. Hence it would be important to further improve the computed stress model in order to determine more precisely if someone is stressed or not and to be able to better analyze the correlation between the stress intensity and the stress type. The current computed stress model is quite simple since it uses only one HRV feature (namely the RMSSD) to determine the stress level. It would be probably more appropriate to consider additional HRV features such as the meanRR or the pRR50, which are both considered to be reliable features for the stress detection in the literature \cite{4,32,31}. A further limitation of this work is that for all users the same stress threshold is applied. A personalization of the stress threshold would be more appropriate, since the HRV is a very individual value (inter alia dependent on the sex, the age and the physical condition).

Performing professional pattern analysis

In section 4.3 the feature combinations of eustress and distress segments have been analyzed by means of patterns in spider charts. However, in this work the patterns were just compared manually. Thus it is conceivable that not all patterns could be identified completely
5.2. LIMITATIONS AND FUTURE WORK

correctly. A professional computer-aided pattern analysis of the spider charts could help to find further distinctive patterns for eustress and distress.

Considering more complex heart rate patterns

The aforementioned pattern analysis of the spider charts revealed that there is not just one general heart rate pattern for eustress and one for distress, as supposed in the findings of the pre-study (see section 2.2.4). Thus there should be considered more complex heart rate patterns to be able to describe eustress and distress accurately. For the deduction of these more complex heart rate patterns a close look into the decision tree classifier may provide a starting point.

Investigating the impact of long eustress periods

In the first place eustress is considered to be something positive that is advantageous for the cognitive processing [24]. But an open question is whether the experience of long periods of eustress has still positive effects or whether there is a point when too much eustress turns negative. This could be for example investigated by inducing eustress explicitly over a longer time range and check periodically the performance of the study participants. It should then be examined, if there is some relation between the length of eustress experience and the performance of the participants. This finding should be also considered when giving persons feedback about their emotional state.

Evaluation of the user interface

In this work a UI was implemented that provides feedback about the emotional state of the users (when experienced the users rest, eustress or distress) and other interesting data (inter alia the heart rate and computed stress history of the users). In the future this UI should be evaluated, in order to assess if the feedback is presented in such a way that it is comprehensible for the users. The UI should then be adjusted according to the findings of the evaluation, so that the users have an optimal user experience.
List of Figures

1.1 Two-dimensional circumplex model of emotions ........................................ 5
1.2 Comparison of an ECG equipment and a fitness tracker ............................... 6

2.1 Sample of an ECG ...................................................................................... 12
2.2 RR-interval time series with low and high variability scores ....................... 16
2.3 A snippet of the heart rate list .................................................................... 18
2.4 The documented stress level and stress types for one day ............................ 18
2.5 A snippet of the output of the first preprocessing step .................................. 19
2.6 Structure of the labeled segments ................................................................ 20
2.7 Structure of the segments after the features were calculated ....................... 20
2.8 Overview of the analyzed data in WEKA .................................................... 21
2.9 Scatter plot visualization of the relationship between SVI\text{surrogate} and VS ... 23
2.10 Scatter plot visualization of the relationship between ApEn and VS ............... 24
2.11 Scatter plot visualization of the relationship between ApEn and SVI\text{surrogate} ... 24
2.12 Scatter plot visualization of the relationship between meanRR and VS ........... 25
2.13 Scatter plot visualization of the relationship between SDRR and pRR50 ............. 26
2.14 Scatter plot visualization of the relationship between CVRR and pRR50 ............ 26
2.15 Heart rate pattern for rest .......................................................................... 30
2.16 Heart rate pattern for eustress ................................................................... 31
2.17 Heart rate pattern for distress .................................................................... 31
2.18 A suppositious decision tree for the classification of eustress and distress ....... 34
2.19 A suppositious separation of eustress and distress training data by means of a maximum marginal hyperplane ...................................................... 35
2.20 A suppositious separation of eustress and distress training data by means of a nonlinear separating hypersurface ................................................. 36
2.21 A snippet of the heart rate list .................................................................... 37
2.22 Structure of the reported activities .............................................................. 37
2.23 Layout of the classification process ............................................................. 37
2.24 Labeling of confirmed and refused segments .............................................. 39
2.25 Layout of the developed system .................................................................. 40
2.26 Outline of the database structure ............................................................... 42
2.27 Mapping from the combination of computed stress level and labeled stress type to computed stress type ................................................................. 44
2.28 Sketch of the home screen ......................................................................... 45
2.29 Sketch of the detail view ............................................................................ 46
2.30 Sketch of the admin view .......................................................................... 47
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>A MongoDB document of the heartrate collection</td>
</tr>
<tr>
<td>3.2</td>
<td>Class diagram of the back-end</td>
</tr>
<tr>
<td>3.3</td>
<td>Structure of the JSON data object received by the server</td>
</tr>
<tr>
<td>3.4</td>
<td>Sequence diagram of the data synchronization procedure</td>
</tr>
<tr>
<td>3.5</td>
<td>Sequence diagram of the method invocations for the data preprocessing and classification</td>
</tr>
<tr>
<td>3.6</td>
<td>Sequence diagram of the preprocessing procedure</td>
</tr>
<tr>
<td>3.7</td>
<td>Sequence diagram of the decision tree classification procedure</td>
</tr>
<tr>
<td>3.8</td>
<td>Sequence diagram of the SVM classification procedure</td>
</tr>
<tr>
<td>3.9</td>
<td>Sequence diagram of the confirmation and refusal of classified segments</td>
</tr>
<tr>
<td>3.10</td>
<td>The menu bar</td>
</tr>
<tr>
<td>3.11</td>
<td>Extract of the home screen</td>
</tr>
<tr>
<td>3.12</td>
<td>Aggregated activities in the home screen</td>
</tr>
<tr>
<td>3.13</td>
<td>Extract of the detail view</td>
</tr>
<tr>
<td>3.14</td>
<td>Timeline with classified segments</td>
</tr>
<tr>
<td>3.15</td>
<td>Timeline with rated segments</td>
</tr>
<tr>
<td>3.16</td>
<td>Spider charts of the detail view</td>
</tr>
<tr>
<td>3.17</td>
<td>Selection of users in the admin view</td>
</tr>
<tr>
<td>3.18</td>
<td>First extract of the admin view</td>
</tr>
<tr>
<td>3.19</td>
<td>Second extract of the admin view</td>
</tr>
<tr>
<td>3.20</td>
<td>Data input via the web portal</td>
</tr>
<tr>
<td>4.1</td>
<td>Calculation of the score for the assessment matrix</td>
</tr>
<tr>
<td>4.2</td>
<td>Assessment matrix for the labeling of exams</td>
</tr>
<tr>
<td>4.3</td>
<td>Accuracy of the three most promising feature combinations</td>
</tr>
<tr>
<td>4.4</td>
<td>Comparison of the differences of HRV features between eustress and distress for each participant of the exam field study</td>
</tr>
<tr>
<td>4.5</td>
<td>Example of a spider chart</td>
</tr>
<tr>
<td>4.6</td>
<td>Number of identified patterns for eustress and distress</td>
</tr>
<tr>
<td>4.7</td>
<td>The identified eustress patterns</td>
</tr>
<tr>
<td>4.8</td>
<td>The identified distress patterns</td>
</tr>
<tr>
<td>4.9</td>
<td>The not assigned patterns</td>
</tr>
<tr>
<td>4.10</td>
<td>Scatter plot visualization of meanRR and RMSSD and their correlation with eustress and distress for all users of the exam field study</td>
</tr>
<tr>
<td>4.11</td>
<td>Comparison of the differences of HRV features between eustress and distress for each participant of the daily field study</td>
</tr>
<tr>
<td>4.12</td>
<td>Scatter plot visualization of meanRR and RMSSD and their correlation with eustress and distress for user 4, 5 and 6 of the daily field study</td>
</tr>
<tr>
<td>4.13</td>
<td>Scatter plot visualization of meanRR and RMSSD and their correlation with eustress and distress for user 1 and 3 of the daily field study</td>
</tr>
<tr>
<td>4.14</td>
<td>Classification accuracy of the daily field study</td>
</tr>
<tr>
<td>5.1</td>
<td>SenceBand</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison of the average values of HRV features during rest, eustress and distress conditions</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Comparison of the average values of HRV features (normalized to lie in the interval [0,1]) during eustress and distress conditions. Furthermore the standard deviation is given in brackets.</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Comparison of the classification performance of the decision tree classifier and the SVM</td>
<td>28</td>
</tr>
<tr>
<td>2.4</td>
<td>Confusion matrix for the two classification schemes decision tree and SVM</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Configuration parameters for the preprocessing procedure</td>
<td>62</td>
</tr>
<tr>
<td>3.2</td>
<td>Configuration parameters for the classification procedure</td>
<td>63</td>
</tr>
<tr>
<td>3.3</td>
<td>Configuration parameters for the daemon service</td>
<td>63</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of the classification performance of the decision tree classifier, the linear SVM and the nonlinear SVM</td>
<td>75</td>
</tr>
<tr>
<td>4.2</td>
<td>Confusion matrix for the three classification schemes decision tree, linear SVM and nonlinear SVM</td>
<td>75</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison of the average values of HRV features (normalized to the interval [0,1]) during eustress and distress conditions. Furthermore the standard deviation is given in brackets</td>
<td>77</td>
</tr>
</tbody>
</table>
LIST OF TABLES
List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMU</td>
<td>Ludwig-Maximilian University</td>
</tr>
<tr>
<td>ECG</td>
<td>electrocardiogram</td>
</tr>
<tr>
<td>HRV</td>
<td>heart rate variability</td>
</tr>
<tr>
<td>GSR</td>
<td>galvanic skin response</td>
</tr>
<tr>
<td>SVM</td>
<td>support vector machine</td>
</tr>
<tr>
<td>ANS</td>
<td>autonomic nervous system</td>
</tr>
<tr>
<td>SNS</td>
<td>sympathetic nervous system</td>
</tr>
<tr>
<td>PNS</td>
<td>parasympathetic nervous system</td>
</tr>
<tr>
<td>meanHR</td>
<td>mean heart rate</td>
</tr>
<tr>
<td>SDHR</td>
<td>standard deviation of the heart rate</td>
</tr>
<tr>
<td>meanRR</td>
<td>mean RR-interval</td>
</tr>
<tr>
<td>SDRR</td>
<td>standard deviation of the RR-intervals</td>
</tr>
<tr>
<td>CVRR</td>
<td>coefficient of variance of the RR-intervals</td>
</tr>
<tr>
<td>RMSSD</td>
<td>root mean square successive difference</td>
</tr>
<tr>
<td>pRR50</td>
<td>number of pairs of adjacent RR-intervals differing by more than 50 ms to all RR-intervals</td>
</tr>
<tr>
<td>ApEn</td>
<td>approximate entropy</td>
</tr>
<tr>
<td>VS</td>
<td>variability score</td>
</tr>
<tr>
<td>SVI</td>
<td>sympathovagal balance index</td>
</tr>
<tr>
<td>SVIsurrogate</td>
<td>surrogate for the sympathovagal balance index</td>
</tr>
<tr>
<td>bpm</td>
<td>beats per minute</td>
</tr>
</tbody>
</table>
Bibliography


