MACHINE LEARNING & DATA ANALYTICS

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Clustering of Stress Responder Types based on Diurnal Cortisol Profiles

Bachelor's Thesis in Medical Engineering

submitted
by

Linda Vorberg

born 16.08.1997 in Aschaffenburg

Written at

Machine Learning and Data Analytics Lab (CS 14)
Department of Computer Science
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Advisors: Robert Richer, M.Sc., Arne Küderle, M.Sc., Prof. Dr. Bjoern Eskofier
(Machine Learning and Data Analytics Lab, FAU Erlangen-Nürnberg),
Sarah Sturmbauer, M.Sc., Prof. Dr. Nicolas Rohleder
(Chair of Health Psychology, FAU Erlangen-Nürnberg)

Started: 03.06.2019

Finished: 04.11.2019

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien des Lehrstuhls für Bachelor- und Masterarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Erlangen, den 04.11.2019

Übersicht

Der zirkadiane Rhythmus hat einen großen Einfluss auf die Funktionen des menschlichen Körpers und richtet sich hauptsächlich nach der täglichen Erdrotation und der damit verbundenen Lichtintensität. Hormone, vor allem Glucocorticoide wie Cortisol, helfen, diesen Rhythmus an den ganzen Organismus weiterzugeben. Allerdings ist der zirkadiane Rhythmus sehr anfällig für Stress, vorwiegend chronischen und Lebenszeitstress. Beide werden durch verschiedene Umstände des täglichen Lebens hervorgerufen, wie beispielsweise Stress am Arbeitsplatz, Armut oder das Pflegen einer älteren oder kranken Person, aber auch durch prägende Ereignisse wie dem Verlust einer nahestehenden Person.
Aufgrund der Entwicklung verschiedener Pathologien, die mit der Unterbrechung des täglichen Rhythmus einhergehen, ist es wichtig, Veränderungen im täglichen Cortisolverlauf zu erkennen. Um auch zukünftigen Erkrankungen vorzubeugen, ist es notwendig, die Verbindung zwischen dem biologischen Rhythmus und psychologischen Variablen, die Aufschluss über Stressbelastung geben können, zu verstehen.
Da dies ein komplexes Phänomen mit einer vielschichtigen Menge an Daten darstellt wurden in dieser Arbeit Methoden des maschinellen Lernens genutzt, um verschiedene Cortisolrhythmen zu identifizieren und mit den psychologischen Variablen aus einem Fragebogen zu verknüpfen.
Mit der Hilfe von Clustering-Algorithmen war es möglich, aus einem Datensatz mit 107 Teilnehmern 3 verschiedene typische Cortisolprofile zu identifizieren – eine ‘normale’ Kurve, eine Kurve mit ‘abgeflachter Aufwachreaktion’ und ‘keine Aufwachreaktion’. Die Studie beinhaltete die Abgabe von Speichelproben verteilt über den Tag, um den Cortisolspiegel zu ermitteln, sowie die Beurteilung von psychologischen Variablen durch das Ausfüllen eines Fragebogens. Obwohl keine statistische Signifikanz beim Vergleich der psychologischen Variablen der gefundenen Gruppen festgestellt werden konnte, gab es dennoch Verbindungen zwischen den Verteilungen der Variablen und den zugehörigen Kurven. Besonders akute körperliche und psychische Krankheit beeinflussten die Variablenverteilung innerhalb der Gruppen stark, wobei jedoch im Verlauf des Tagesprofils kein Unterschied zu gesunden Probanden zu sehen war. Daher ist die Untersuchung des Zusammenhangs zwischen verschiedenen Erkrankungen und Stress ein möglicher Ansatzpunkt für weitere Forschung.
Eine Vorhersage des Cortisolverlaufes konnte allerdings nicht zuverlässig getroffen werden. Der Klassifizierungsansatz war vor allem durch fehlende Annotationen limitiert und ist daher ebenfalls ein weiterer Aspekt für zukünftige Arbeiten. Damit könnte eine Voraussage des Cortisolrhythmus tatsächlich ohne aufwändige Speichelproben möglich sein.

Abstract

One of the strongest regulators within our body is the circadian timing system that is mainly controlled by the daily rotation of the earth and the resulting light pattern. Hormones, especially glucocorticoids like cortisol, help to synchronise this pattern throughout the whole body. However, this diurnal rhythm of cortisol is very sensitive to stress, especially life and chronic stress. Both are caused by a combination of different circumstances of daily life, like work stress, poverty or conditions like looking after an elderly or sick person as well as striking life events, such as interpersonal loss or humiliation.

Due to the development of different pathologies that accompany the disruption of this circadian pattern, it is important to detect alterations in the cycle of cortisol. To prevent future disease, it is necessary to understand the linkages of the biological cycle with the psychological variables that reveal information about stress exposure.

As this interaction is a complex phenomenon with multiple layers of data, *Machine Learning* methods were applied in this work to identify different cortisol rhythms and link them to survey-assessed variables.

With the help of clustering algorithms it was possible to determine 3 different types of diurnal rhythms from a dataset with 107 subjects – 'normal' cycle, 'flattened Cortisol Awakening Response (CAR)' and 'no CAR'. The study included the collection of saliva samples throughout the day to observe the level of cortisol and the assessment of psychological variables with the help of a questionnaire. Although there was no statistical significance in the psychological variables between the groups that were found, linkages between the distribution of these variables and the corresponding curves could be identified. Especially the influence of acute physical and mental disease had a large impact on the distribution of variables within the groups whereas no alteration could be seen in the cycle of cortisol compared to disease-free subjects. Therefore, the investigation of the influence of disease on the conception of stress is a topic for further research.

A prediction of the type of cortisol cycle could not be made yet as missing labels in the data lead to a bad performance of the classification algorithms. In future work, this approach could be continued and optimized until a prediction of the cortisol rhythm, based on psychological variables without the assessment of saliva, is possible.

Contents

Chapter 1

Introduction

Stress affects the whole body – that is evident regarding the cascade of stress mediators that begins in our brain, goes down to our kidneys and then reaches nearly every cell [Fri09]. This response is triggered in situations of inescapable threat and helps the individual to survive. The most important hormone responsible for initiating a stress response is cortisol. Cortisol is a glucocorticoid produced by the adrenal cortex and is one of few hormones that are necessary for survival [Smy97]. It is a biomarker that is frequently used in psychobiological research as it responds to both acute stressors and chronic stress. Additionally, it is easy to assess, as it is possible and reliable to measure in saliva and does thus not require cumbersome and invasive blood serum analysis [Ada17].

While stress systems are vital for the individual in order to survive, they also have the potential to damage the organism [Roh19]. The exposure to chronic stress is most likely to result in permanent changes in biological, physiological and emotional responses and to increase the risk for diseases like cardiovascular diseases or even cancer [Coh07].

A circadian rhythm, as it is also found within other functions of the human body, such as heart rate, blood pressure, body temperature or blood glucose level [Sch14] plays a vital role in cortisol regulation. The cortisol level follows the circadian rhythm, i.e. changes over the course of one day which is called *Diurnal Cortisol Rhythm*. An alteration of this rhythm has been proposed to be a mediator between chronic stress and poor mental and physical health outcomes in past research [Ada17]. However, varying psychological factors and types of stress lead to other outcomes in the cortisol profile than what is proposed to be normal [Ada17]. Therefore it is important to identify the linkages between the diurnal rhythm of cortisol and its implications on mental and physical health. The knowledge gained from understanding these complex interactions can be used to improve the circadian regulation and prevent the development of chronic stress and

potential following psychopathology, e.g. by altering health behaviours, applying stress reduction methods or foster better sleep [Ada17].
As the interaction of stress mediators and psychological determinants constitutes quite a complex phenomenon, this problem has to be addressed with other methods than statistical analysis that are able to test selected hypotheses or analyze trends and correlations, which has mainly been done by psychologists so far [Ada17]. Therefore, new methods that can process many different levels of data and cover hidden linkages have to be explored. A promising approach that is not spread widely in psychology yet is the use of modern *Machine Learning (ML)* techniques which have the potential to uncover more complex relationships. Especially the automatic distinction of different classes (also referred to as classification) is one of the most common application fields of *ML*. The utilisation of classification algorithms in order to classify a diurnal profile into one of different, previously defined groups, but also the application of clustering algorithms to find populations with similar rhythms (without the need of defining classes beforehand) could be potential strategies. As the utilization of these methods in the field of stress research has by far not been investigated sufficiently, there is huge potential to uncover hidden relationships by further research.
For that reason, this work attempts to apply different machine learning methods to identify different stress responder types in data from a study collected by the *Chair of Health Psychology of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)*. The dataset consists of diurnal cortisol profiles, measured by acquiring salivary cortisol samples over the course of two consecutive days, and survey-assessed demographic and psychological variables. Furthermore, it is examined whether a prediction of the stress responder type based on the diurnal cortisol profile is possible.

The thesis is structured as followed: After an outline of the medical background in Chapter 2, Chapter 3 sums up the current state of research in the field of health psychology, with an emphasis on the application of machine learning methods to this topic. In Chapter 4 a detailed description of the methods is provided, including a description of the dataset used in this thesis, and the machine learning pipelines for both unsupervised and supervised methods (feature extraction, preprocessing, clustering/classification and evaluation). After the presentation of the obtained results in Chapter 5, the different approaches are compared and discussed in detail in Chapter 6. The thesis is concluded by Chapter 7, which sums up the key findings and presents an outlook with possible points of contact for future work.

Chapter 2

Medical Background

This chapter gives an overview of the medical background of the pathways that are responsible for stress hormones and health outcomes. The stress reaction is initiated in the brain and affects the whole body, whereby the two major components responsible for spreading the stress signalling cascade throughout the body are the *Hypothalamic-Pituitary-Adrenal Axis (HPA Axis)* and the *Sympathetic Nervous System (SNS)*.
States of stress can be "experiences in daily life, including daily hassles as well as major life events and abuse or trauma" [McE08] that can occur at different stages in life. Chronic stress and life stress have a considerable impact on the course of the diurnal cycle which is part of a complex circadian timing system [Sto01], whereas acute psychosocial stress causes a so-called *Fight-or-Flight Response* in the body, together with a release of different hormones to help the individual to cope effectively with environmental threats [Bre15].

2.1 Stress Signalling Pathways

In general, the release of cortisol is increased in states of heavy physical or mental strain and case of pain or dropping blood pressure. The purpose of this elevated level of cortisol is the provision of energy substrates by inhibiting the absorption of glucose in fat cells on the one hand and increasing the production of glucose in the liver to raise the glucose concentration in the blood plasma, on the other hand [Lan11].
The reaction to stress propagates through two different pathways and starts with the activation of neurons in the hypothalamus. The *SNS* then triggers the *Fight-or-Flight Response*, which includes increased tension of muscles, a defensive posture, the rise of blood pressure, sweating and an increased heart rate. At the same time, *Corticotropin Releasing Factor (CRF)*, a polypeptide that

also follows a circadian rhythm, is released from the hypothalamus, the initiator of the stress response. *CRF* triggers the distribution of *Adrenocorticotrophic Hormone (ACTH)*, a peptide hormone that regulates the release of corticoids, from the anterior pituitary glands which are located at the base of the brain [Lan11]. This ultimately leads to the release of *Cortisol* from the adrenal cortex. In turn, the increased level of cortisol inhibits the further production of *CRF* and *ACTH* in order to regulate the concentration of plasma cortisol as presented in Figure 2.1.
On the downside, *glucocorticoides* like cortisol delay cell growth and wound healing [Lan11]. For that reason, increased cortisol levels over a longer time period, e.g. due to chronic stress or a defect in cortisol regulation can negatively affect the human body. According to Cohen et al. chronic stress, which is stress that lasts for more than 3 months, heightens the risk of depression, cardiovascular diseases, diabetes, autoimmune diseases, upper respiratory infections, and poorer wound healing [Coh12].

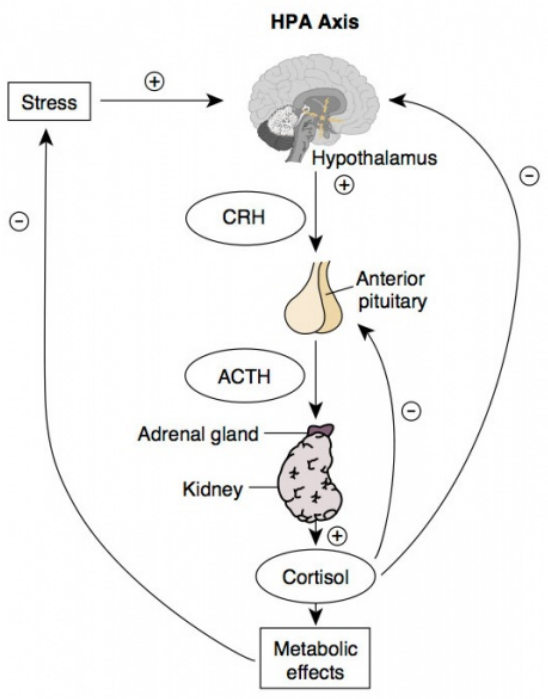


Figure 2.1: **HPA Axis** with the hormones involved, *negative feedback* of Cortisol suppressing the release of *CRH* and *ACTH* [Hil]

2.2 Diurnal Cortisol Rhythm

A fundamental property of living matter is the presence of a circadian rhythm, which can be influenced by internal and external factors, e.g. light. In humans, this rhythm spans approximately 24 hours and is synchronized with the daily rotation of the earth. The main regulator is the diurnal light pattern which is received by the human body through the eye and is then transmitted to *Suprachiasmatic Nuclei (SCN)* neurons in the hypothalamus. These neurons act as the central clock of the organism and are influenced by clock gene expression. In general, up to 10% of the human genome is under circadian control [Sch14].
Hormones, especially glucocorticoids play a role in the rhythmic synchronization of multiple peripheral clocks of the body. The *HPA Axis* receives input from the *SCN* which target the hypothalamus and initiate a cascade of hormones (*CRF, ACTH* and cortisol). The *HPA Axis* also cooperates with the *SNS*, that can modulate circadian rhythms locally. The whole process of entrainment, which means the alignment of the organisms' endogenous circadian rhythm to the external rhythm is shown in Figure 2.2 [Sch14].
Besides the hormonal cycle, the *SCN* also regulate the diurnal rhythm of body temperature, blood pressure and heart rate.
A disruption of this circadian rhythm is likely to happen when experiencing jetlag, sleep disturbances or doing shift-work [Ada17].

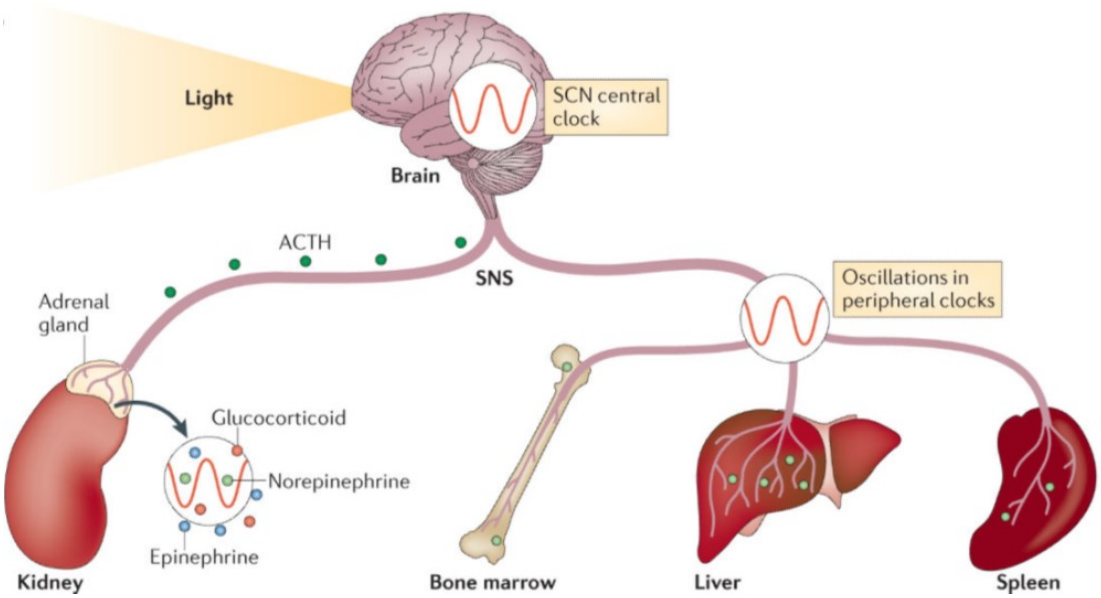


Figure 2.2: ***Entrainment* and synchronization**, *SCN* as the master clock of the organism, *HPA axis* setting a common phase [Sch14]

An example of a typical 24-hour cortisol rhythm can be seen in Figure 2.3. The concentration of cortisol rises slowly during the night in response to suprarenal activation by *ACTH*, which is controlled by *SCN*. The level increases by around 50-60% within 30 minutes after waking [Pru97] which is called the *Cortisol Awakening Response (CAR)*. It is supposed to start the mobilization of energy for the metabolism and prepare the individual for the upcoming day and possible stressful situations [Ste16]. It plays an important role in the course of the whole cycle as alterations of this response to awakening can be related to changes in *HPA Axis* activity. An enhanced *CAR* is e.g. found in people who report job stress and workoverload [Ste16].
After the peak level at 30 minutes after the time of waking, the concentration drops rapidly within the first hours of wakefulness, decreases more slowly until bedtime and reaches its minimum value in the first half of the night. [Ada17].

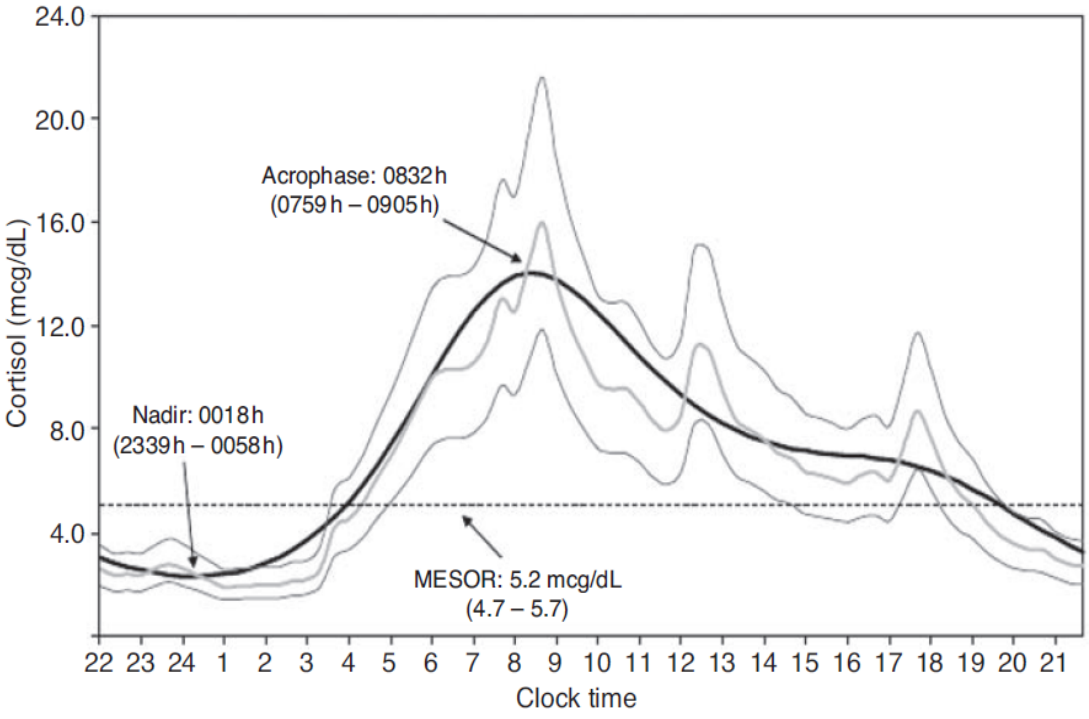


Figure 2.3: **24h Cortisol rhythm**, minimum value at 00:18, maximum value (CAR) at 08:32 [Cha10]

As the central circadian biology is influenced by clock gene expression in the *SCN*, genetic factors could affect *HPA Axis* functioning and the following diurnal cycle of cortisol. But there is also evidence that dysregulation of this rhythm is caused by psychosocial stress. This *stress-related circadian dysregulation* found in the altered cortisol cycle might give a hint on changes

within other circadian regulated functions of the body [Ada17]. As there are different variations in the cycle, e.g. normal, flat or inconsistent, between individuals due to genetic, developmental and psychosocial factors [Ada17], the diurnal cortisol rhythm is studied intensely as presented in the next chapter.

Chapter 3

Related Work

Since the *CAR* is stated to be a reliable indicator for changes in *HPA Axis* activity which can, for example, be related to chronic stress and its level is hardly impacted by age, smoking or the use of oral contraceptives, it has been investigated widely [Wüs00].
Adam et al. extended the research about altered *CAR* in stress-related diseases, especially *Major Depressive Disorder (MDD)*, simply known as depression. In their work, they state that the size of the *CAR* is a significant potential risk factor for future *MDD*, especially a heightened *CAR* may play a role [Ada10]. On the other hand, a decreased *CAR* is associated with self-focused rumination and less improvement of bad mood after distraction which constitutes vulnerability for future depression [Kue07]. Another stress-related disease is burnout which has been investigated by several groups and lead to contradictory findings. DeVente et al. and Grossi et al. obtained results about an enhanced *CAR* among participants with burnout compared to healthy controls [Gro05, De 03], whereas Chida et al. reported in their meta-analysis that burnout is characterized by a reduced *CAR*. They explain the contrary findings by sleep disturbances that can occur with burnout, so that the cortisol level might arouse near consciousness and before the person defines to be awake [Chi09]. Nevertheless, cortisol reactivity and basal levels were found out to be similar for healthy subjects and patients with burnout [De 03].
Posttraumatic Stress Disorder (PTSD), which is a mental disorder and is characterized by the re-experiencing of a traumatic event and emotional numbing was studied in connection with the *CAR* by Wessa et al. [Wes06]. They compared cortisol values of 48 trauma-exposed subjects (29 with and 19 without PTSD) and 15 non-exposed controls. *PTSD* patients showed a significantly lower cortisol increase after waking than the other groups, but baseline cortisol levels directly after waking did not differ.
Further research has also been done in the field of work-related stress. Steptoe et al. found out

that free cortisol levels in the morning were significantly higher in school teachers with high job strain, defined as a combination of high job demand and low job control, compared to teachers with low job strain [Ste00]. Similar findings were reported by Schulz et al. who compared cortisol levels of subjects who were chronically stressed due to work overload and non-stressed ones. They also stated a significantly larger increase in cortisol after waking for the chronically stressed group [Sch98]. Furthermore, the effect of work overload plays a role in the difference in the cortisol rhythm between workdays and weekends. There is a clear difference in the *CAR* between workday and weekend, that is associated with workload and worry [Sch04, Tho06, Wil05]. Schlotz et al. performed a study with 219 participants. Subjects who reported higher work overload and worrying showed a stronger increase and enhanced cortisol levels after waking on weekdays. However, this effect could not be observed on weekends [Sch04]. Similar findings around the time of waking were made for 'early-risers' that show a higher *CAR* than people who get up later [Kud06, Edw01]. Besides, shorter sleep is associated with a heightened *CAR* [Kum09].
Besides work-related stress, other environmental factors like *Socioeconomic Status (SES)*, a measure for the social standing of an individual in relation to others, have been shown to impact *HPA Axis* activity. Previous work has already associated *SES* with increased psychosocial stress. Therefore, Desantis et al. investigated the impact of low *SES* during different developmental periods on the diurnal cortisol profile [Des15]. The lowest *CAR*, highest bedtime levels, lowest total cortisol levels during the day were found amongst subjects with low *SES* from infancy through early adulthood. This goes along with the findings of Cohen et al. who also reported flatter diurnal rhythms with less of a decline in the evening for subjects with low *SES* [Coh06]. In contrast, Wright et al. report that a larger *CAR* is related to low *SES* [Wri05]. Another aspect of circumstances of daily life is the caregiving for an elderly or sick relative which comes along with increased stress and burden. These circumstances are associated with an altered *CAR* as well [De 05, Buc04].
Within the last years, *ML* methods have emerged in research topics in the field of Psychology and have been used to explore uncovered relationships. They constitute a new alternative to conventional statistical methods. For example, Smets et al. compared six different *ML* methods for distinguishing psychological stress from rest periods. The stress detection was based on physiological signals like *Electrocardiogram (ECG)* or temperature during a stress test. The stress and rest periods were best classified by Bayesian networks (84.6 % accuracy) and generalized *Support Vector Machines (SVMs)* (82.7 % accuracy) [Sme16].
A lot of research has been done by Galatzer-Levy et al., for example in their work about the 'Relevance of Machine Learning to the Study of Stress Pathology, Recovery, and Resilience' they

present different *ML* methods that could help to identify risk or classify individuals as healthy or ill. Besides, a lot of their work is about *PTSD*. They tried to predict *PTSD* after trauma in a study with 957 trauma survivors with the use of *SVMs* and other *ML* methods, such as *Random Forests*. The information was collected directly in the emergency room and 10 days past the event. The data set consisted of the type of traumatic event and injury, as well as psychometric assessment, and social support [GL14]. An *Area Under the Receiver Operating Characteristics curve (AUC)* of 0.82 was reached, which could even be improved to an AUC value of 0.93 in a later work from 2017 with the use of *SVMs* and graph induction algorithms [GL17]. A group from the Machine Learning and Data Analytics Lab in Erlangen with Abel et al. investigated whether it is possible to classify stress responder groups based on biological markers. Acute stress was induced by the *Trier Social Stress Test (TSST)* on two consecutive days. Cortisol and *Interleukin-6 (IL-6)* samples were collected before and after the stressor and used to classify subjects into different stress responder groups. The best results were achieved with linear *SVMs* for both cortisol (92.2 % $\pm$ 9.7 % for four different classes) and *IL-6* (91.2 % $\pm$ 6.3% for three different classes) [Abe19]. Nevertheless, the relation of stress, especially chronic and life stress, and the diurnal cortisol rhythm has not been investigated with *ML* methods yet and therefore this thesis advances research in this field.

Chapter 4

Methods

In this chapter, the study in which the data for this work was acquired is presented first. Second, the ML methods used in this thesis are explained. In general, two approaches were performed, unsupervised and supervised machine learning. Both are realized including preprocessing, feature extraction (which is the same for both models), selection of algorithms and evaluation.

4.1 Data Acquisition

The data acquisition was performed by researchers from the *Chair of Health Psychology* of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). The study took place over the course of almost 2 years, June 2017 – July 2019 and was carried out with 171 subjects, 78 men and 93 women. The intention was to explore the determinants of life stress. Therefore, all participants had to come to the laboratory to register at the beginning. Then, saliva samples had to be collected at two consecutive days at home to assess the cycle of cortisol. Due to missing data, some subjects had to be excluded for further evaluation. Finally, this leads to 107 subjects, 47 males and 60 females. The mean age was 29.67 years as shown in Table 4.1.

Health Data

All participants gave information about their size, weight and diseases. Other relevant health data that were acquired are information about smoking habits, medication, and medical treatment.

Table 4.1: Demographic and health information

age [years]	29.67 ± 10.25
BMI $[\mathrm{kg/m^2}]$	24.12 ± 5.26
Smoker [yes/no]	[20/87]

Psychological Variables

During the study, all subjects filled out several questionnaires about recent stressors, anxiety, depression, rumination but also about stressors at different stages in life, for example during childhood or adolescence as well as questions about fundamentals such as education, housing or work. Afterwards, scores and sums were determined from the different questionnaires, e.g. *PSS (Perceived Stress Scale)*, *STADI (State-Trait-fear-Depression-Inventory)* or *FSOZU (Questionnaire about Social Support)*. Questions about life stress were assessed via the *STRAIN Index (Stress and Adversity Inventory)*, an online system for systematically assessing lifetime stress exposure [Sla18]. Essential variables of the *STRAIN* are the *total count of stressors* (further referred to as *StressCT*) and *events of early adversity* (further referred to as *EATotCT*). The total count of stressors includes all types of stressors such as acute life events, e.g. getting fired from a job, chronic difficulties, like caregiving for an older person and stressors in different life domains [Sla19]. Events of early adversity are traumatic experiences that occur before the age of 18, e.g. the divorce of the parents. The most important variables which also include variables from the *STRAIN* are shown in Table 4.2. A full list of all variables assessed can be found in Appendix A.1.

Table 4.2: **Most important items from the questionnaire**

PHQ	Physical health complaints/ symptoms
K6	Mental health complaints/ symptoms
StressCT	Count of cumulative stressors across the life span
StressTH	Severity of cumulative stressors across the life span
EATotCT	Count of stressors during childhood (before the age of 18)
EATotTH	Severity of stressors during childhood
t1_ADSL_sum	Depressive symptoms within the last weeks
t1_PSS_sum	Perceived stress within in the last 4 weeks
t1_STADI_S	Current state of fear and depression
t1_STADI_T	Tendency to states of fear and depression
t1_ERQ_suppress	Strategy suppression
t1_ERQ_reapp	Strategy reappraisal
t1_FSOZU_sum	Perceived social support

Biological Variables

The subjects were told to take the first sample directly after waking up. The next ones were then supposed to be taken 30 minutes after, then 4 hours, 9 hours and 13 hours after the time of waking, as presented in Figure 4.1. In addition, the subjects were instructed to write down the exact times when the saliva samples were taken. These samples were analysed for cortisol and alpha-amylase in a laboratory. Additionally, *IL-6*, a marker for inflammation, was determined from blood in the laboratory at the beginning. An example of diurnal cortisol rhythm that was picked randomly from the dataset is shown in Figure 4.2

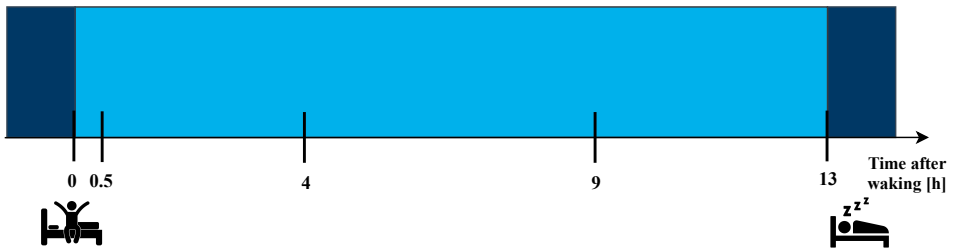


Figure 4.1: **Saliva collection times**

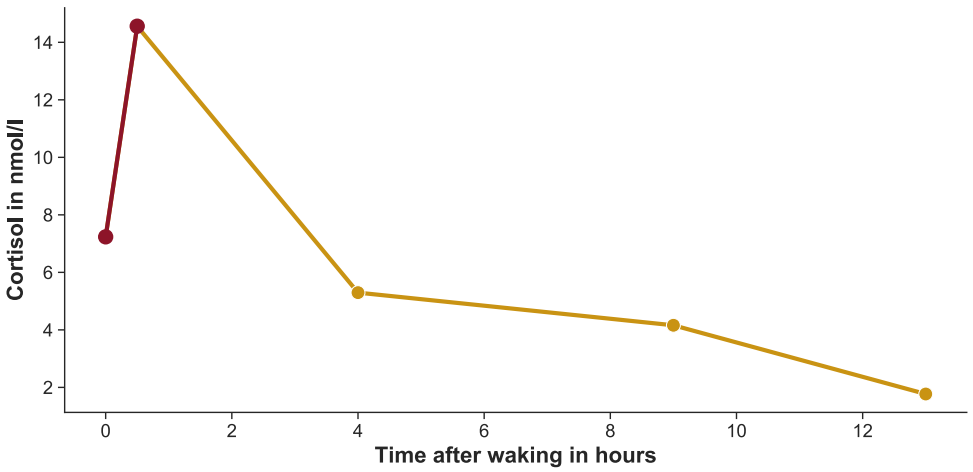


Figure 4.2: **Example of a Diurnal Cortisol Rhythm** from the dataset, *CAR* marked in dark red

4.2 Feature Extraction

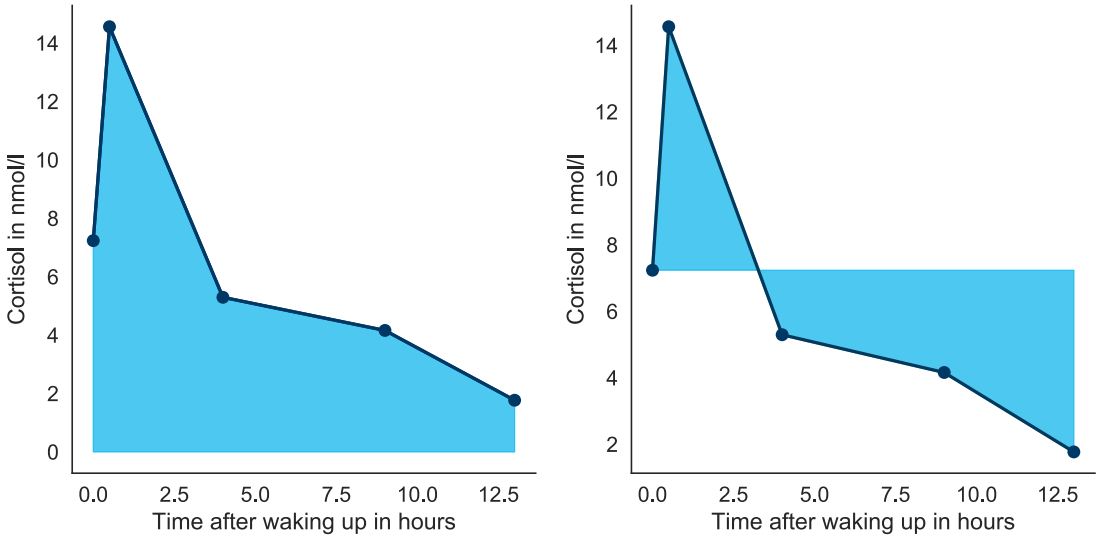
In order to apply machine learning techniques to a given problem a typical approach for facilitating the subsequent learning process is to extract relevant and non-redundant features from the dataset. These features were obtained from the samples of the diurnal cortisol rhythm. For every subject, the *Area Under the Curve (AUC)* was computed. According to Pruessner et al. there are two possible definitions for the term "Area under the Curve" [Pru03]:

- The *Area under the curve with respect to ground*(AUC_g) (Eq: 4.1)
- The *Area under the curve with respect to increase*(AUC_i) (Eq: 4.2)

Area under the curve with respect to ground

The *Area under the curve with respect to ground* represents the cortisol output during the day, more specific from the point of waking, when taking the first saliva sample until the point of time of the last cortisol sample. The AUC_g (4.1) is computed based on the summation of trapezoids

$$AUC_g = \sum_{i=1}^{n-1} \frac{(m_i + m_{(i+1)}) \cdot t_i}{2} \tag{4.1}$$



(a) with respect to ground (AUC_g) (b) with respect to increase (AUC_i)

Figure 4.3: **Area under the curve** of one cortisol rhythm. Measurements m_1 to m_5 at time points t_1 to t_5 are denoted as blue dots.

with t_i denoting the time intervals between the measurements m_i and $m_{(i+1)}$. In this case t and m range from 1 to 5, which is also visible in Figure 4.3a where the markers represent the measurements.

Area under the curve with respect to increase

The *Area under the curve with respect to increase* emphasizes changes over time. It is more a measure of decrease than an area as its values are mainly negative due to the decrease of the cortisol level over the day. An example can be seen in Figure 4.3b. The AUC_i is computed by subtracting the baseline (i.e. the value of the first sample) from all measurement points [Pru03].

$$AUC_i = AUC_g - m_1 \cdot \sum_{i=1}^{n-1} t_i = \sum_{i=1}^{n-1} \left(\frac{(m_{(i+1)} + m_i)}{2} - m_1 \right) \cdot t_i \tag{4.2}$$

Slope

Another common feature for characterizing the diurnal cortisol rhythm is the *slope*, which describes the degree of change of the indicator [Ada17], in this case, the level of salivary cortisol. The slope is computed by building the difference quotient between two measurement points m_i and m_k.

$$s = \frac{m_n - m_0}{t_n - t_0} \tag{4.3}$$

with m_n and m_0 representing the last and the first measurement. In this work, two different variants were used as features – the slope over the whole day and the slope of the *CAR*, which is the incline between the first two measurements. As some of the subjects didn't take the first sample exactly 30 minutes after waking, but e.g. after 25 minutes, the value of this second measurement was interpolated to exactly 30 minutes via linear interpolation, because even a few minutes may impact this value considerably [Clo04].

As the Cortisol Awakening Response (*CAR*) is a very important part of the whole diurnal cycle, *CAR*-specific features were computed for the characterisation of different cortisol awakening responses:

- The *Area under the Cortisol Awakening Response with respect to ground* (AUC_{CAR_g})
- The *Area under the Cortisol Awakening Response with respect to increase* (AUC_{CAR_i})

Both were computed with the trapezoidal rule but only for the first three measurements.

All features were computed per subject and separately for the first and the second day. Another approach was to compute the mean over both days and the difference between the two days. As the use of both feature extraction variants is ambiguous, either the features of day 1 and day 2 separately or the combination of mean and difference were used.
Since the feature ranges are highly different, they were normalized using z-score normalization which subtracts the mean value of the feature distribution from each sample and divides it by the standard deviation (see Equation 4.4). This is necessary for some clustering and classification algorithms to work properly.

$$z = \frac{(x - \mu)}{\sigma} \tag{4.4}$$

After z-score normalization, all feature distributions have a mean value of 0 and a standard deviation of 1.

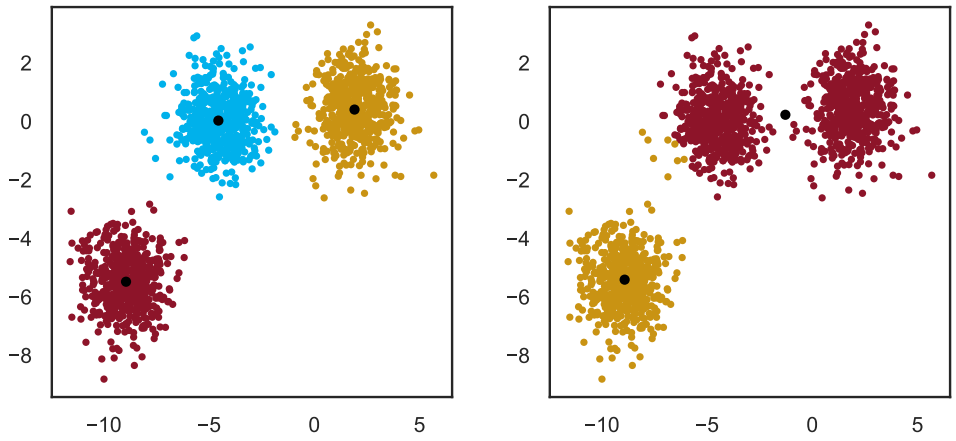
4.3 Unsupervised Learning

Since in previous studies mainly statistical methods were used to evaluate a single hypothesis, this work attempts to utilize more complex *ML* methods to analyse more intricate relations. First, unsupervised methods were used in order to find meaningful clusters within the study population. The goal in unsupervised learning is to find groups of similar examples when no class membership is known [Bis07]. In a first step, different clustering algorithms are applied to features extracted from the cortisol samples in order to find different groups of cortisol patterns. Afterwards, it was analyzed whether different cortisol profiles can predict differences in psychological variables between these groups.

4.3.1 Clustering Algorithms

KMeans

The *KMeans algorithm* attempts to find groups in a set of unlabelled data. The number of clusters k that will be found has to be determined before. The *KMeans* algorithm initializes the k clusters by setting k random points in the data set to form the initial cluster centres. The points closest to each centre are identified and assigned to the corresponding cluster. Next, the mean values of each cluster are determined and set as new cluster centres. These two steps are repeated until convergence [Has17]. One iteration of the KMeans algorithm can be described by the



(a) $k = 3$ leads to desired clustering result

(b) $k = 2$ still converges, but leads to adverse outcome

Figure 4.4: Randomly generated blobs clustered with *KMeans* and different number of clusters

minimization of the objective function J in Equation 4.5, which represents the sum of squared Euclidean distances of each data point to its cluster center.

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|x_n - \mu_k\|^2$$
$$\text{with: } r_{nk} = \begin{cases} 1 & if x_n \in k \\ 0 & else \end{cases} \tag{4.5}$$

It represents the sum of the square of the distances of each data point to its cluster centre μ_k. For every observation x_n a set of binary indicator variables $r_{nk} \in \{0, 1\}$, with $k = 1, ...K$ describes which cluster K the point x_n is assigned to in order to only sum up the distances between samples and the cluster centers they belong to, respectively [Bis07]. In this work, the set of features was used as input data in different variations, day 1 and day 2 separately and the combination of mean and difference between both days. Since the number of different clusters in the dataset is unknown, different values of k, k $\in [2, 4]$, were used for the unsupervised learning process. As the example in Figure 4.4 yields, the number of clusters that should be found by the clustering algorithm is crucial for the result of the clustering process. An inadequate number of clusters still leads to the algorithm to converge, but the results might be misleading.

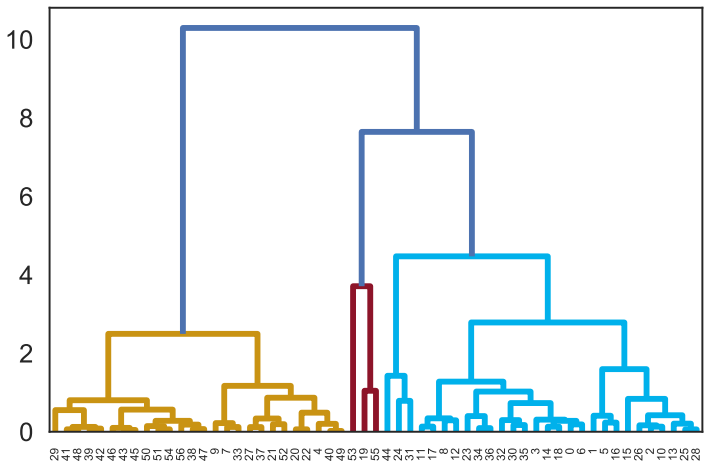


Figure 4.5: **Dendrogram of Agglomerative Clustering** with *linkage* = *ward*, vertical height of the dendrogram shows the *Euclidean distances* between the points of the horizontal axis

Hierarchical Clustering

In comparison to the *KMeans* algorithm *Hierarchical Clustering* does not need a pre-defined number of clusters that should be found, but the specification about the dissimilarity between the groups has to be set before. The hierarchy of clusters is represented as a tree with the root forming a unique cluster which contains all the other clusters (see Figure 4.5).
In this work a bottom-up strategy of hierarchical clustering, called *Agglomerative Clustering* was used. Hereby, a pair of selected clusters is merged into a single one [Has17]. Therefore, a measure of dissimilarity has to be defined before in order to decide which two clusters are closest. The distance between the observations used in this case is called *ward* and minimizes the variance of the clusters that are going to be merged. The dissimilarity between clusters A and B is expressed by the formula

$$d(A,B) = \sqrt{\frac{2 * |A||B|}{|A| + |B|}} * \|\overrightarrow{c_A} - \overrightarrow{c_B}\|_2 \tag{4.6}$$

with A as centroid of cluster $\overrightarrow{c_A}$ and B as centroid of cluster $\overrightarrow{c_B}$ respectively. As shown in Figure 4.5 the points closest to each other are determined first. This is e.g. represented by the connection of point 25 and 28 on the very right bottom which then form a cluster. The absolute euclidean distance is displayed on the vertical axis. The process of finding the closest point or cluster is performed until all points are joint in a single cluster, represented by the root of the tree.

4.3.2 Evaluation

The result of the clustering process can be evaluated by the validity measure *Silhouette Coefficient*. The *silhouette* $S(o)$ of an object o from cluster A is defined as the difference between the distance of object o to the *closest* cluster B and the distance of object o to cluster A, weighted by the maximum of both distances (see Equation 4.7).

$$S(o) = \frac{dist(B, o) - dist(A, o)}{\max\{dist(B, o), dist(A, o)\}} \tag{4.7}$$

This leads to silhouette values between -1 and +1, with a value of:

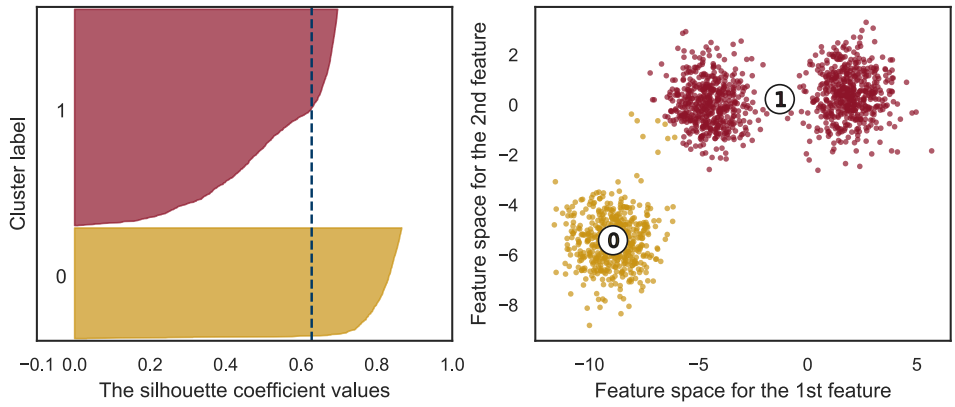
- +1 indicating that o is assigned to the correct cluster
- 0 indicating that o lies equally far away from both clusters
- -1 indicating that o is possibly assigned to the wrong cluster (i.e. closer to cluster B than to cluster A)

From the silhouette, the *silhouette coefficient* s, which is the arithmetric mean of all silhouettes of a cluster C, i.e. the silhouettes of all objects o belonging to cluster C, can be computed for either a single cluster or the whole dataset (Equation 4.8) [Rou87].

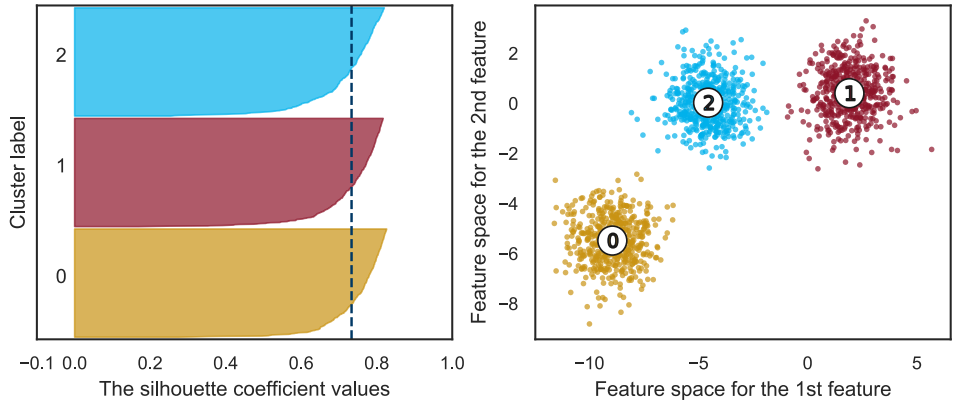
$$s_C = \frac{1}{n_C} \sum_{o \in C} s(o) \tag{4.8}$$

For visualization, a silhouette plot, which plots the silhouette of each object o is used, which helps to find the correct number of clusters as well. The silhouette plot in Figure 4.6 shows that the mean coefficient over all samples is higher for 3 clusters than for 2, which implies that the data is better represented by 3 clusters than by 2. The mean silhouette coefficient serves as a measure for the quality of the clustering process.

The investigation of the differences in the psychological variables between the groups is done by applying a *Analysis of Covariance (ANCOVA)*. The psychological variables, e.g. *Total Count of Stressors* serve as dependent variable and disease and cortisone medication as covariates. In general, an *ANCOVA* determines the correlation between the covariates and the dependent variable and removes the variance, which is associated with the covariates from the dependent variable scores [Rut01].
The results of the *ANCOVA* consist of *p-values* and *F-values*. A *p-value* below the level of



(a) Silhouette plot for 2 clusters and visualization of clustered data



(b) Silhouette plot for 3 clusters and visualization of clustered data

Figure 4.6: **Silhouette plots** for randomly generated blobs, *mean silhouette coefficient* presented by the vertical dashed line

significance (α), which is usually 5%, allows to reject the null hypothesis. In this case, it means that we can assume that the subjects belong to different groups – the groups that were found by the clustering algorithm. But, as the *ANCOVA* was performed with $m = 13$ different variables, a multiple testing correction had to be applied. This was done to prevent the incorrect rejection of the null hypothesis, due to the occurrence of a rare event which is more likely to happen with multiple hypotheses. The chosen method was *Bonferroni* correction, so the null hypothesis can be rejected for every

$$p \leq \frac{\alpha}{m} = \frac{5\%}{13} = 0.00385 \tag{4.9}$$

4.4 Supervised Learning

Another machine learning approach that was utilized in this thesis is supervised *ML*. Different *classifiers* were applied to predict the group of stress responder a person belongs to, based on their cortisol profile.

4.4.1 Preprocessing

Data Labeling

For the classification approach, each sample needs to be assigned to a class in advance (also referred to as *labelling*). During the classification process, the algorithm then attempts to learn the class membership based on the provided training data. The performance of the resulting model is then usually evaluated on a separate dataset. As the group that a person belongs to, which would serve as a label, is not present in the dataset, a bootstrapping approach to generate labels based on survey-assessed variables, such as the total count of stressors a person had experienced, had to be performed. In this work, two different approaches were realized to assign labels to specific subjects.
One way is to predict the label by clustering the variable into k different groups. This is done by performing *KMeans* on a single variable from the questionnaire. For $k = 2$, this would for example split the population into *High Stress* and *Low Stress* groups. If the clustering of one variable yielded groups that consisted of less than 3 subjects, this variable was rejected for further classification.
Another attempt for grouping the data is to divide them by performing percentile-based splits. In the case of two groups data is split by the median – the 50th percentile – into a *Low Responder* (below median) and a *High Responder* (above median) group. These two approaches were performed with the most important survey-assessed variables from Table 4.2, which were determined by psychological experts. This lead to different labels which were then used separately for the classification process.

Oversampling

To overcome the problem of an imbalanced class distribution, a synthetic minority oversampling technique is used to balance the classes after the labelling process. Otherwise, the classifier mainly trains on the majority class which leads to poor performance. The algorithm chosen is *SMOTE* from the `imbalanced-learn` library [Wan06]. This oversampling method synthetically

```
    def cluster_var(self, variable, n_clusters):
        data = self.df_full[variable]
        x = pd.DataFrame(data, index=self.
           dict_df_wide['cort'].index)
        ### remove subjects with no input for this variable
        x = x.replace(0, np.nan)
        x = x.dropna()
        ### clustering with KMeans
        k = KMeans(n_clusters= n_clusters).fit_predict(x)
        res = pd.DataFrame(x)
        res = res.rename(columns={0: variable})

        k = pd.DataFrame(data=k, index = x.index)
        k = k.rename(columns={0: 'group'})
        ### DataFrame 'r' with input variable and cluster group
        r = pd.concat([res, k], axis=1)

        return r
```

Listing 4.1: Generation of Labels by Clustering

creates additional samples for the minority class. By searching for the k nearest neighbours of x_i, the nearest neighbour x' is chosen and via

$$x - new = x_i + (x' - x - i) \times \delta$$

a new sample x_{new} is generated and contributes to the minority class. δ is a random number $\in [0, 1]$ [Zhe15]. In this work, a value of $k = 3$ was chosen for the oversampling algorithm.

Feature Selection

One common problem when training classification models with a high number of features is that it tends to overfit, i.e. that it learns the structures of the training data well, but performs weakly on the testing data [Caw10]. Additionally, an increase in features leads to a strong increase in the complexity of the classification problem. One approach to counteract these issues and to reduce the number of features used for classification is performing a feature selection beforehand. This is also important as single features may carry good classification information when treated separately but not together in a combined feature vector [The09]. The method chosen for feature selection in this work was *Select K Best* [Ped11] with the score function `f_classif` which

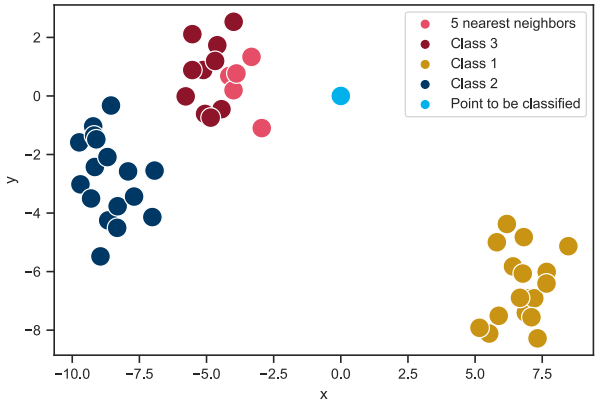


Figure 4.7: **Nearest Neighbour** Classification, here: $k = 5$

selects the k best features with the best *Analysis of Variance (ANOVA) F-value*. The *F-value* is the variance between groups divided by the variance within groups. A large *F-value* indicates that the difference of the means between the groups is higher than the variation of the individual observations in the group [Kim14]. The selection algorithm was performed with 1 to 12 features and the classifier was trained with the new input of selected features.

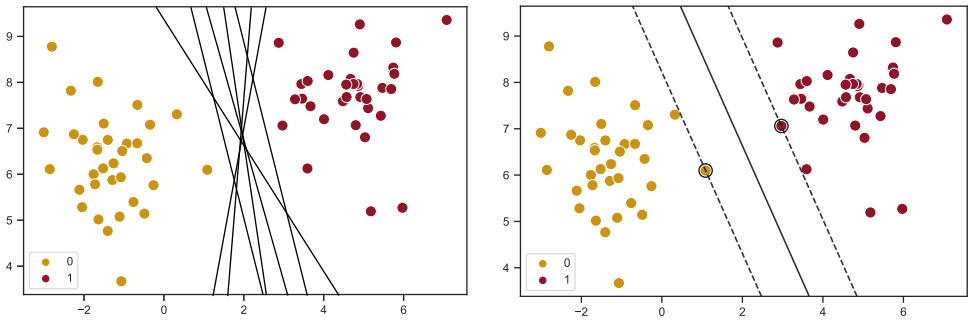
4.4.2 Classification Algorithms

K Nearest Neighbours

The first classifier used is the *K Nearest Neigbbor (kNN)* classifier [Ped11]. For a query point x_0, the k nearest training points are found and x_0 is assigned to the class which forms the majority of the k nearest neighbours of x_0 [Has17]. For the distance, the Euclidean distance $d(i) = \|x_{(i)} - x_0\|$ is used. In Figure 4.7, the light blue point has to be assigned to a group. Firstly, the distances to the k nearest neighbours of the groups around are determined and the point is assigned to class 3 based on the majority vote (5 out of 5 neighbours belong to class 3, 0 out of 5 neighbours belong to class 1 and 2). The only parameter that needs to be tuned for the *kNN* classifier is the number of neighbours k.

Support Vector Machines

When trying to separate two classes, it seems very easy to find a decision boundary when the classes are clearly separable (Figure 4.8). Thereby, an infinite number of possible linear decision boundaries can be found, but the challenge is to find the optimal one that also reaches good results with new, unseen data. *Support Vector Machines (SVM's)* help to solve this problem as they train a



(a) Possible decision boundaries for the separation of 2 classes

(b) Support Vector Machines Classifier, decision boundary as solid line, margin as dashed line

Figure 4.8: Classification of 2 clearly separable classes

decision boundary that maximizes the margin, leading to an optimal hyperplane that separates the data [Gun98]. An example of two classes, with some of the possible decision boundaries, is shown in Figure 4.8a. In Figure 4.8b the optimal boundary is presented by the solid line in the middle. The margin is the distance between it and the nearest points of each class, which is represented by the dashed line. However, by maximizing the margin, the rate of misclassification is increasing as well. The parameter C of the *Support Vector Classifier (SVC)* algorithm, the so-called *penalty parameter*, defines how large or small the margin should be [Ped11]. A large value of C will hardly allow any misclassification and leads to an overfitting boundary. A small value for C increases the margin and leads to a smoother boundary. An optimal value for C as well as a kernel function have to be chosen. The kernel function defines the inner product in the transformed space, because as the data set used may not be linearly separable anymore, a linear hyperplane can be found when being mapped into a higher dimensional space [Gun98]. Possible kernels are, *linear, polynomial kernel* $(1 + \langle x, x' \rangle)^d$ and a *radial basis function kernel* $(exp(-\gamma \|x - x'\|^2))$ [Has17]. The selection of the best parameters results in the *optimal separating hyperplane* which is not just a line anymore for a more complex problem with higher dimensional data.

Random Forest

The last classifier used is the *Random Forest* classifier. A *Random Forest* is composed of many individual decision trees, where one tree consists of data that is continuously split up according to a certain parameter. For each possible outcome (i.e. class) one branch is created [Utg97]. The decision trees form an ensemble of trees where every tree separately predicts a class. The final class membership of one sample is determined by majority vote of all trees. To set up such a

Table 4.3: **Parameters to be optimized with Randomized Search for *Random Forest Classifier***

Parameter	Description	Value range
max_depth	maximum depth of the tree	[80, 90, 100, 110, 200]
min_samples_leaf	minimum number of samples required to be at a leaf node	[3, 4, 5]
min_samples_split	minimum number of samples required to split a node	[8, 10, 12]
n_estimators	number of estimators (i.e. number of trees in the forest)	[100, 200, 300, 1000]

forest of trees bootstrap aggregating is used in order to create a large collection of de-correlated trees [Has17]. Each new training set is drawn from the original training set with replacement and a tree is grown on it using random feature selection [Bre01]. For growing the tree the best split point has to be picked in each step to split the node into two daughter nodes until the minimum node size is reached. At each node, a predictor generates a splitting variable which splits up the training set into different subsets [The09]. The splitting criterion used in this case is *Gini Impurity* which measures how often a randomly chosen element would be labelled incorrectly. The parameters to be optimized for the *Random Forest Classifier* are displayed in Table 4.3. The value ranges were adapted to the classifier's performance for further evaluation.

4.4.3 Evaluation

Cross Validation

In order to evaluate the classification performance, the data was split up into a training and a complimentary test set. This has to be done to see how the model performs on data that it has not seen before. The training set is used to train the model which is then tested on the unseen data from the test set. The cross-validation method chosen here is *Stratified K Fold* [Ped11] where the data is split up k times. This leads to an overall accuracy which is the average classification accuracy of the k folds. Hereby, the stratified partitioning leads to the same proportion of classes in each subset [Mul00]. The cross-validation split is also presented in Figure 4.9 where the data is split up via *Stratified K Fold*. The training set is then again used for cross-validation in the next step.

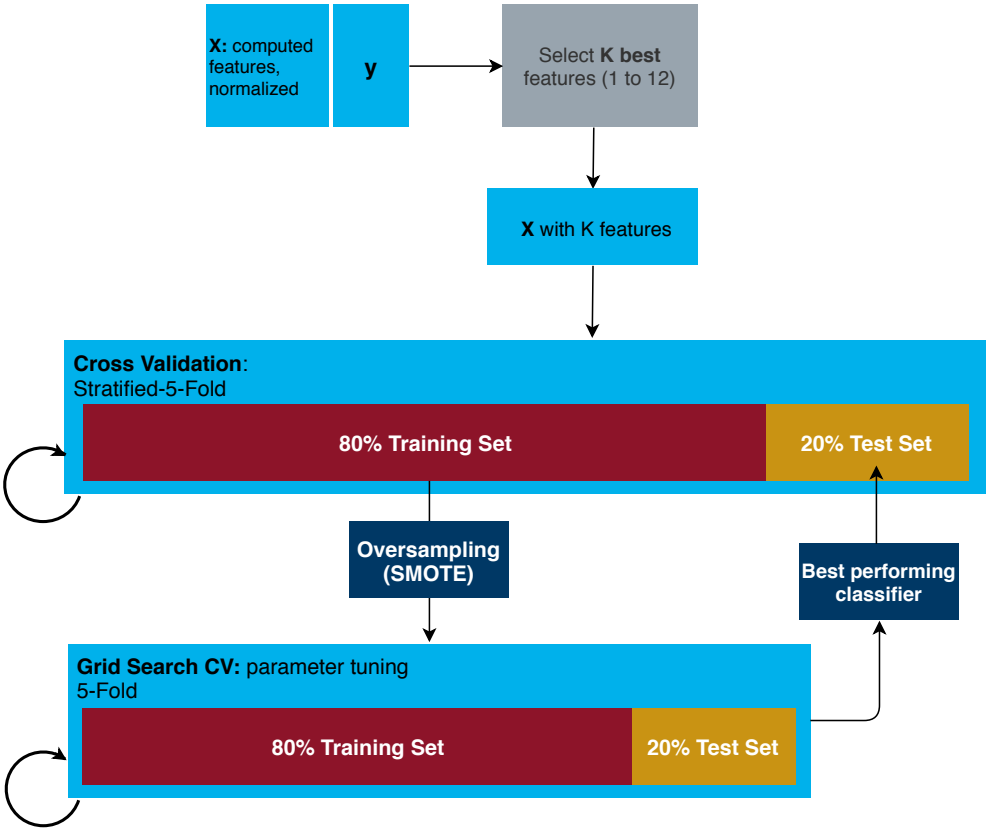


Figure 4.9: Cross Validation and Feature Selection Pipeline

Parameter Tuning

To find the best parameters for the *kNN* and *SVMs* classifiers, a *Grid Search* algorithm [Ped11] was applied. During grid search, every possible combination of parameters is systematically evaluated in order to find the best-performing parameter set. It evaluates the best number of neighbours k for the *kNN* classifier as well as the best C value and kernel function for the *Support Vector Classifier*. The best hyperparameters for the *Random Forest Classifier* are found via *Randomized Search* [Ped11], as it takes only a small fraction of the computation time of *Grid Search*.

Classification Performance

The performances of the classification algorithms can be evaluated by their mean accuracy. Accuracy is defined as

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{4.10}$$

[Gar09] and is computed on every test data from the cross-validation split.
Another good measure for the performance of a classification algorithm is *Sensitivity* and *Specificity*. They are both computed from the values of the confusion matrix which contains *True Positives (TP), True Negatives (TN), False Positives (FP)* and *False Negatives (FN)*. *Sensitivity* is defined as

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4.11}$$

and means that, e.g. within a test for a certain disease, a person with disease is correctly classified as 'diseased'. On the other hand, *Specificity* is defined as the ability of the test to classify a healthy person as 'healthy' and is computed via

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4.12}$$

Chapter 5

Results

In this chapter, the results obtained from the previously explained methods are presented. The outcomes of the unsupervised learning approach are summed up first. The second part deals with the results from the classification process and the corresponding preprocessing steps.

5.1 Unsupervised Learning

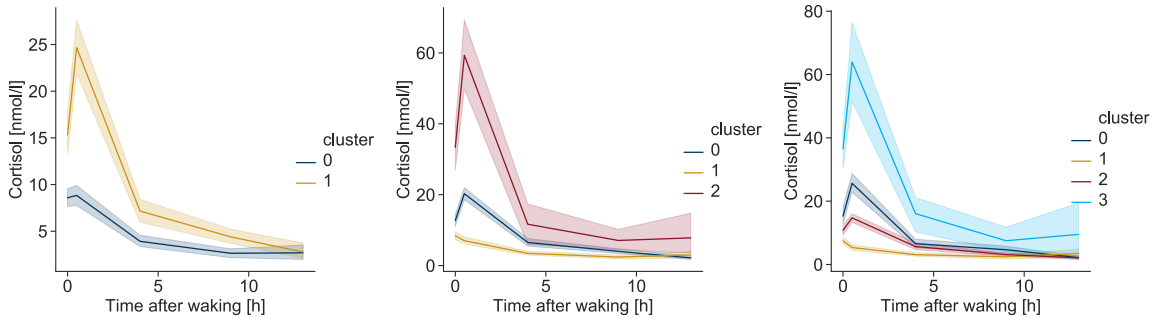
Silhouette Analysis

The clustering process was performed with both features computed *per day* and their *mean and difference*, *KMeans* and *Agglomerative Clustering* and different groups of clusters. First, the correct number of clusters had to be chosen which was done with a silhouette analysis. In Table 5.1 the mean silhouette scores from the different combinations of features, feature computation and algorithms are displayed. It is visible that there is just a slight difference between the scores of different numbers of clusters and also between the algorithms. A concrete example of the silhouette analysis is given in Appendix B.1 where the silhouette coefficient is plotted for the features AUC_i and *Slope* for a number of 2, 3 and 4 clusters found with *KMeans Clustering*.

In Figure 5.1, the results for different number of clusters seem quite similar. However, after consultation with psychological experts $n = 4$ clusters and clustering with *KMeans* based on AUC_{CAR_g} seem to be the most interpretable result. As displayed in Figure 5.1c, it leads to 3 clusters that are in line with previous findings from literature – cortisol profiles with totally flat curves (i.e. no CAR at all), profiles with a flattened *CAR* and profiles with a higher, "regular" CAR. Cluster 3 (displayed as light-blue in Figure 5.1c) only consists of two members which can be considered as 'outliers' as their rise of cortisol in the morning exceeds the typical increase stated in literature (about 50-60%) by far [Pru97]. Therefore, in the further analysis, the outlier group is

Table 5.1: **Mean silhouette coefficient** for clustering with *KMeans* or *Agglomerative Clustering*, based on one or two features, for `n_clusters` $\in [2, 4]$, maximum of each configuration highlighted in *italic*

			Kmeans		**Agglomerative**	
	Feature computation based on		days	mean/difference	days	mean/differnce
1 Feature	**n_clusters:**	**2**	*0.476*	*0.466*	*0.454*	*0.547*
		3	0.399	0.415	0.377	0.400
		4	0.377	0.441	0.378	0.418
2 Features	**n_clusters:**	**2**	*0.399*	*0.369*	*0.413*	*0.376*
		3	0.306	0.320	0.320	0.293
		4	0.285	0.303	0.286	0.272



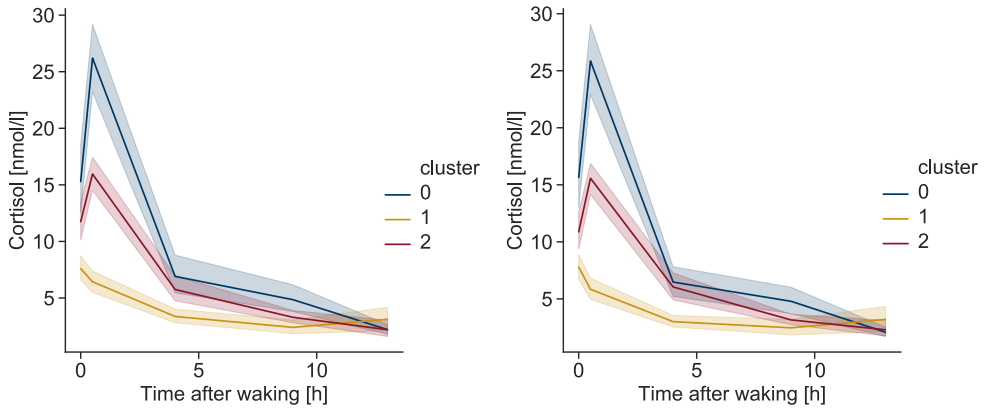
(a) **2 clusters**, class distribution [65/42]

(b) **3 clusters**, class distribution [53/51/3], cluster no. 2 contains outliers

(c) **4 clusters**, class distribution [26/37/42/2], cluster no. 3 contains outliers

Figure 5.1: **Clustering results of cortisol profiles** with different number of clusters. Data were clustered based on feature AUC_{CAR_g} per day and *KMeans* algorithm. Results show that *n = 4 clusters* lead to the most reasonable profiles, that are in line with the current state of research.

excluded which leads to a total cluster number of n = 3 with 105 subjects. In comparison, when a number of only 3 clusters is applied from the beginning, different clusters are found as presented in Figure 5.1b: profiles with totally flat curves (51 subjects), profiles with slightly flattened curves 53 subjects) and 3 outliers which show an abnormal, very steep *CAR*. With this process, the 3 rhythms that are also known from literature could not be found. Additional clustering results with other features and different numbers of clusters can be found in Appendix B.2.



(a) Clustered with *KMeans*, class distribution [23/45/37]

(b) Clustered with *Agglomerative clustering*, class distribution [23/40/42]

Figure 5.2: **Clustering results of cortisol profiles** with different clustering algorithms and n = 3. Clustering based on AUC_{CAR_g} per day, similar clusters are found, but class distribution is slightly different

Clustering Algorithms

The different curves, determined by *KMeans* and *Agglomerative Clustering* based on feature AUC_{CAR_g} for each day separately are displayed in Figure 5.2. Hereby, the clustering algorithms nearly find the same clusters, there is just a slight difference in the distribution. The 2 outliers were already excluded. For that reason, further analysis only looks at the clusters obtained from the *KMeans* algorithm.

Correlation with Variables

As next step, the outcome of the clustering process is correlated with the most relevant variables assessed in the questionnaire. This is done by performing an *ANCOVA* with the groups found by the clustering algorithm.

The outcome of the *ANCOVA* with 'disease' as covariate is shown in Table 5.2. Results show that the control variable 'disease' has a considerable impact on the distribution of the variables analyzed in this thesis. This is also visible in the boxplot of variable *StressTH* in Figure 5.3 which shows the variable distribution for the different clusters for people with disease and without. More boxplots of other variables assessed via *STRAIN*, can be found in B.3, B.4 and B.5. Amongst subjects with disease, the range is much wider within each cluster group. In consultation with

Table 5.2: **Results of ANCOVA** between the 3 groups found with *KMeans* and control variable '*disease'*. p-values below the corrected significance level are highlighted in *italic*

Variable	Source	p-value	F-value
PHQ	cluster	0.0427	3.253
	disease	0.0345	4.591
K6	cluster	0.8890	0.118
	disease	0.0223	5.387
StressCT	cluster	0.3845	0.965
	disease	*0.0007*	11.962
StressTH	cluster	0.6520	0.430
	disease	*0.0001*	16.099
EATotCT	cluster	0.7278	0.319
	disease	0.7675	0.088
EATotTH	cluster	0.5424	0.615
	disease	0.3014	1.079
t1_ADSL_sum	cluster	0.2404	1.446
	disease	*0.0002*	14.613
t1_PSS_sum	cluster	0.6175	0.484
	disease	*0.0002*	14.547
t1_STADI_S	cluster	0.5050	0.688
	disease	0.0181	5.769
t1_STADI_T	cluster	0.9990	*0.001*
	disease	0.0040	8.691
t1_ERQ_suppress	cluster	0.1909	1.683
	disease	0.4331	0.619
t1_ERQ_reapp	cluster	0.2208	1.533
	disease	0.7849	0.075
t1_FSOZU_sum	cluster	0.9505	0.051
	disease	*0.0036*	8.878

psychological experts, this implies that people with a disease would have to be regarded separately. This analysis would be more complicated as 'disease' means any physical or mental disease and would require a more specific investigation. Therefore, people who stated to have any form of '*disease'* are excluded for the final part of the analysis, which leads to 50 subjects who declared to be disease-free. As a last step, an *ANOVA* is performed on the 13 psychological variables, as subjects with disease are not considered and no control variable is necessary anymore. The results of the *ANOVA* are presented in Table 5.3. No variable lead to a significant outcome with a *p-value* below 0.385%.

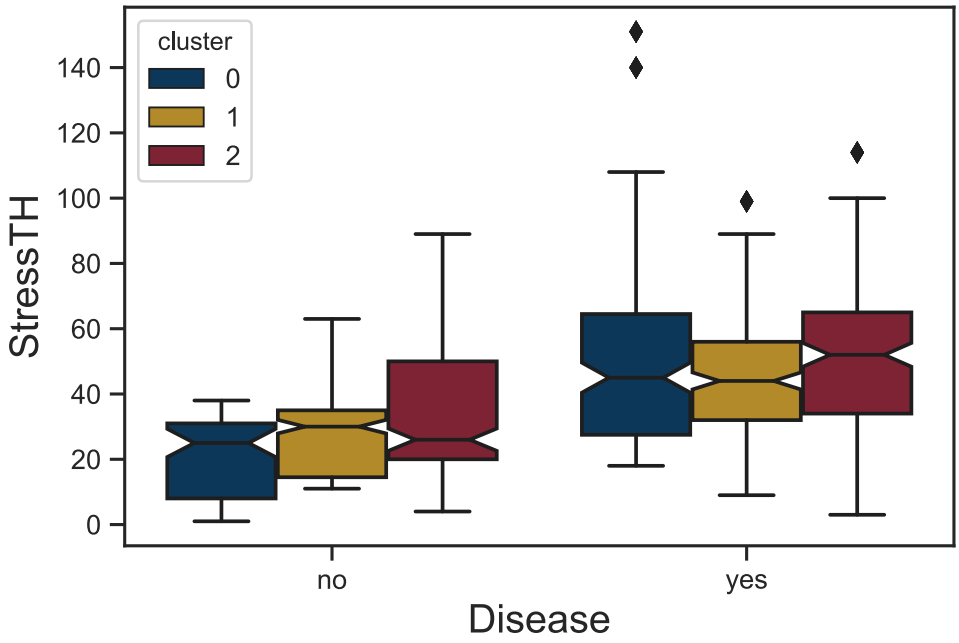


Figure 5.3: **Boxplot of variable *StressTH*** for 3 clusters, the notch in the boxplot marks the **confidence interval** (95%), the diamonds present **outliers**. Boxplots show that the variable range is wider for people with disease

Table 5.3: **Results of ANOVA** for most relevant variables, no significant p-values (<0.385%)

Variable	**p-value**	**F-value**
PHQ	0.3475	1.081
K6	0.8136	0.207
StressCT	0.7794	0.251
StressTH	0.9912	0.009
EATotCT	0.4958	0.712
EATotTH	0.5803	0.551
t1_ADSL_sum	0.9933	0.007
t1_PSS_sum	0.2676	1.356
t1_STADI_S	0.7327	0.313
t1_STADI_T	0.3247	1.152
t1_ERQ_suppress	0.2365	1.487
t1_ERQ_reapp	0.7031	0.355
t1_FSOZU_sum	0.1962	1.686

5.2 Supervised Learning

Feature Selection

After the generation of labels, the classification pipeline began with the selection of the best features. The results of the *Select K Best* algorithm depend on the labelling process, whether it was done by clustering or percentile split. Additionally, they are dependent on the variable that it was split on, as this determines the distribution of the features within a group. In Table 5.4 the *F-values* of all features are presented for the computation for each day separately and the mean and difference between the days. The labels were generated by clustering based on the variable *Total Count of Stressors* as this variable delivers good results in the whole classification process. Hereby, it is visible that the features' scores were not significantly better for one or the other approach. Therefore, they are both investigated further.
The feature selection algorithm was performed with $k \in [1, 12]$ but the number of features highly varied between the different variables and classifiers. For the further presented classification results, for each variable and classifier, the best combination of 1 to 12 features was used, i.e. the subset of features that lead to the highest accuracy.

Table 5.4: **Scores (ANOVA F-values)** for variable *'Total Count of Stressors'* for feature computation methods, maximum scores of each approach highlighted in **bold**

	F-values (ANOVA)			
Feature	*per day*		*mean/difference*	
AUC_g	day 1	0.0256	mean	0.0237
AUC_g	day 2	0.0147	difference	0.0106
AUC_i	day 1	1.1266	mean	2.0510
AUC_i	day 2	1.9379	difference	1.2455
AUC_{CAR_g}	day 1	0.2601	mean	0.1965
AUC_{CAR_g}	day 2	0.0935	difference	2.0282
AUC_{CAR_i}	day 1	0.2404	mean	1.1473
AUC_{CAR_i}	day 2	1.5785	difference	0.3168
$Slope$	day 1	3.3181	mean	**4.6918**
$Slope$	day 2	**3.7636**	difference	0.1495
$Slope_{CAR}$	day 1	0.3010	mean	0.7562
$Slope_{CAR}$	day 2	0.9652	difference	0.3023

Classification

When searching for the best performing classifier, the best parameters for all three classifiers were found with *Grid Search CV* and *Randomized Search CV* [Ped11]. The list of parameters with the corresponding accuracies for the *Support Vector Classifier* are shown in Table 5.5. A random fold from the *Stratified K-Fold* was picked, with the average accuracy over the 5 cross-validation folds within the *Grid Search* algorithm. The highest accuracies throughout all 5 cross-validation folds were obtained by applying the *radial basis function* kernel and a *C* value of 100.
In general, the best results were obtained for the classification of 2 classes with labels based on clustering, so the focus lies on this part of the classification results. The accuracy of the algorithms varied between the variables as well. The most significant results of the different classifiers are presented for both feature computation methods individually in Table 5.6 and Table 5.7.
For the classification with features computed for each day, *Random Forest Classifier* performed best, for the other approach *SVMs* lead to the highest accuracy.

Table 5.5: **Grid Search Results for SVM** with variable '*Total Count of Stressors*', best result highlighted in *italic*

C	**kernel**	**Mean test score (%)**
1	linear	64.3
	poly	66.7
	rbf	65.9
10	linear	68.3
	poly	69.0
	rbf	*81.0*
100	linear	70.6
	poly	77.0
	rbf	*81.0*

Table 5.6: **Classification Results**: Variable with the highest Accuracy in % (*Mean* ± *Standard Deviation*). Classification based on feature computation with mean and difference of the 2 days, best performing classifier highlighted in *italic*.

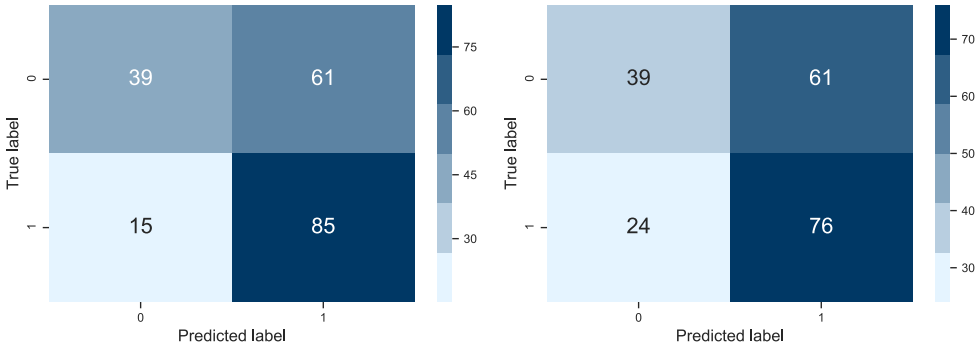
Variable	Classifier	Accuracy	Number of features used
Total Count of Stressors	SVM	*77.7 ± 7.5*	10
	kNN	71.9 ± 5.6	9
	Random Forest	71.4 ± 10.0	12
Count of Stressors during Childhood	SVM	*69.0 ± 5.6*	9
	kNN	58.9 ± 13	1
	Random Forest	68.0 ± 8.8	4

Table 5.7: **Classification Results**: Variable with the highest Accuracy in % (*Mean* ± *Standard Deviation*). Classification based on feature computation for each day seperately, best performing classifier highlighted in *italic*.

Variable	Classifier	Accuracy	Number of features used
Total Count of Stressors	SVM	68.5 ± 14.3	10
	kNN	65.6 ± 11.9	7
	Random Forest	*69.4 ± 12.3*	7
Count of Stressors during Childhood	SVM	73.0 ± 4.8	12
	kNN	71.2 ± 7.3	1
	Random Forest	*74.6 ± 10.6*	1

The confusion matrices of *SVMs* and *Random Forests* are shown in Figure 5.4 for the mean and difference feature computation and in Figure 5.5 for the features per day, respectively. From the confusion matrix *Sensitivity* and *Specificity* can obtained. For example *SVMs* classify a person from class 1 ('positive' class) as such with a probability of 81%, whereas just 32% of the subjects from the 'negative' class (class 0) are detected as such.

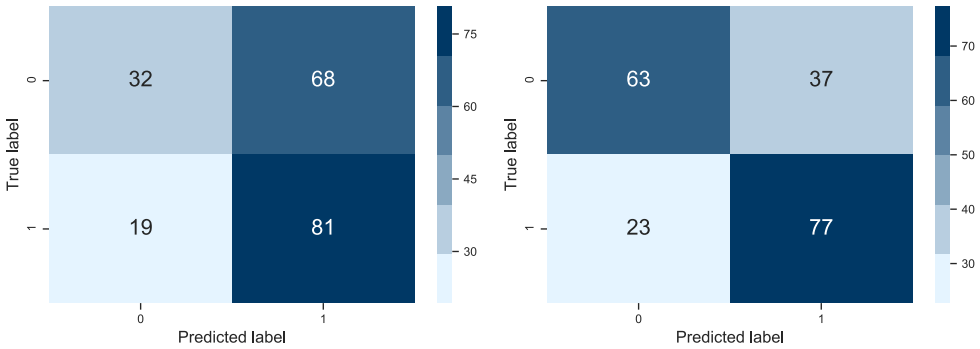
It is visible that mostly more than half of class '0' is misclassified as class '1'. Only *Random Forest* based on feature's mean and difference achieves to predict the classes correctly with an accuracy of 78% for class '0' and 63% for class '1'.



(a) Support Vector Classifier, Mean Accuracy 77.7±7.5%

(b) Random Forest, Mean Accuracy 71.4±10%

Figure 5.4: **Confusion matrix in %**, 2 classes, variable *StressCT*, Feature's mean and difference, class distribution ['0': 28/ '1': 79]



(a) Support Vector Classifier, Mean Accuracy 73.0±4.8%

(b) Random Forest, Mean Accuracy 74.6±10.6%

Figure 5.5: **Confusion matrix in %**, 2 classes, variable *EATotCT*, Features per day, class distribution ['0': 19/ '1': 88]

Chapter 6

Discussion

6.1 Unsupervised Learning

The first step of this work was to perform unsupervised learning and to find different groups of diurnal cortisol rhythms. At the beginning, the appropriate number of groups had to be found. It turned out that deciding for the number of groups only based on the highest score from the silhouette did not lead to the best solution. According to the scores, 3 clusters would have been the best choice, but when looking at the curves, the detected groups were found to be not reasonable. Together with psychological experts, the appropriate number could be found by investigating the differences in the cortisol profiles of the clusters. A descriptive analysis revealed that out of the 4 classes one group can reliably be considered as 'outliers' whereas the other 3 classes can be considered as 'expected', as they are in line with profiles previously reported in literature. The outliers showed an increase of cortisol of up to 90% after waking up, which is a much stronger response than normal [Pru97]. To investigate whether this happened due to measurement errors or whether a certain condition is related to such an enhanced response, a larger dataset with more subjects belonging to this group would be needed.
When choosing the clustering algorithm, quite similar results were obtained for both *KMeans* and *Agglomerative Clustering*. In further steps of variable correlation, it turned out that *KMeans* lead to better outcomes, even though the performance highly depends on the variable and the features used for clustering. Besides, due to the lower complexity of *KMeans*, this was the algorithm of choice.
From the *ANOVA* with different features and feature computation methods it turned out that the use of one feature computed for each day separately was the best option. During the discussion with psychological experts, the clustering of the curves based on the *Area under the Cortisol*

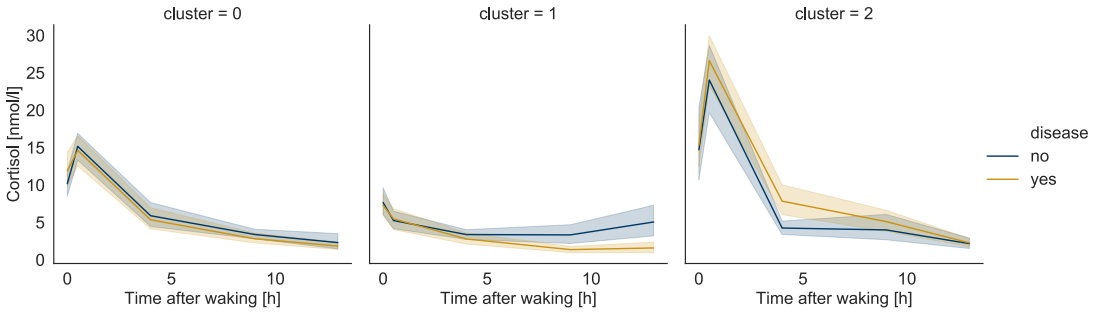


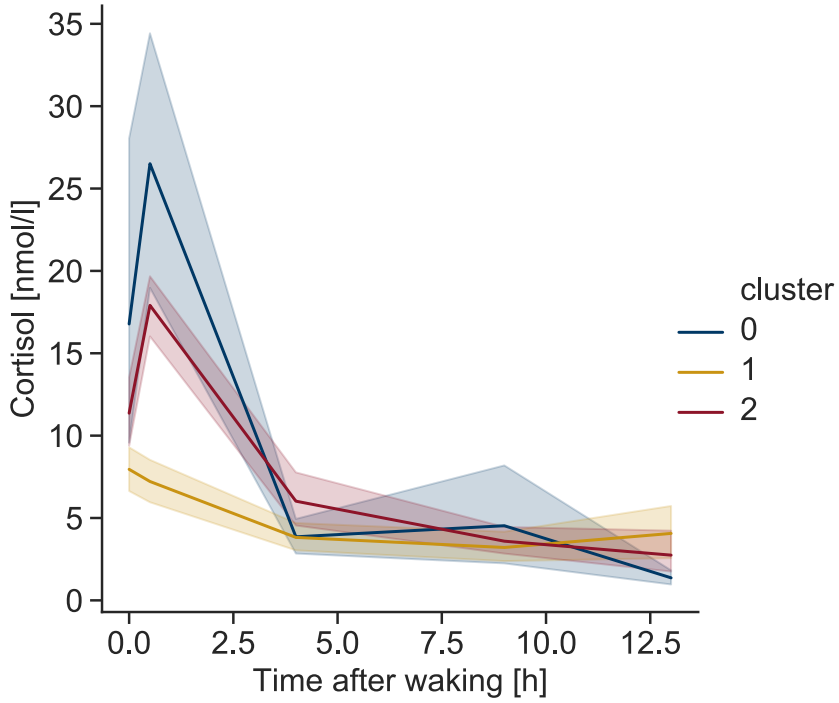
Figure 6.1: **Cortisol rhythms over all subjects** per cluster. Subjects with disease hardly differ from healthy subjects.

Awakening Response (AUC_{CAR_g}) was found out to work best. The *CAR* is a very important part of the diurnal rhythm and mainly characterizes the whole cycle. That is why the groups can be distinguished best based on this feature. The combination of many features produced unclear outcomes with no 'typical' courses of cortisol.

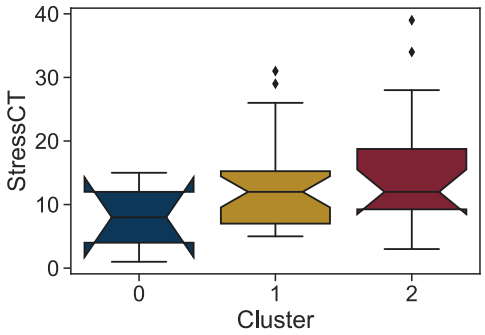
During the process of variable correlation, it was proved that the factor 'disease' has a very large impact on the distribution of the psychological variables. As seen in the boxplot (Fig.: 5.3), the range of the variable values is very wide. A possible explanation is the loose definition of 'disease' and hence the heterogeneity of this variable. Disease can include e.g. an acute cold but also depression which lead to very different outcomes in the questionnaire. This could also be seen in the results of the *ANCOVA*, where the *p-values* for every variable differed very much when controlling for disease. However, it is interesting that *disease* does not impact the course of the diurnal rhythm as shown in Figure 6.1. Here, the curves are hardly distinguishable, a difference is only observable in survey-assessed variables. To find out how disease impacts the diurnal cortisol rhythm and the psychological variables, different groups of diseases need to be defined which then have to be investigated separately. With a simple *'yes or no'* variable, no analysis of the linkages between the cycle of cortisol and the psychological variables can be performed.

In the last step, the analysis of differences among the groups 'disease' and 'non-disease' using ANOVA, no statistically significant results were found. None of the *p-values* was below the corrected level of significance (0.0385%). This outcome does not allow to reject the null hypothesis for any of the variables. Nevertheless, the clusters found can be explained with the boxplots of the variables *StressCT* and *StressTH* in Figure 6.2 very well. The subjects that belong to the 'regular' curve group (cluster '0') show the overall lowest values for both *StressCT* and *StressTH* and the range of values for both variables is not as wide as for the other two groups. According to previous work, people who show an altered *CAR* or even no *CAR* experience more

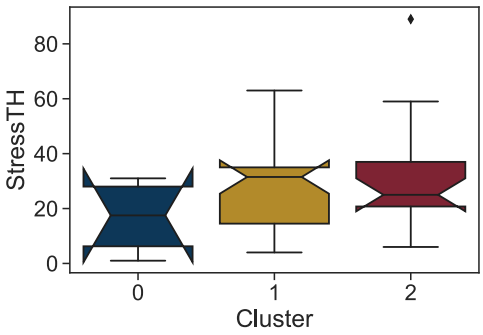
stressors and also rate them with a higher severity [Ste16, Fri09]. When looking at the two altered curves and corresponding variable distributions it can be assumed that the severity of the stressors experienced is decisive for the group that shows a completely flat curve without a *CAR*, as the values of that variable began at a higher level and reached higher than for the group with a flattened



(a) Final clusters, subjects with disease excluded, 3 *'typical'* curves: *'regular'* curve: cluster '0', *'flattened CAR'*: cluster '1' and *'no CAR'*: cluster '2'



(b) Boxplot for variable *StressCT* for 3 clusters



(c) Boxplot for variable *StressTH* for 3 clusters

Figure 6.2: **Results of clustering process** for healthy subjects and corresponding boxplots. The notch in the boxplot marks the **confidence interval** (95%), the diamonds present **outliers**

CAR. So, despite the non-significant results form statistics there is still evidence from descriptive analysis for a relationship between the biological and the psychological variables which have to be investigated further.

6.2 Supervised Learning

The results of the classification process showed that no reliable prediction of the group of diurnal cortisol profiles could be made. Even with the right combination of features and parameter optimization, the highest accuracy achieved was 77.7 ± 7.5 % (for *SVMs*). Nevertheless, the best classification was made by *Random Forests* classifier for variable *EATotCT*, because it reached both the highest sensitivity and specificity. The problem with the other classifiers was that despite the quite high accuracy only 39% or less of the members of class '0' could be predicted. This could be explained by the fact that class '0' was the minority class and the classifier tends to learn better on the majority class (class '1') and more often misclassifies the other class. When creating the labels by dividing the variables by their median, balanced classes were created. However, the problem with this approach was that often the same value for a variable was present in both classes due to the strict median split and values existing multiple times for different subjects. Therefore, the classification with these labels has not been investigated further.

Another explanation for the bad classification performance is the labelling process. A well-known paradigm in machine learning is that a classifier only performs as good as the labels provided. However, in this work, no ground truth labels were provided and, therefore, had to be generated. When the labels are not correct, the classifier cannot perform well. This also becomes visible regarding a pair plot of the features used for classification (Figure: 6.3). The groups based on labelling by clustering the variable *StressCT* are marked in different colours. The figure shows that the groups are not distinguishable at all based on the computed features. Also in Figure 6.4 the cortisol profiles look quite the same for both groups determined by the labelling process.

Accordingly, to improve the result of classification, another way to provide labels has to be found. Another possibility, which can be investigated in future work, would be the usage of other features which might differentiate the different groups better.

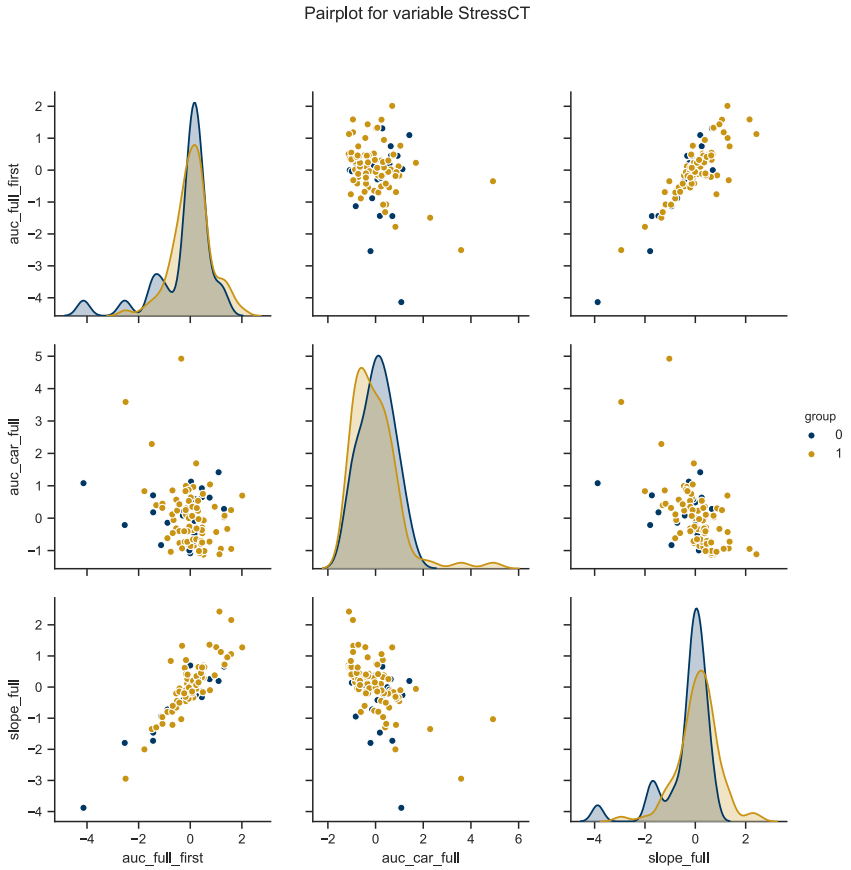


Figure 6.3: **Pairplot of features** used for classification with SVM and variable *StrsessCT* for labelling. The groups found by clustering are not clearly separable in the features of cortisol rhythm.

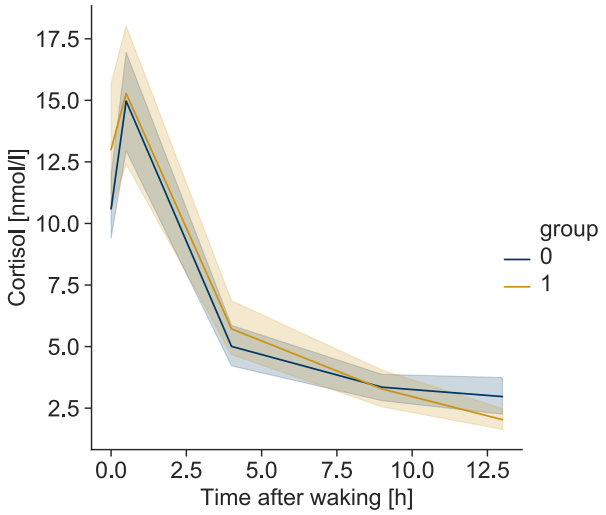


Figure 6.4: **Cortisol Profiles** of the groups determined by clustering variable *StressCT*

Chapter 7

Conclusion and Outlook

The goal of this thesis was to identify different stress responder types based on their diurnal cortisol profile with the help of different *ML* methods and the interpretation of the outcome with regard to the current state of research. Furthermore, it had to be examined whether a prediction of the cortisol profile based on psychological survey-assessed variables is possible. The base of this work was a dataset from a study with 107 subjects that included the diurnal course of cortisol on two consecutive days, and the scores obtained from different psychological questionnaires.
The exploration of diurnal cortisol rhythms and their relation to psychological variables was performed with 2 different machine learning approaches. The supervised approach – consisting of classification and prediction of the cortisol profiles – was not possible with the algorithms used in this work. Neither *SVMs*, *kNN* nor *Random Forests* were able to reliably predict labels that were generated from the survey-assessed variables with simple methods. The main problem was the absence of labels that would serve as ground truth in the data and the imbalance of classes resulting from the performed labelling process. In future work, another possibility to get concrete labels, e.g. from expert knowledge and the exploration of other algorithms could improve classification and make subsequent prediction possible.
Within the other approach, the clustering methods applied were able to separate typical cycles of cortisol as they are known from current literature. Even though the statistical analysis did not result in significant findings, relationships between the psychological variables and the cortisol rhythm could be identified. A huge part of the study population had to be excluded from the final analysis due to a high prevalence of some type of disease. Further research could be dealing with different types of diseases and their outcome in biological and psychological variables, as in this work it could be seen that this factor has a considerable impact. Besides, future studies might

help to reveal which psychological circumstances lead to cortisol profiles that were identified as ‘outliers’, e.g. an enhanced *CAR* that is nearly twice as high as normal.

Appendix A

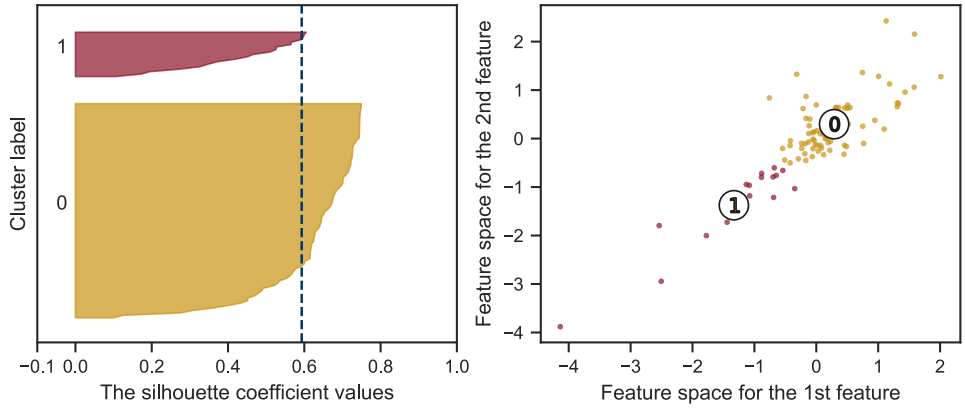
Additional Tables

Table A.1: **Stress variables** assessed via STRAIN Index

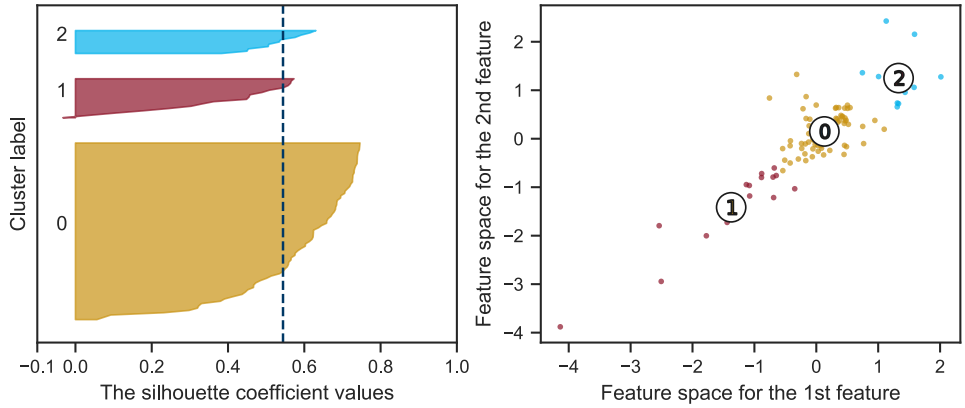
Variables assessed via STRAIN Index	
Physical Helath Complaints/Symptoms	
Mental Health Complaints/Symptoms	
Core and Transition to College	Total Count of Stressors
Core and Transition to College	Total Severity of Stressors
Core	Total Count of Stressors
Core	Total Severity of Stressors
Prenatal	Total Count
Early Adversity	Count of Acute Life Events
Early Adversity	Severity of Acute Life Events
Adulthood	Total Count
Adulthood	Total Severity
Housing	Total Count
Housing	Total Severity
Education	Total Count
Education	Total Severity
Work	Total Count
Work	Total Severity
Treatment/Health	Total Count
Treatment/Health	Total Severity
Marital/Partner	Total Count
Marital/Partner	Total Severity
Reproduction	Total Count
Reproduction	Total Severity
Financial	Total Count
Financial	Total Severity
Legal/Crime	Total Count
Legal/Crime	Total Severity
Death	Total Count
Death	Total Severity
Life-Threatening Situations	Total Count
Life-Threatening Situations	Total Severity
Possessions	Total Count
Possessions	Total Severity
Interpersonal Loss	Total Count
Interpersonal Loss	Total Severity
Physical Danger	Total Count
Physical Danger	Total Severity
Humiliation	Total Count
Humiliation	Total Severity
Entrapment	Total Count
Entrapment	Total Severity
Role Change/Reversal	Total Count
Role Change/Reversal	Total Severity

Appendix B

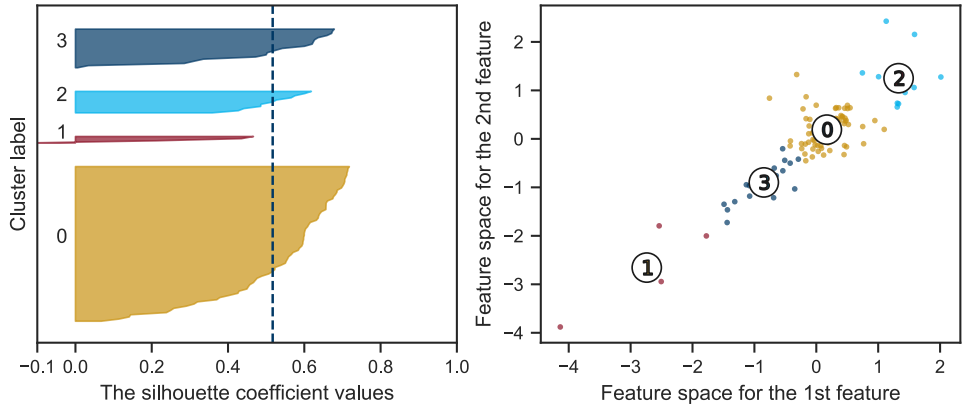
Additional Figures



(a) For 2 clusters there are no negative values, the highest mean silhouette coefficient is achieved

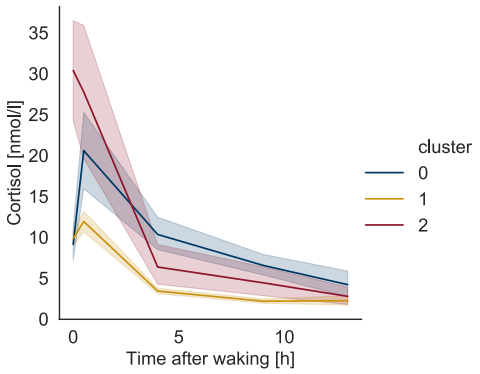


(b) For 3 clusters there are only a few samples with a negative coefficient, mean silhouette score is nearly as good as for 2 clusters

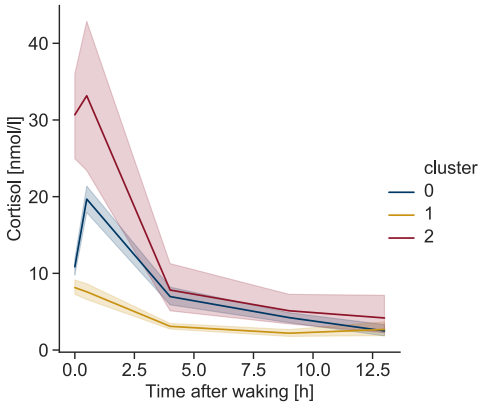


(c) With 4 clusters, cluster nnumber one only contains very few values

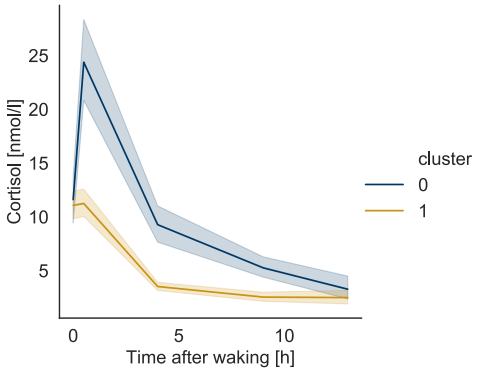
Figure B.1: **Silhouette plots** for *KMeans* clustering with mean of features: AUC_i and **Slope**. Mean silhouette coefficient presented by the vertical dashed line.



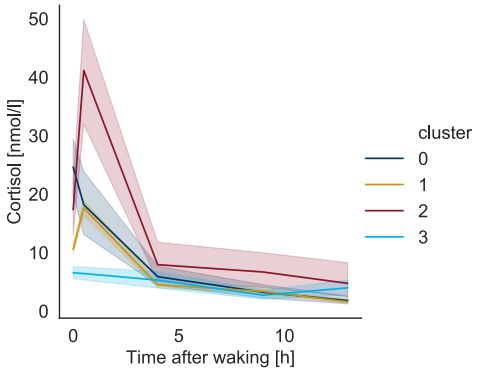
(a) 3 clusters, **Features:** AUC_g, AUC_i, class distribution [24/75/8]



(b) 3 clusters, **Features:** AUC_{CAR_g}, $Slope$,AUC_g, AUC_i, class distribution [47/51/9]



(c) 2 clusters, **Features:** AUC_{CAR_i}, $Slope$, class distribution [31/76]



(d) 4 clusters, **Features:** $Slope$, $Slope_{CAR}$ class distribution [13/47/7/40]

Figure B.2: **Clustering results of cortisol profiles** with *Kmeans*, different number of clusters and different combination of features, all computed for each day seperately.

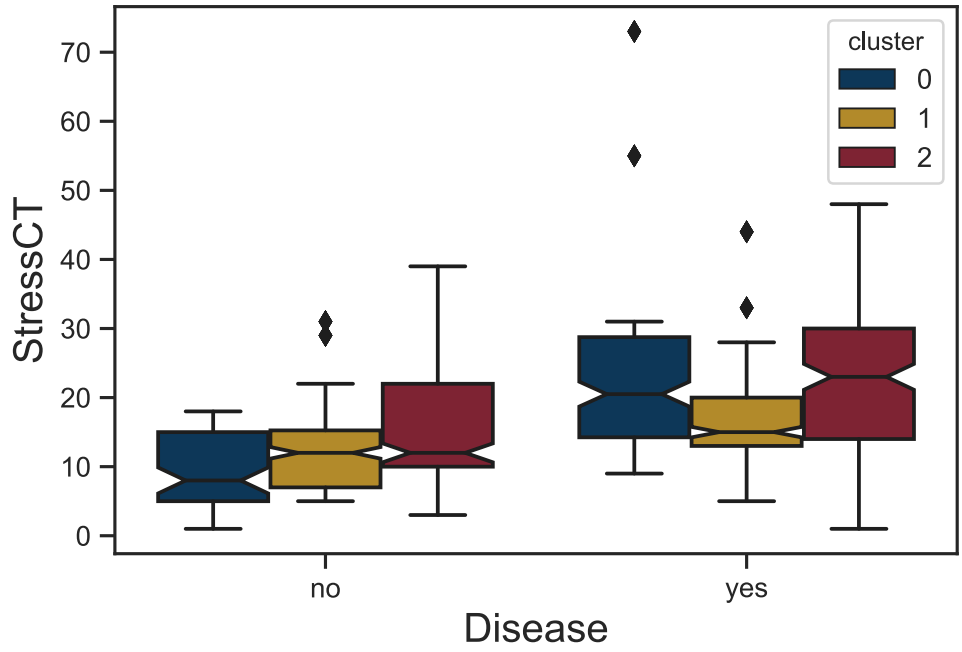


Figure B.3: **Boxplot of variable *StressCT*** for 3 clusters, the notch in the boxplot marks the **confidence interval** (95%), the diamonds present **outliers**.

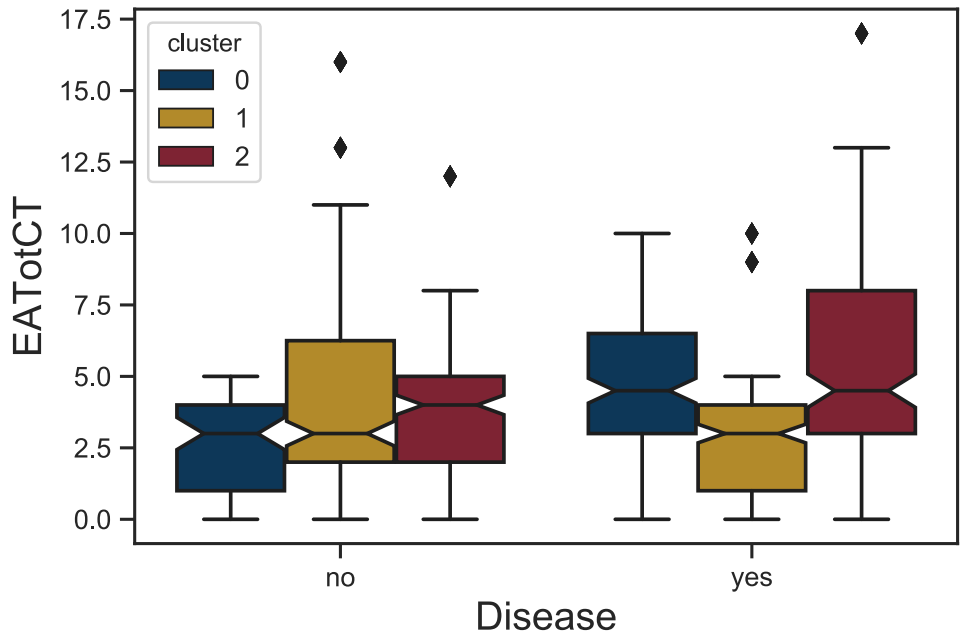


Figure B.4: **Boxplot of variable *EATotCT*** for 3 clusters, the notch in the boxplot marks the **confidence interval** (95%), the diamonds present **outliers**. The values of the healthy group are the highest for the completely flat curve (cluster 1).

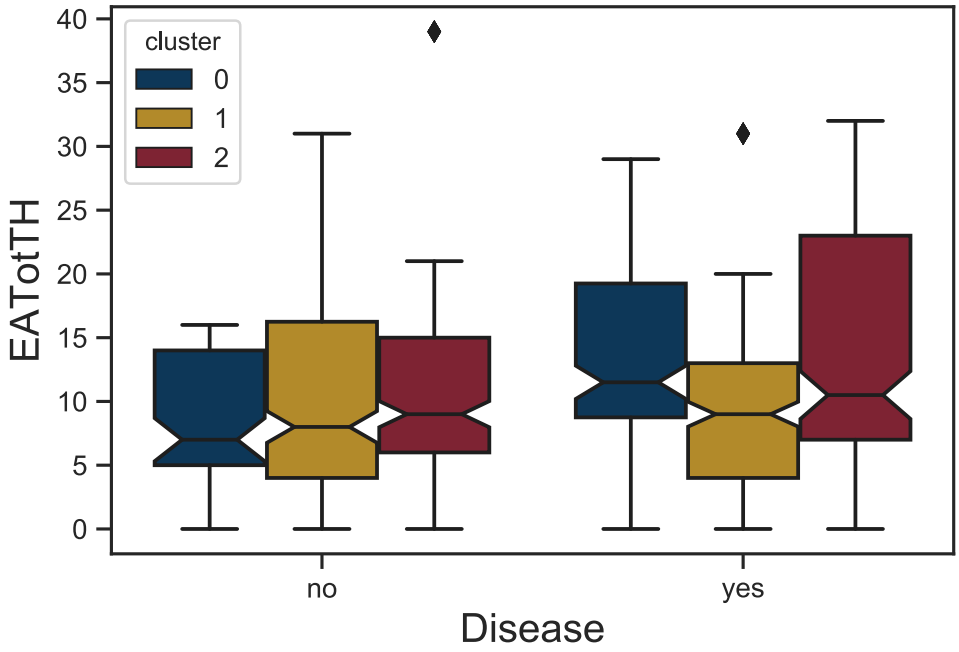


Figure B.5: **Boxplot of variable *EATotTH*** for 3 clusters, the notch in the boxplot marks the **confidence interval** (95%), the diamonds present **outliers**. The values of the healthy group are the highest for the completely flat curve (cluster 1).

Glossary

ACTH Adrenocorticotrophic Hormone

ANCOVA Analysis of Covariance

ANOVA Analysis of Variance

AUC Area Under the Receiver Operating Characteristics curve

AUC Area Under the Curve

CAR Cortisol Awakening Response

Cortisol is a steroid hormone and belongs to the class of glucocorticoides

CRF Corticotropin Releasing Factor

ECG Electrocardiogram

HPA Axis Hypothalamic-Pituitary-Adrenal Axis

IL-6 Interleukin-6

kNN K Nearest Neigbbor

MDD Major Depressive Disorder

ML Machine Learning

PTSD Posttraumatic Stress Disorder

SCN Suprachiasmatic Nuclei

SES Socioeconomic Status

SNS Sympathetic Nervous System

SVMs Support Vector Machines

TSST Trier Social Stress Test

List of Figures

List of Tables

Bibliography

[Abe19] L. Abel, R. Richer, A. Küderle, S. Gradl, B.M. Eskofier, and N. Rohleder. Classification of Acute Stress-Induced Response Patterns. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare - PervasiveHealth'19*, pages 366–370, New York, New York, USA, 2019. ACM Press.

[Ada10] E.K. Adam, L.D. Doane, R.E. Zinbarg, S. Mineka, M.G. Craske, and J.W. Griffith. Prospective prediction of major depressive disorder from cortisol awakening responses in adolescence. *Psychoneuroendocrinology*, 35(6):921–931, 2010.

[Ada17] E. K. Adam, M. E. Quinn, R. Tavernier, M. T. McQuillan, K. A. Dahlke, and K. E. Gilbert. Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. *Psychoneuroendocrinology*, 83:25–41, 2017.

[Bis07] C. M. Bishop. *Machine Learning and Pattern Recoginiton*. Springer, 2007.

[Bre01] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–33, 2001.

[Bre15] J. G. Breines, C. M. McInnis, Y. I. Kuras, M. V. Thoma, D. Gianferante, L. Hanlin, X. Chen, and N. Rohleder. Self-compassionate young adults show lower salivary alpha-amylase responses to repeated psychosocial stress. *Self Identity*, 14(4):390–402, 2015.

[Buc04] T. W. Buchanan, S. Kern, J. S. Allen, D. Tranel, and C. Kirschbaum. Circadian regulation of cortisol after hippocampal damage in humans. *Biological Psychiatry*, 56(9):651–656, 2004.

[Caw10] G. C. Cawley and N. L.C. Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11:2079–2107, 2010.

[Cha10] S. Chan and M. Debono. Review: Replication of cortisol circadian rhythm: New advances in hydrocortisone replacement therapy. *Therapeutic Advances in Endocrinology and Metabolism*, 1(3):129–138, 2010.

[Chi09] Y. Chida and A. Steptoe. Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. *Biological Psychology*, 80(3):265–278, 2009.

[Clo04] A. Clow, L. Thorn, P. Evans, and F. Hucklebridge. The Awakening Cortisol Response: Methodological Issues and Significance. *Stress*, 7(1):29–37, 2004.

[Coh06] S. Cohen, J. E. Schwartz, E. Epel, C. Kirschbaum, S. Sidney, and T. Seeman. Socioeconomic status, race, and diurnal cortisol decline in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Psychosomatic Medicine*, 68(1):41–50, 2006.

[Coh07] S. Cohen, D. Janicki-Deverts, and G. E. Miller. Psychological Stress and Disease. *Journal of the American Medical Association*, 298(14):1685–1687, 2007.

[Coh12] S. Cohen, D. Janicki-Deverts, W. J. Doyle, G. E. Miller, E. Frank, B. S. Rabin, and R. B. Turner. Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. *Proceedings of the National Academy of Sciences*, 109(16):5995–5999, 2012.

[De 03] W. De Vente, M. Olff, J. G. C. Van Amsterdam, J. H. Kamphuis, and P. M. G. Emmelkamp. Physiological differences between burnout patients and healthy controls: Blood pressure, heart rate, and cortisol responses. *Occupational and Environmental Medicine*, 60:54–61, 2003.

[De 05] M. E. De Vugt, N. A. Nicolson, P. Aalten, R. Lousberg, J. Jolle, and F. R. J. Verhey. Behavioral problems in dementia patients and salivary cortisol patterns in caregivers. *Journal of Neuropsychiatry and Clinical Neurosciences*, 17(2):201–207, 2005.

[Des15] A. S. Desantis, C. W. Kuzawa, and E. K. Adam. Developmental origins of flatter cortisol rhythms: Socioeconomic status and adult cortisol activity. *American Journal of Human Biology*, 27:458–467, 2015.

[Edw01] S. Edwards, P. Evans, F. Hucklebridge, and A. Clow. Association between time of awakening and diurnal cortisol secretory activity. *Psychoneuroendocrinology*, 26:613–622, 2001.

[Fri09] E. Fries, L. Dettenborn, and C. Kirschbaum. The cortisol awakening response (CAR): Facts and future directions. *International Journal of Psychophysiology*, 72:67–73, 2009.

[Gar09] S. García, A. Fernández, J. Luengo, and F. Herrera. A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. *Soft Computing*, 13:959–977, 2009.

[GL14] I. R. Galatzer-Levy, K. I. Karstoft, A. Statnikov, and A. A. Shalev. Quantitative forecasting of PTSD from early trauma responses: A Machine Learning application. *Journal of Psychiatric Research*, 59:68–76, 2014.

[GL17] I. R. Galatzer-Levy, S. Ma, A. Statnikov, R. Yehuda, and A. Y. Shalev. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Translational Psychiatry*, 7, 2017.

[Gro05] G. Grossi, A. Perski, M. Ekstedt, T. Johansson, M. Lindström, and K. Holm. The morning salivary cortisol response in burnout. *Journal of Psychosomatic Research*, 59:103–111, 2005.

[Gun98] S. R. Gunn. Support Vector Machines for classification and regression. Technical report, 1998.

[Has17] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2017.

[Hil] Mark Hill. UNSW embryology. `https://embryology.med.unsw.edu.au/embryology/index.php`. Accessed: 29.10.2019.

[Kim14] H. Kim. Analysis of variance (ANOVA) comparing means of more than two groups. *Restorative Dentistry & Endodontics*, 39(1):74–77, 2014.

[Kud06] B. M. Kudielka, I. S. Federenko, D. H. Hellhammer, and S. Wüst. Morningness and eveningness: The free cortisol rise after awakening in 'early birds' and 'night owls'. *Biological Psychology*, 72:141–146, 2006.

[Kue07] C. Kuehner, S. Holzhauer, and S. Huffziger. Decreased cortisol response to awakening is associated with cognitive vulnerability to depression in a nonclinical sample of young adults. *Psychoneuroendocrinology*, 32:199–209, 2007.

[Kum09] M. Kumari, E. Badrick, J. Ferrie, A. Perski, M. Marmot, and T. Chandola. Self-reported sleep duration and sleep disturbance are independently associated with cortisol secretion in the Whitehall II study. *Journal of Clinical Endocrinology and Metabolism*, 94(12):4801–4809, 2009.

[Lan11] F. Lang and P. Lang. Hormonale Regulation. In *Basiswissen Physiologie*, pages 235–268. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[McE08] B. S. McEwen. Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. *European Journal of Pharmacology*, 583:174–185, 2008.

[Mul00] M. Mullin and R. Sukthankar. Complete Cross-Validation for Nearest Neighbor Classifiers. In *ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning*, pages 639–646, 2000.

[Par08] R. Parikh, A. Mathai, S. Parikh, G. C. Sekhar, and R. Thomas. Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, 56:45–50, 2008.

[Ped11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Pru97] J. C. Pruessner, O. T. Wolf, D. H. Hellhammer, K. von Auer, S. Jobst, F. Kaspers, C. Kirschbaum, and A. Buske-Kirschbaum. Free Cortisol Levels After Awakening: A Reliable Biological Marker For The Assessment Of Adrenocortical Activity. *Life Science*, 61(26):2539–2549, 1997.

[Pru03] J. C. Pruessner, C. Kirschbaum, G. Meinlschmidt, and D. H. Hellhammer. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change (multiple letters). *Psychoneuroendocrinology*, 28:916–931, 2003.

[Roh19] N. Rohleder. Stress and inflammation - The need to address the gap in the transition between acute and chronic stress effects. *Psychoneuroendocrinology*, 105:164 – 171, 2019.

[Rou87] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[Rut01] A. Rutherford. *Introducing Anova and Ancova: A GLM Approach*. ISM (London, England). SAGE Publications, 2001.

[Sch98] P. Schulz, C. Kirschbaum, J. Prüßner, and D. Hellhammer. Increased free cortisol secretion after awakening in chronically stressed individuals due to work overload. *Stress Medicine*, 14:91–97, 1998.

[Sch04] W. Schlotz, J. Hellhammer, P. Schulz, and A. A. Stone. Perceived Work Overload and Chronic Worrying Predict Weekend-Weekday Differences in the Cortisol Awakening Response. *Psychosomatic Medicine*, 66:207–214, 2004.

[Sch14] C. Scheiermann, Y. Kunisaki, and P. S. Frenette. Circadian control of the immune system. *Natrue reviews Immunology*, 13(3):190–198, 2014.

[Sla18] G. M. Slavich and G. S. Shields. Assessing Lifetime Stress Exposure Using the Stress and Adversity Inventory for Adults (Adult STRAIN): An Overview and Initial Validation. *Psychosomatic Medicine*, 80:17–27, 2018.

[Sla19] G. M Slavich. Stressnology : The primitive (and problematic) study of life stress exposure and pressing need for better measurement. *Brain Behavior and Immunity*, 75:3–5, 2019.

[Sme16] E. Smets, P. Casale, U. Großekathöfer, B. Lamichhane, W. De Raedt, K. Bogaerts, I. Van Diest, and C. Van Hoof. Comparison of Machine Learning Techniques for Psychophysiological Stress Detection. In *Communications in Computer and Information Science*, volume 604, pages 85–98, 2016.

[Smy97] J. M. Smyth, M. C. Ockenfels, A. A. Gorin, D: Catley, L. S. Porter, C. Kirschbaum, D. H. Hellhammer, and A. Arthur. Individual Differences in the diurnal Cycle of Cortisol. *Science*, 22(2):89–105, 1997.

[Ste00] A. Steptoe, M. Cropley, J. Griffith, and C. Kirschbaum. Job strain and anger expression predict early morning elevations in salivary cortisol. *Psychosomatic Medicine*, 62:286–292, 2000.

[Ste16] A. Steptoe and B. Serwinski. *Cortisol awakening response*. Elsevier Inc., 2016.

[Sto01] A. A. Stone, J. E. Schwartz, J. Smyth, C. Kirschbaum, S. Cohen, D. Hellhammer, and S. Grossman. Individual differences in the diurnal cycle of salivary free cortisol: A replication of flattened cycles for some individuals. *Psychoneuroendocrinology*, 26:295–306, 2001.

[The09] S. Theodoridis and K. Koutrumbas. *Pattern Recognition*. Academic Press, 2 edition, 2009.

[Tho06] L. Thorn, F. Hucklebridge, P. Evans, and A. Clow. Suspected non-adherence and weekend versus week day differences in the awakening cortisol response. *Psychoneuroendocrinology*, 31:1009–1018, 2006.

[Utg97] P. E. Utgoff, N. C. Berkman, and J. A. Clouse. Decision Tree Induction Based on Efficient Tree Restructuring. *Machine Learning*, 29:5–44, 1997.

[Wan06] J. Wang, M. Xu, H. Wang, and J. Zhang. Classification of imbalanced data by using the smote algorithm and locally linear embedding. In *2006 8th international Conference on Signal Processing*, volume 3, Nov 2006.

[Wes06] M. Wessa, N. Rohleder, C. Kirschbaum, and H. Flor. Altered cortisol awakening response in posttraumatic stress disorder. *Psychoneuroendocrinology*, 31:209–215, 2006.

[Wil05] E. Williams, K. Magid, and A. Steptoe. The impact of time of waking and concurrent subjective stress on the cortisol response to awakening. *Psychoneuroendocrinology*, 30:139–148, 2005.

[Wri05] C. E. Wright and A. Steptoe. Subjective socioeconomic position, gender and cortisol responses to waking in an elderly population. *Psychoneuroendocrinology*, 30:582–590, 2005.

[Wüs00] S. Wüst, J. Wolf, D. H. Hellhammer, I. Federenko, N. Schommer, and C. Kirschbaum. The cortisol awakening response - normal values and confounds. *Noise & health*, 7:77–85, 2000.

[Zhe15] Z. Zheng, Y. Cai, and Y. Li. Oversampling method for imbalanced classification. *Computing and Informatics*, 34:1017–1037, 2015.