MACHINE LEARNING & DATA ANALYTICS

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Classification and Prediction of Acute Stress-Induced Response Patterns

Bachelor's Thesis in Medical Engineering

submitted
by

Luca Abel

born 27.02.1996 in Malsch

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. Björn Eskofier
(Machine Learning and Data Analytics Lab, FAU Erlangen-Nürnberg)
Prof. Dr. Nicolas Rohleder
(Chair of Health Psychology, FAU Erlangen-Nürnberg)

Started: 01.09.2018

Finished: 31.01.2019

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 31.01.2019

Übersicht

Das Potential von Maschinellem Lernen ermöglicht neue Ansätze zur Analyse psychologischer Daten. Machinelles Lernen bietet viele Vorteile im Vergleich zu traditionellen statistischen Methoden. In der Gesundheitspsychologie besteht großes Interesse daran, Prozesse zu erforschen, die durch akuten Stress ausgelöst werden. Daher wird mit dieser Arbeit eine Methode vorgestellt verschiedene Stressreaktionsmuster automatisch zu klassifizieren. Die Basis dafür bilden Daten, die von einem Experten der Stressforschung 'gelabelt' wurden. Drei wichtige stresssensible Marker wurden dabei berücksichtigt: Cortisol und alpha-Amylase im Speichel, sowie Interleukin-6 (IL-6) im Blutserum. Hierfür wurde eine zuverlässige Stressbelastung durch einen standardisierten Test, den *Trier Social Stress Test*, sichergestellt.

Mit Hilfe von *Support Vector Machines* konnten hohe Genauigkeiten bei der Zuordnung der Klassen erreicht werden. Bei der Klassifikation von vier verschiedenen Reaktionsmustern im Cortisol lag die mittlere Genauigkeit bei $92.2\ \% \pm 9.7\ \%$. Drei verschiedene IL-6 Reaktionen konnten mit einer mittleren Genauigkeit von $91.2\ \% \pm 6.3\ \%$ zugeordnet werden. Des weiteren lag die mittlere Genauigkeit für die Klassifizierung von vier verschiedenen Amylase-Mustern bei $61.2\ \% \pm 12.3\ \%$.

Außerdem wurden die statistischen Zusammenhänge der Muster der jeweiligen Parameter untersucht. Mögliche interessante Korrelationen treten für folgende Kombinationen auf: 'Amylase keine Gewöhnung' - 'Cortisol keine Antwort' (r = 0.28, p = 0.004), 'Amylase keine Antwort' - 'Cortisol antizipierter Stress' (r = 0.27, p = 0.006), 'IL-6 keine Gewöhnung' - 'Cortisol keine Antwort' (r = 0.25, p = 0.01), 'IL-6 Sensibilisierung' - 'Cortisol antiziperter Stress' (r = 0.26, p = 0.007) und 'IL-6 Sensibilisierung' - 'Amylase keine Antwort' (r = 0.25, p = 0.008).

Limitierungen traten dabei unter anderem aufgrund der relativ kleinen Datenmenge (63 Personen mit vollständigen Cortisol-Daten) auf. Des weiteren wurde die Differenzierung der einzelnen Klassen durch eine große Varianz innerhalb der Klassen erschwert.

Der in dieser Thesis vorgestellte Ansatz zur Klassifizierung verschiedener Stressreaktionsmuster kann in zukünftigen Arbeiten dazu verwendet werden ein System zur Vorhersage dieser Muster zu entwickeln.

Abstract

The potential of machine learning techniques enables new possibilities for the analysis of psychological data. Machine learning has many advantages compared to the traditional statistical methods. In the field of health psychology, there is a huge interest in exploring the processes which are triggered by acute stress on a systemic level. Therefore, this work introduces a method for the automatic prediction of stress response classes based on learning from data labeled by an expert. Three important stress-sensitive markers were used: Salivary cortisol, salivary alpha-amylase, and the Interleukin-6 concentration in the blood serum. The stress was induced using a standardized test protocol, the Trier Social Stress Test.

With the help of *Support Vector Machines* it was accomplished to label the classes with a high accuracy. For the classification of four different cortisol responder types a mean accuracy of $92.2\ \% \pm 9.7\ \%$ was accomplished. Three different IL-6 reactions could be classified with a mean accuracy of $91.2\ \% \pm 6.3\ \%$ and four types of amylase responses with a mean accuracy of $61.2\ \% \pm 12.3\ \%$.

Furthermore, the connections of the classes of different markers were explored statistically. Possible interesting correlations were found for ‘amylase non-habituation’ - ‘cortisol non-responder’ (r = 0.28, p = 0.004), ‘amylase non-responder’ - ‘cortisol anticipatory stress’ (r = 0.27, p = 0.006), ‘il6 non-habituation’ - ‘cortisol non-responder’ (r = 0.25, p = 0.01), ‘il6 sensitization’ - ‘cortisol anticipatory stress’ (r = 0.26, p = 0.007) and ‘il6 sensitization’ - ‘amylase non-responder’ (r = 0.25, p = 0.008).

Limitations, which occured in this work were especially caused by the small sample size (63 subjects with full cortisol data) and a high variance in the different classes.

For future work, the classification approach introduced in this thesis could be easily adapted for the use in a recommender system, in order to make stress response pattern classification faster and easier for new data.

Contents

Chapter 1

Introduction

Stress is a hidden epidemic – the World Health Organization estimates that mental diseases, including stress-related disorders, will be the second leading cause of disabilities by the year 2020 [Kal02]. The economic impact of stress is huge, with an estimated loss of about $42 billion for absence and treatment of stress-related illness in the U.S. [Kal02]. Acute stress, such as psychosocial stress, which is defined as stress experienced as a result from social interaction with others, induces strong biological responses. Through the secretion of stress hormones and widely spread autonomic innervation, these responses have the potential to reach nearly every cell in the organism [Sap00]. Whereas adequate stress responses are crucial for a correct physiological reaction, defective stress response patterns might be linked with the biological and physiological markers of aging like DNA damage, over-expression of inflammatory genes, and declines in cognitive functioning [Roh09]. There is evidence from previous studies that different response patterns exist, and some have shown to be associated with inflammatory diseases [Sch03]. Inflammation also signals into the central nervous system, where it might be responsible for age-related cognitive decline and mood disturbances [Wea02]. For further research, it is important to classify these patterns in order to identify such defective responses automatically.

This is why *Machine Learning (ML)* becomes interesting for psychology, because recognizing patterns and assigning them to different classes (also known as *classification*) is one of the main use cases of *ML*. Whereas *ML*-based techniques have been widely used for years in other fields of medicine and health care, the application for psychological problems like diagnostics, treatment, and research is rather rare. Nevertheless, there is huge potential in this interdisciplinary challenge. Most psychological publications are based on relatively simple statistical methods, compared to the possibilities of *ML*.

In this work, different classification algorithms were used to examine whether the classification of stress response patterns based on biological variables is possible. The biological data was recorded in a preceding study using the *Trier Social Stress Test* (further described in Section 4.1.1) for reliable induction of psychosocial stress. Additionally, this work attempts to find connections between the patterns of different biological markers through statistical analysis.

In the following chapter, the medical background of the acute stress response is outlined. Chapter 3 gives an overview about previous works using machine learning for health psychology. Additionally, related work in the field of stress response, which are based on similar datasets, are discussed. The next chapter (no. 4) describes the methods used for stress pattern classification and pattern correlation. This includes a description of preprocessing and feature selection as well as the learning and evaluation processes. Furthermore, the statistical methods used in order to find connections between the patterns are explained. All the results, which are shown in Chapter 5, will be discussed in Chapter 6 along with the discussion of the methods used. A conclusion and an outlook on how this work can be used in further research will be given in Chapter 7.

Chapter 2

Medical Background

According to Hans Selye, the pioneer of stress research, stress is defined as the "unspecific reaction of an organism to all kinds of demands" [Udr88]. These demands vary from all kinds of psychological stress, which are typically situations with low predictability, low controllability and novelty [Kir94]. Some common examples are exams, job interviews or dentist treatments. Nevertheless, stress reactions are not only a psychological phenomenon, but can also follow on different kinds of physical exercise, like running or bicycle riding [Nat09].

In order to understand what is happening in the body when a subject is under acute or chronic stress two important pathways have to be investigated: The *Sympathetic Nervous System (SNS)* and the *Hypothalamic–Pituitary–Adrenal Axis (HPA-Axis)*. Through these stress is sending signals about the threatened state of the organism to the target tissues [Roh09].

The *SNS* is a part of the *Autonomous Nervous System (ANS)*, which controls nearly all physiological functions. An example is the cardio-circulatory system, which is controlled through adaptations of the vessels or the heart rate. Furthermore, reflexes like the pupil reflex are controlled by the *ANS* [Lan11d]. In general, the *SNS* stimulates effects which are important for the 'fight-or-flight' response, while the antagonist, the parasympathetic nervous system, mainly inhibits these effects.

Another mechanism, which is also regulated by the *ANS*, is the secretion and composition of saliva. Therefore, the concentration of *Salivary Alpha-Amylase (sAA)*, which is one of most important enzymes in the saliva, is an indirect indicator of *Automatic Activation* [Roh09]. *Automatic Activation* is usually high during psychological stress [Roh09]. Alpha-amylase is the enzyme that is responsible for the first digestion of carbohydrates during the chewing process [Lan11b].

The second mechanism is the *HPA-Axis*. It can be defined as an interactive neuroendocrine unit comprising of the hypothalamus, the pituitary gland and the adrenal cortex [Orb13]. The

HPA-Axis plays a key role in stress reaction and *Homeostasis* as well. As visualized in Figure 2.1, the hypothalamus releases *CRH*, which is stimulating the production of *ACTH* in the anterior pituary. An increase in *ACTH* results in the production and secretion of cortisol in the adrenal cortex [Orb13]. *Cortisol* is a glucocortoid belonging to the class of steriod hormones. The purpose of glucocortoids is to stimulate the conversion of proteins to glucose.

The *HPA-Axis* is an example of a negative feedback loop. Cortisol secretion limits itself through feedback to the hypothalamus and the anterior pituitary [Orb13]. Increased cortisol production has numerous effects like enhancement of vascular activity, reduced immune responses, stimulated gluconeogenesis and inhibition of nonessential functions [Orb13]. The typical reaction to acute stress is that the cortisol level increases compared to baseline, with a peak about 15 minutes after stress. Then the cortisol level slowly declines until it recovers to baseline after about 2 hours. There are some hints that a dysregulation of the *HPA-Axis* is connected to various physiological and psychological illnesses [Orb13].

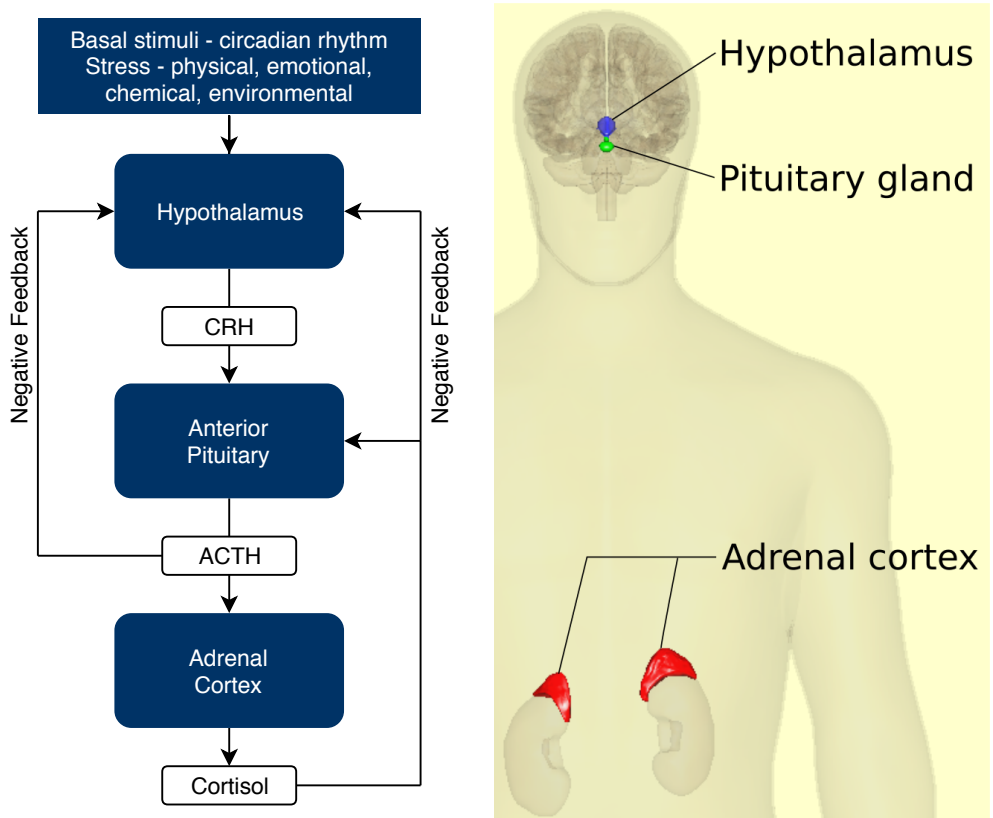


Figure 2.1: Schematic representation of the *Hypothalamic–Pituitary–Adrenal Axis*. *CRH*: *Corticotrophin Releasing Hormone*, *ACTH*: *Adrenocorticotrophic Hormone*, left: modified from [Orb13], right: [Ana15]

Another activator of the *HPA-Axis* is the pro-inflammatory hormone *Interleukin-6 (IL-6)* [Mas93]. Interleukins are one of the five main groups of cytokines, proteins that regulate the growth and differentiation of cells. Interleukins are messenger substances of the cells involved in the immune reaction. An overview of important interleukins with their place of production and their effects can be found in Table 2.1.

Table 2.1: Interleukins and their Effects. Modified from [Lan11a].

Interleukin	Place of Production	Main Effect(s)
IL-1	Macrophages	Stimulation of T helper cells
IL-2	T cells	Proliferation and maturing of T cells, stimulation of B cells
IL-3	T cells	Stimulation of hematopoiesis (no inflammatory effects)
IL-4	T cells	Growth and differentiation of B cells, growth of T cells
IL-5	T helper cells	Differentiation of B cells
IL-6	*Macrophages, T cells*	*Maturing of B cells*
IL-7	T cells	Proliferation of T and B cells
IL-8	T cells	Activation of granulocytes
IL-10	T helper cells	Inhibition of T helper cells, differentiation of B cells

IL-6 is not the only cytokine effecting the *HPA-Axis*. IL-1 and IL-10 also have a stimulating influence on the *HPA-Axis* [Dun07]. Acute psychosocial stress is inducing a short-time rise in cytokines in healthy adult humans according to a study of Miller et al. [Mil05]. Additionally, elevated baseline *IL-6* levels had been connected to a risk of decline in cognitive function [Wea02].

Chapter 3

Related Work

Machine Learning is one of the most important methods in todays research. It has become ubiquitous and one often encounters results of Machine Learning processes in daily life without even noticing. The improvement potential for fields like industry, medicine, commerce, etc. from these methods is huge, but there are only few examples for the usage of Machine Learning in psychology, especially health psychology.

The work of Galatzer-Levy et al. is focusing on *Posttraumatic Stress Disorder (PTSD)*. They used *Support Vector Machines (SVMs)* along with Markov Boundary feature selection to predict non-remitting *PTSD* from information collected within 10 days after a traumatic event [GL14]. In their work, they reached about 95% mean accuracy (AUC = 0.77). Another interesting outcome of this work is that the prediction accuracy remains the same (AUC = 0.77) for a smaller subset of features, compared to the prediction based on the whole dataset [GL14]. Based on these findings, Galatzer-Levy et al. published another work using Machine Learning for PTSD prediction. Their results show that a high PTSD prediction accuracy (AUC = 0.82) is possible when using SVMs along with combined clinical, neuroendrocrine, psychophysiological and demographic information [GL17].

Another publication by Galatzer-Levy et al. used latent class analysis to find patterns of *PTSD Comorbidity*. They found three relevant classes, a class characterized by predominantly comorbid mood and anxiety disorders, a class characterized by predominantly comorbid mood, anxiety, and substance dependence and a relatively pure low-comorbidity PTSD class [GL13]. These should be further investigated to improve diagnosis and treatment [GL13].

Another *ML* application (patent US7805396B2 by Wagner and Martin [A]), proposed a method to classify normal and abnormal diurnal cortisol secretion patterns with only two saliva samples per day. The computerized method they used is based on a fuzzy logic algorithm.

Similar datasets as the one used in this thesis already brought up some new findings on how physiology influences the acute stress response. For instance, the biological sex influences the cortisol response, according to a study of Kirschbaum et al., the average cortisol level is two times higher among men compared to women [Kir92a]. McInnis et al. further discovered that subjects with adipositas show higher *IL-6* responses and a less efficient *HPA-Axis* habituation [McI14]. Smoking leads to an increased overall cortisol level, compared to non-smokers with equal environmental stimuli [Kir92b]. Similary, glucose intake leads to a stronger cortisol reaction and low glucose blood levels prevent the stress response through the *HPA-Axis*, while protein and fat intakes show no influence [GB02]. Another interesting work of La Marka et al. has shown that the 'Cold Face Test' has an influence on the stress response [Mar11]. For the 'Cold Face Test', the vagus nerve, which is an essential part of the parasympathetic nervous system, is stimulated with cold by applying a cold pack onto a subjects face. They found that a faster response to this stimulus is associated with a lower stress response, expressed in the form of lower cortisol values and an enhanced mood [Mar11]. Finally, there is evidence from the research of Wüst et al. that common polymorphisms in the glucocorticoid receptor gene lead to significantly higher cortisol stress responses [Wüs04]. The term polymorphism is used to describe certain mutations in the genotype.

Besides the physiological influences, several psychological impacts on the stress response were also discovered. The work of Kuras et al. showed that healthy adults with childhood adversity have an increased alpha-amylase response [Kur17]. Another paper suggests that participants with higher self-compassion show lower *IL-6* responses when performing the *TSST* [Bre14]. Furthermore, there is evidence that post-stress rumination predicts *HPA-Axis* responses to repeated acute stress [Gia14]. Post-stress rumination after a first *TSST* was associated with greater cortisol responses both on the intial and the consecutive test, indicating non-habituation to the stressor [Gia14].

A recently published work of Fiksdal et al. implies that symptoms of anxiety are linked to a attenuated cortisol stress response [Fik19]. In addition, they found exaggerated responses for subjects with depression symptoms. All subjects in this study had no psychiatric diagnosis [Fik19].

Another work, which analyzed the correlation between the activity of the *HPA-Axis* and the inflammation response, showed that a stronger *HPA-Axis* habituation is inversely related to inflammatory sensitization [Tho17].

Additionally, there are some evidences that defective stress response patterns lead to chronic diseases. The work of Weaver et al. suggests that a higher *IL-6* level leads to a decline in cognitive functions [Wea02]. Further studies showed that insufficient glucocorticoid signaling can be found

in patients with stress-related neuropsychiatric disorders [Rai03]. Psychological stress might also contribute to atherosclerosis for subjects with a lack of habituation in *IL-6* responses [vK06].

Chapter 4

Methods

At the beginning of this chapter, the dataset used in this thesis is explored. In the next section, the classification process is described including preprocessing, feature extraction, evaluation methods, and classification algorithms. The chapter is concluded by a section describing the statistical approach of finding connections within the patterns of different stress-reactive variables.

4.1 Data Acquisition

The dataset used in this thesis was recorded and kindly provided by Rohleder et al. in the context of a study regarding the relation of age-related diseases and psychosocial stress [Roh09].

4.1.1 Study Design

For the study 100 healthy subjects have been recruited [Roh09]. Table 4.1 shows the mean age, height, and weight of the study population.

Table 4.1: Demographic Information

Age [years]	37.8 ± 18.4
Height [cm]	169.1 ± 11.1
Weight [kg]	73.3 ± 14.7

All subjects have been exposed to the *Trier Social Stress Test (TSST)* on two consecutive days. The *TSST* has proven as a standard procedure for inducing stress in adult human participants [Dic04] and has also been successfully used with elderly subjects [Roh02]. The *TSST*

consists of an anticipatory phase (about 10 minutes) and a test period (10 minutes), where the subjects have to perform free speech and mental arithmetic tasks in front of an audience. The audience normally consists of at least three persons. The free speech part is typically structured like a job interview [Kir93]. Additionally, video and audio recordings of the subject are taken in order to induce more stress and for later evaluation. A typical set-up of the *TSST* with an audience of only two people is shown in Figure 4.1. Along with much information assessed with questionnaires, which are further described in the following sections, blood and saliva samples have been taken at specific times before and after the test. The timepoints can be found in Table 4.2.

Figure 4.1: Set-up of the Trier Social Stress Test [Fri15].

4.1.2 Demographic Information

The recorded demographic information consists of age, gender, ethnicity, race, education, income, and many more. Furthermore, two questionnaires have been used, which are assessing the subjective and the objective social economic status.

4.1.3 Psychological Variables

Different psychological variables have been assessed with questionnaires. The resulting scores include information about depression, vital exhaustion, chronic stress, self-compassion, and

many more. The underlying questionnaires are standard scales in psychological research, such as the Maastricht Vital Exhaustion Scale [App87], the Trier Inventory for Chronic Stress (Short Version) [Sch98] and the Perceived Stress Scale [Coh83].

4.1.4 Health Variables

Different details about the health status of the individual subjects were collected as well. In addition to information about the fitness (BMI, body fat percentage), there is a detailed record of medications as well as infectious and chronic diseases.

4.1.5 Biological Variables

Biological variables include information about cortisol, amylase, and *IL-6* levels. These are the relevant variables for this thesis.

Table 4.2: Biological Variables

	Number of full datasets	Sample type	Time relative to TSST [min]
Cortisol	63	Saliva	-45, -1, +1, +10, +30, +60, +120
Amylase	77	Saliva	-45, -1, +1, +10, +30, +60, +120
IL-6	71	Blood	-1, +30, +120

As depicted in Table 4.2, seven saliva samples per day and subject have been collected for cortisol and amylase. In contrast, only three samples per day and subject were collected for *IL-6*, since *IL-6* levels are declining very slowly after stress exposure. Additionally, less blood samples were acquired due to economic reasons and to keep the level of additional stress induced by blood sampling as low as possible. The *IL-6* concentration in the blood serum was determined with a high-sensitive commercial ELISA technique (R&D Systems) [Roh09].

4.2 Stress Pattern Classification

The first goal of this Bachelor's Thesis was the classification of different stress responder types based on the biological variables described in Table 4.2.

The typical reaction to acute stress on two consecutive days is *habituation* (Figure 4.2) [Kir94]. On the first day, a high increase in cortisol levels can be observed. The peak of the response is approximately 15 minutes after the *TSST*, a return to baseline level is recognized about 2 hours after the stress impulse. On the second day, cortisol levels are only slightly increased and return to baseline quite fast.

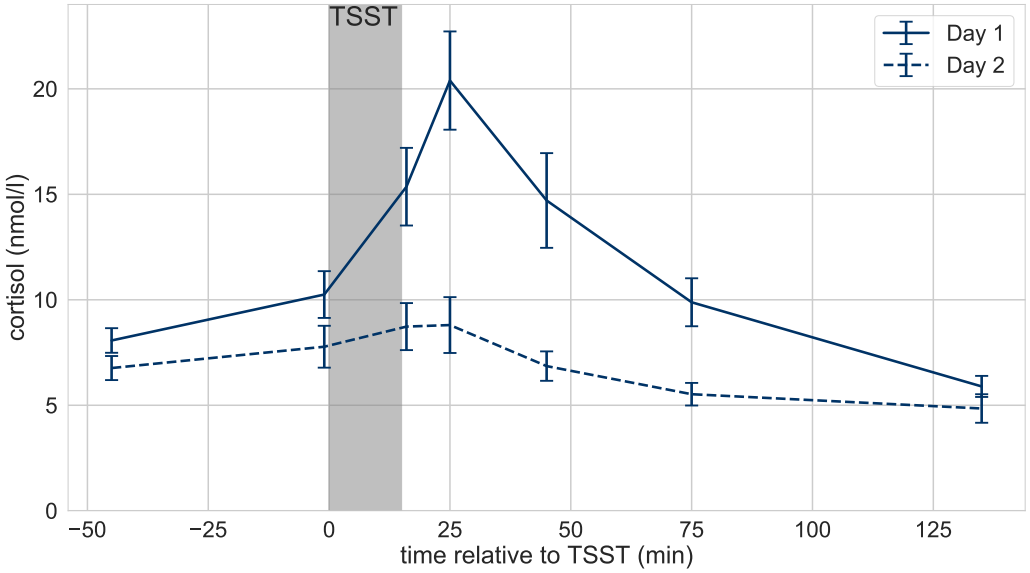


Figure 4.2: Healthy cortisol reaction (habituation) to acute stress on two consecutive days. Mean and standard error for all subjects labeled with 'habituation'.

Class labels for each of the three variables, which were created by an expert, served as ground truth for supervised learning. The classes are derived from previous work of McEwen et al., which suggested five different types of acute stress responses [McE98]. A short description of the classes can be found in Table 4.3. In the responses of *sAA* only the classes 1-4 can be found, for *IL-6* only the classes 1-3 exist. The mean levels of all five cortisol classes on the two consecutive days of the study can be found in Figure B.3.

Table 4.3: Stress responder classes

Class Nr.	Class	Description
1	**Habituation**	High response on the first day, low response on the second day
2	**Non-Habituation**	High responses on both days
3	**Sensitization**	Low or none response on day 1, high response on day 2
4	**Non-Responder**	Low or none response on both days
5	**Anticipatory Stress**	Highest response before the *TSST* on one of the two days

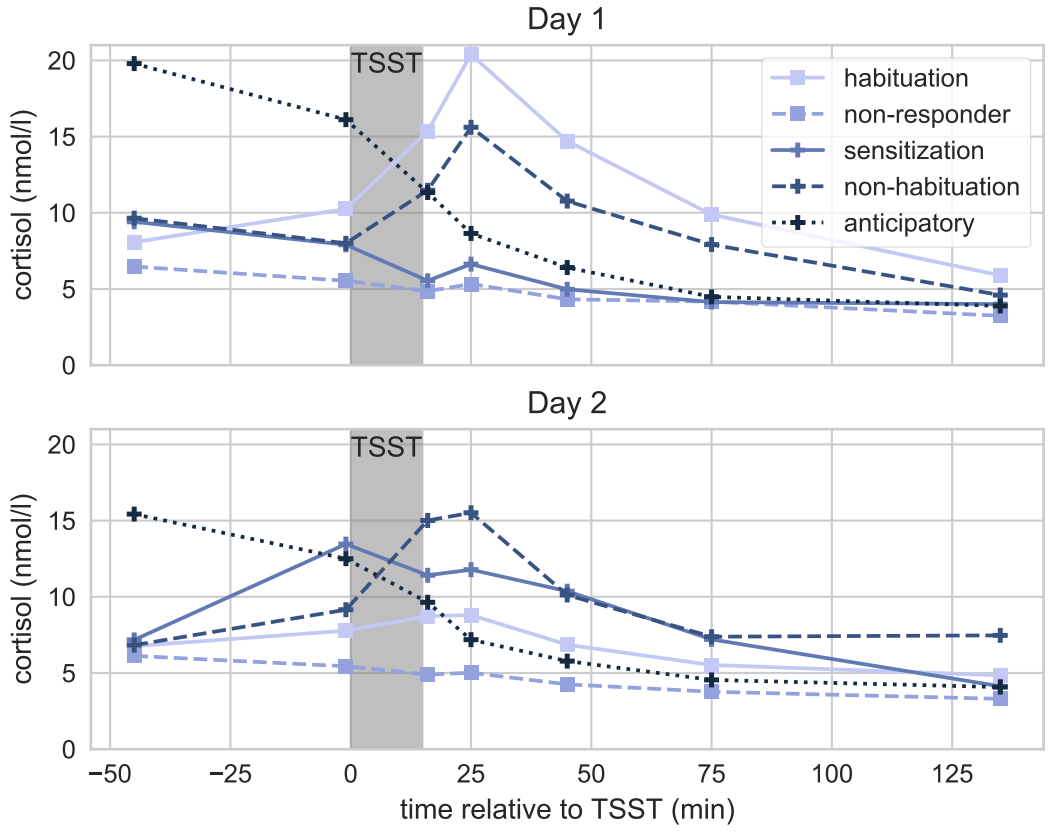


Figure 4.3: Mean cortisol levels for each class. Standard errors for each individual classes are shown in Figure B.3.

An overview of the learning process is outlined in Figure 4.4. After data acquisition, which was not part of this work, the next step was feature extraction and selection. Subsequently, the

resulting feature set was used to train a classifier that tries to discriminate the different classes. The performance of the classifier was then evaluated in the end [Nie83].

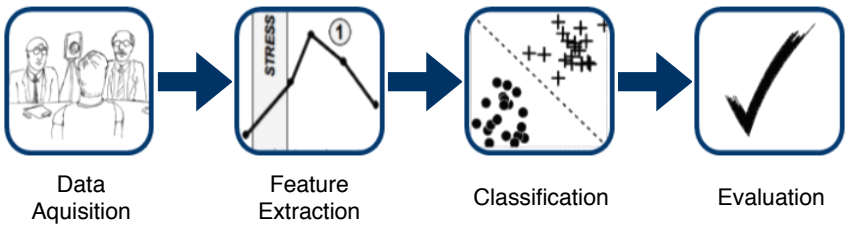


Figure 4.4: General Machine Learning Pipeline

4.2.1 Preprocessing and Feature Extraction

All of the following steps were performed equally for each of the three biological variables. For preprocessing the variables were scaled, such that the highest measured value is 1 and the lowest measured value is 0.

Additionally to the scaled raw data, the features listed in Table 4.4 were calculated. The 'X' markers indicate if the feature was computed for boths days individually and for both days combined or only for one of these options.

Feature extraction in general, can lead to better generalization in the learning process, if the generated features are less sensitive to variance within the classes. The features listed in Table 4.4 were used, since they should replicate the aspects an expert uses for labeling.

Table 4.4: Features extracted from the biological data

Feature	Description	Day1	Day2	Both
Min-Max Difference	Difference bet. minimum and maximum	X	X	X
Max-Max Difference	Difference bet. maxima of the two days			X
Time of Maximum	Location of maximum on the time axis	X	X	
Mean	Mean value (see Eq. 4.1)	X	X	X
Std. Deviation	Standard deviation (see Eq. 4.2)	X	X	X
Skewness	Measure of the asymmetry	X	X	X
Kurtosis	Measure of the 'tailedness'	X	X	X

Mean $\bar{x}$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.1}$$

Standard deviation σ:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{4.2}$$

For feature selection the Select-K-Best algorithm from the Python library *scikit-learn* was used [Ped11]. The *Analysis of Variance (ANOVA)* F-value served as score function. Select-K-Best is reducing the feature set by only selecting the k features with the highest score.

Feature selection is important to reduce the complexity of the classification problem, and can lead to better classification results, because the classifier does not learn features, which do not represent the classes well. Feature selection was performed separately from the pipeline showed in Figure 4.6.

4.2.2 Classification and Evaluation

Cross-validation

An important method, for evaluating the classification perfomance is cross-validation. Cross-validation was used multiple times in this thesis. For example, a 5-Fold cross-validation splits up the dataset 5 times randomly, as visualized in Figure 4.5. For each split, a classifier is fitted to the training set and evaluated on the test set. The data already used as a test set will not be used in further test sets. Every subject is included in the test set exactly once. *Stratified* cross-validation ensures that each fold has the same percentage of samples per class as the overall data. This is important in order to have a training set that has the same properties as the overall data and to prevent the issue of possible folds which might not include samples of all classes in the training set.

Template Matching

The first classification approach used in this thesis was template matching. For this method, templates are generated for each class. The templates should represent the classes well in order to allow reliable classification.

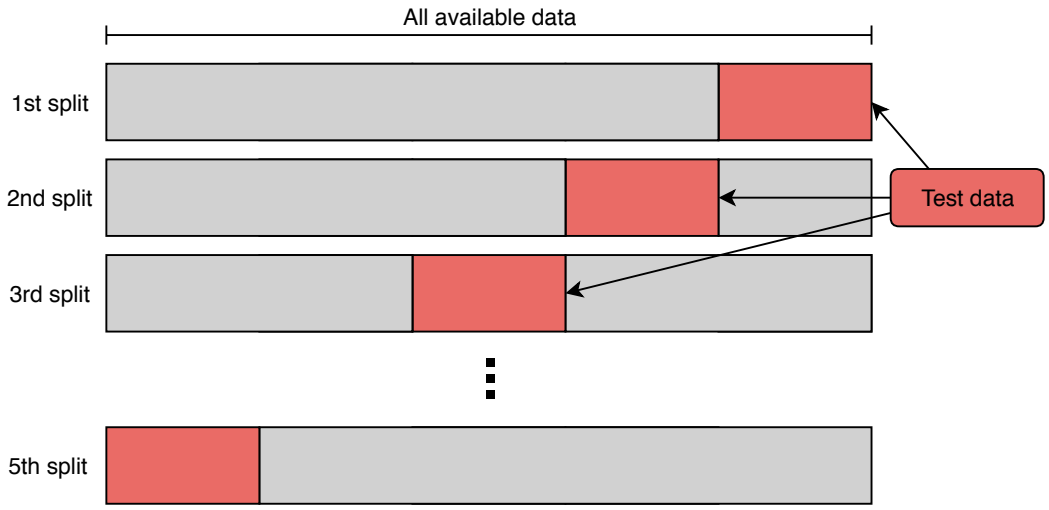


Figure 4.5: Visualization of cross-validation.

In the beginning, the raw biological data was normalized and split up in training and a test set using Stratified-5-Fold cross-validation. The templates of each class were then generated by calculating the mean values of the separate classes in the test set. Therefore, new templates for each class were generated in every cross-validation fold. For the classification of one sample the correlation between the sample and each of the class templates was computed. New samples were then assigned to the class with the highest correlation coefficient. In the end, the accuracy was computed by counting the number of correctly classified samples, divided through the number of samples.

Listing 4.1: Template matching method

```
import numpy as np
from sklearn.model_selection import StratifiedKFold
from scipy.stats import pearsonr

# set up of the cross-validation
n_splits = 5
skf = StratifiedKFold(n_splits)
def template_matching(data, classes, classnames):
    # get number of classes
    n_cla = len(classnames)
    # variable for calculating the accuracy
```

```
12      mean_acc = 0
13
14      # cross-validation split
15      for train_index, test_index in skf.split(data, classes):
16          train, test, class_train, class_test = data[train_index], data
                [test_index], classes[train_index], classes[test_index]
17          # generate the templates
18          templates = get_all_templates(train, class_train, n_cla)
19          # compute correlations
20          corr = np.empty((n_cla,len(test)))
21          for m in range(n_cla):
22              for n in range(len(test)):
23                  corr[m,n] = pearsonr(test[n,:], templates.iloc[:,m])
                        [0]
24          # find highest correlation
25          pred = corr.argmax(axis=0)
26
27          acc = 0
28          for j in range(len(pred)):
29              # check for true labels
30              if pred[j] == class_test[j]:
31                  acc += 1
32          acc /= len(pred)
33          mean_acc += acc
34      mean_acc /= n_splits
35      return mean_acc
```

For the next approach, the pipeline shown in Figure 4.6 was developed, which is described in detail in the following paragraphs.

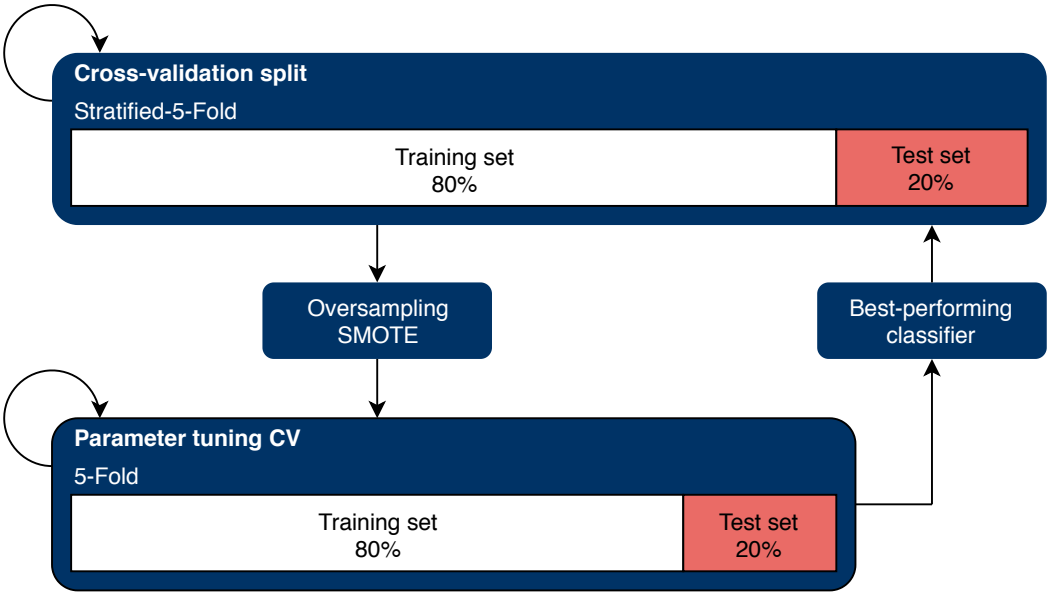


Figure 4.6: Flowchart of the classification pipeline. *CV*: *Cross-validation*

In the first step, the dataset was split up into a training and a test set, called cross-validation split in Figure 4.6. A typical size, which was also chosen in this thesis, is 80% of all subjects for the training set and the remaining 20% for the test set. The split was performed using a Stratified-5-Fold cross-validation.

Oversampling

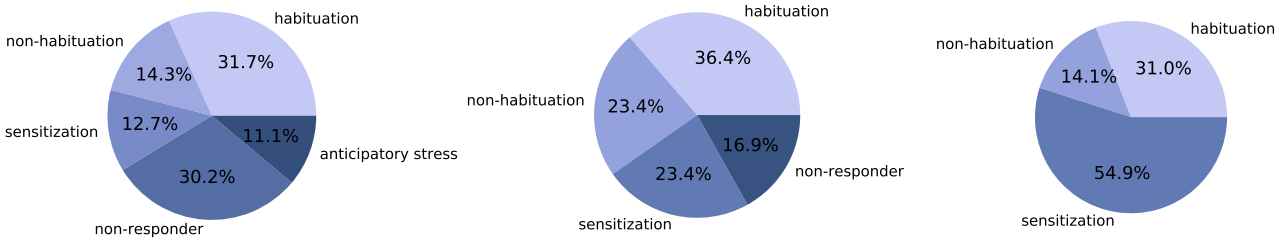
As shown in Figure 4.7, the classes in the dataset are highly unbalanced, which is not only the case for the shown cortisol classes, but also for *sAA* and *IL-6* classes. Unbalanced classes have a bad influence on the performance of the classifier, because the classifier tends to learn only the majority class. Therefore, oversampling was used to balance the classes. The oversampling technique of choice in this work is *SMOTE* [Cha02], used from the *imbalanced-learn* Python module [Lem17]. Oversampling is performed by selecting each sample from the minority class and adding synthetic samples, which are generated by randomly interpolating between one sample and one of the k-nearest neighbors [Cha02].

Mathematically, this can be described as:

$$x_{new} = x_i + \lambda(x_{zi} - x_i) \tag{4.3}$$

where x_{new} is the generated, synthetic sample, x_{zi} is one of the k-nearest neighbor of the sample x_i and λ is a random factor in the range $[0, 1]$.

For this work $k = 3$ nearest neighbors were used for *SMOTE*. The advantage of *SMOTE* compared to other oversampling techniques, like random oversampling, is that it creates a larger and less specific decision region, which leads to better generalization of the resulting classifier. [Ped11].



(a) Cortisol **(b)** Amylase **(c)** IL-6

Figure 4.7: Distribution of stress responder classes

As shown in Figure 4.6, the oversampled training data was split up again in the next step, using standard 5-Fold cross-validation. This cross-validation fold is relevant for parameter tuning. A grid search was applied for finding the optimal parameters for each classifier. Grid search is the technique of systematically trying out all parameter combinations in a predefined parameter space. Because the classes are balanced from the previously performed oversampling, there is no need for stratified cross-validation.

In the next step, the best performing classifier was then applied on the test set from the first cross-validation split, resulting in five test scores, one for each fold. Finally, the mean accuracy, which is the most important measure for the classifying performance, was computed as the mean over all test scores.

With this pipeline it is ensured that no test data 'leaks' into the training set, which would lead to over-optimistic results caused by overfitting [Nas07]. A short extract from the classification pipeline implemented in Python, using only *SVM* classification, is shown in Listing 4.2.

Classifiers

For the classification task three different classifiers were used. All classifiers, which were applied in this thesis, are provided by the *scikit-learn* Python library [Ped11].

The first classifier is the k-Nearest-Neighbors classifier. This classifier uses non-generalizing learning. It does not construct a general internal model, but simply stores instances of training data [Ped11].

The classification is done by simply applying a majority vote of the k nearest neighbors of the sample, where k is an integer value. For basic *kNN* classification, 'nearest' is specified as the smallest euclidean distance. In Figure 4.8 an example for *5 NN* Classification is shown for two features and two classes. The blue and green dots represent training instances of the two classes, whereas the red dot represents a new test sample. The five nearest neighbors of the test sample are denoted by a red outline. Four of them are blue and one is green, therefore the new sample would be assigned to the 'blue' class. In order to find the optimal k for the given dataset a grid search needs to be performed. In this thesis, the parameter space for the grid search was $k \in [1, 20]$.

Figure 4.8: Visualization of *kNN* classification

Another type of classifier that was applied were *Support Vector Machines*. A *SVM* constructs a hyper-plane in the multi-dimensional feature space which separates the samples of the different classes. The best separation is achieved by the hyper-plane that has the largest distance to the nearest training data points of any class. This distance is called *functional margin*. In general, a higher margin results in a lower generalization error of the classifier. A visual representation of

this can be found in Figure 4.9, where the functional margin is the distance between the dotted lines. In this simple example, the feature space is only two-dimensional, therefore the hyper-plane is a one-dimensional line. With larger feature spaces, the number of dimensions increases.

Because of the multi-class character of the data, the 'one-against-one' approach was used. 'One-against-one' is describing the method of training one binary SVM for each pair of classes to seperate the data [Chi02]. Two different kernel functions were evaluated in the grid search, Linear ($\langle x, x' \rangle$, where $\langle\rangle$ denotes the scalar product) and *Radial Basis Function (RBF)* ($\exp(-\gamma||x - x'||^2)$). Another parameter for SVM classification, which needs to be optimized in a grid search, is C. Large C values cause a smaller margin hyper plane, if that plane classifies more training points correctly. Therefore, large C's lead to worse generalization. C was varied between the values $[10^0, 10^1, 10^2, 10^3, 10^4]$. For *RBF*, the additional parameter γ was set to 0.01 or 0.001. This results in 15 different parameter combinations that were applied in grid search.

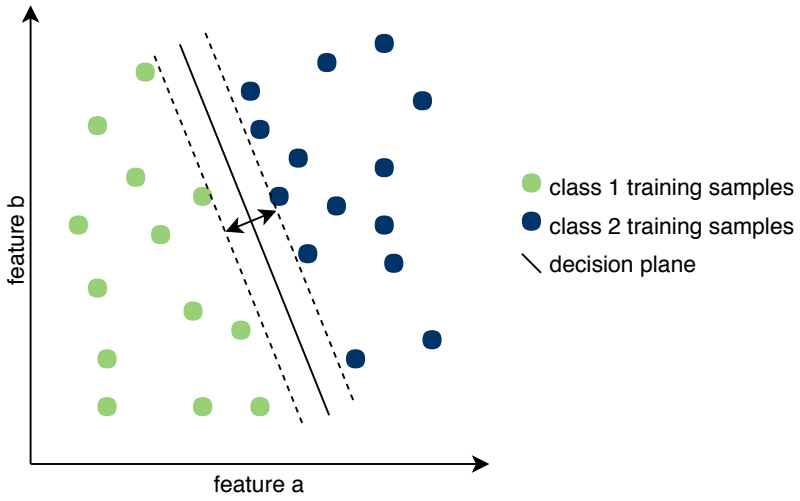


Figure 4.9: Visualization of *SVM* classification

The third classifier applied on the data is the *Random Forest* classifier. A Random Forest classifier constructs a number of *Decision Trees* on sub-samples of the dataset. *Decision Trees* are built by learning simple decision rules inferred from the data. The decision rules for each steps are selected as the variables that best split the given set of items. There are different metrics that define 'best split', such as *Gini impurity*, *Information gain* or *Variance reduction*. For the decision trees used in this thesis Gini impurity was used, which is a measure of how often a element could be incorrectly classified [D'A11]. An example for a simple decision tree can be found in Figure 4.10, a whole decision tree for cortisol classification is shown in Figure B.3.

There are many parameters for Random Forest classification which can be optimized, for example:

- number of trees in the forest
- maximum number of features considered for splitting each node
- maximum number of levels in each decision tree (depth)
- minimum number of data points placed in a node before the node is split

In order to keep computational complexity as low as possible, randomized search was used for parameter optimization. In contrast to grid search, randomized search is using a random parameter set in a predefined number of iterations, instead of systematically trying out every parameter combination. An overview of the parameter space is shown in Table 4.5, detailed description of these parameters can be found in the *scikit-learn* documentation of the `RandomForest` class [Ped11].

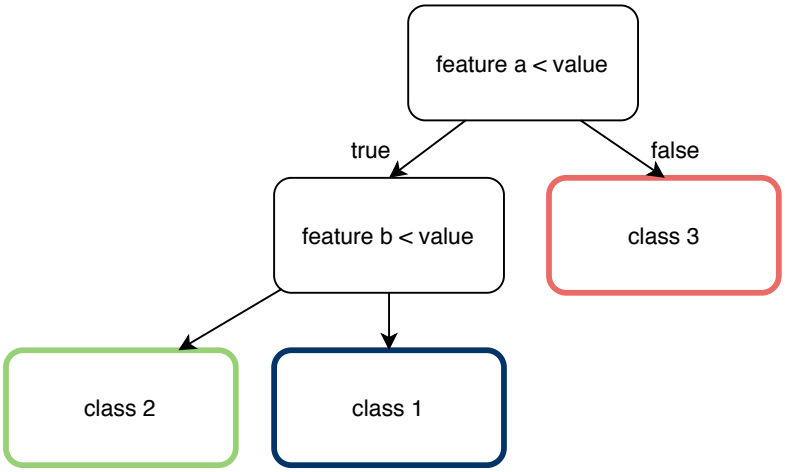


Figure 4.10: Visualization of a *Decision Tree*

Table 4.5: Parameters for *Random Forest* grid search.

Parameter	Possible values
n_estimators	200 to 2000, steps of 10
max_features	[‘auto’, ‘sqrt’]
max_depth	10 to 100, steps of 10
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 2, 4]
bootstrap	[True, False]

Listing 4.2: Extract from classification pipeline

```
from sklearn.model_selection import StratifiedKFold, GridSearchCV
from imblearn.oversampling import SMOTE
# Set-up of SMOTE
smo = SMOTE(k_neighbors = 3)
# Stratified 5 Fold cross-validation split
n_splits = 5
skf = StratifiedKFold(n_splits)
for train_index, test_index in skf.split(data,classes):
    train, test, class_train, class_test = data[train_index], data[
       test_index], classes[train_index], classes[test_index]
    # Oversampling
    train, class_train = smo.fit_resample(train, class_train)
    # Grid search using 5 Fold cross-validation
    gridsvm = GridSearchCV(svm,param_gridsvm, cv = n_splits, iid =
       True, return_train_score = False)
    gridsvm.fit(train, class_train)
    # Evaluation of best-performing classifier on the test data
    scores_svm[i] = gridsvm.best_estimator_.score(test,class_test)
```

4.3 Stress Pattern Correlation

4.3.1 Preprocessing

For this task, only subjects with expert labels for all three biological markers (cortisol, amylase, and IL-6) could be considered, which resulted in a dataset consisting of 71 subjects. This data was then reorganized such that every column only holds the information of one class, so that a '1' denotes that the subject belongs to the respective class, '0' otherwise (as also depicted in Table 4.6). The *ground truth* labels were used for this task.

Table 4.6: Example of preprocessing for pattern correlation. Left: before preprocessing; Right: after preprocessing.

Subject ID	Cortisol class
1	habituation
2	non-habituation
3	sensitization
4	habituation
5	non-responder
6	non-habituation
⋮	⋮

Subject ID	Hab.	Non-hab.	Sen.	Non-resp.
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	1	0	0	0
5	0	0	0	1
6	0	1	0	0
⋮	⋮	⋮	⋮	⋮

4.3.2 Statistics

In order to find correlations between the patterns of the three biological variables, the Pearson correlation coefficient was calculated column-wise. Therefore, one correlation coefficient was computed for every class combination of the five cortisol, four amylase, and three IL-6 classes. The Pearson correlation coefficient r_{xy} is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{n\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.4}$$

where n is sample size, x_i and y_i are the individual sample points and $\bar{x}$ and $\bar{y}$ are the sample means.

Along with the Pearson correlation, the 2-tailed probability value (p-value) was calculated. The 2-tailed p-value can be calculated by:

$$p = 2 \cdot \min\{P(X \leq x \mid H),\ P(X \geq x \mid H)\} \tag{4.5}$$

where X is the random variable, H is the statistical hypothesis and $P()$ is the probability.

The combination of a non-zero Pearson correlation coefficient and a p-value smaller than the significance level reveals statistically significant combinations. The significance level α was set to 1 % ($\alpha = 0.01$), meaning there is a 99 % ($1 - \alpha = 0.99$) chance that the correlations found are not caused by coincidence.

Additionally, the number of class combinations of two variables was counted for all three permutations. This was then normalized to get percentual values.

Chapter 5

Results

All results, which have been collected in this thesis, are presented in this chapter. The first section deals with the results obtained for the stress pattern classification. In the next section, findings from the analysis of the pattern correlations are presented.

5.1 Stress Pattern Classification

Template Matching

With the template matching approach the achieved classification accuracies were 57.6 % for cortisol, 54.9 % for amylase and 30.9 % for IL-6.

Feature Selection

Feature selection showed that all additional features, that are listed in Table 4.4, performed worse than raw samples. When applying the *K-Best* feature selection for $k \in [1, 14]$, the remaining features have always purely consisted of the raw samples. As shown for cortisol in Table 5.1, samples achieved higher scores than the additional features computed on the raw samples. Similar results were obtained for the two other variables *sAA* and *IL-6*. Neither choosing more features than the samples, nor choosing less could improve the classification performance. Therefore, the following classification results are all based on the 14 (6 for *IL-6*) samples of the biological markers.

Table 5.1: Feature selection results for cortisol

Feature	Score (*ANOVA* f-value)	Feature	Score (*ANOVA* f-value)
Sample 1	14.5	**Min-Max Day 1**	4.3
Sample 2	6.5	**Min-Max Day 2**	3.6
Sample 3	8.8	**Max-Max**	4.4
Sample 4	11.8	**locMax Day 1**	3.6
Sample 5	6.5	**locMax Day 2**	4.2
Sample 6	8.0	**Mean Day 1**	2.3
Sample 7	5.5	**Mean Day 2**	2.5
Sample 8	11.3	**Mean All**	1.2
Sample 9	4.8	**Std. Dev. Day 1**	2.5
Sample 10	4.9	**Std. Dev. Day 2**	1.1
Sample 11	5.5	**Std. Dev. All**	2.3
Sample 12	4.4	**Skewness Day 1**	1.1
Sample 13	4.5	**Skewness Day 2**	0.7
Sample 14	9.6	**Skewness All**	0.9
		Kurtosis Day 1	1.1
		Kurtosis Day 2	0.8
		Kurtosis All	0.9

Classification

Table 5.2 lists the classification performance of the three classifiers implemented in this thesis. It can be observed that Support Vector Machines performed best throughout all biological variables. Because the highest accuracy scores were observed with *SVMs*, the parameter optimiza-

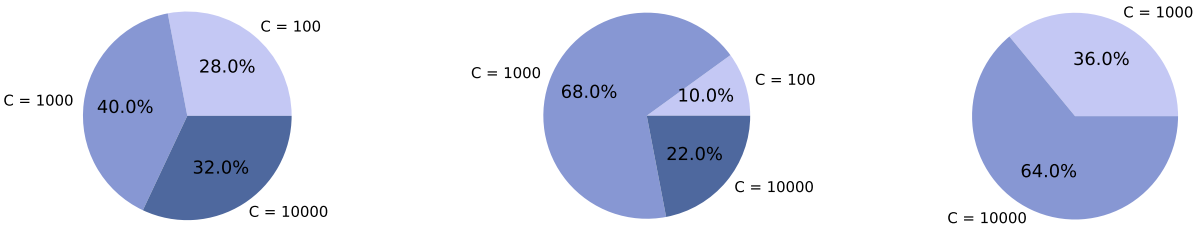
Table 5.2: Accuracies of different classifiers in % (Mean $\pm$ Standard Deviation). The best performing classifier is highlighted in *italic*.

Classifier	*SVM*	k-NN	Random Forest
Cortisol	*78.1 $\pm$ 11.2*	64.6 ± 12.2	66.2 ± 11.0
Amylase	*61.2 $\pm$ 12.3*	45.4 ± 10.1	44.8 ± 9.9
IL-6	*91.2 $\pm$ 6.3*	84.3 ± 9.2	77.4 ± 8.9

Table 5.3: Parameter optimization for cortisol response *SVM* classification. Example from a random cross-validation split. The best parameter combination is denoted in *italic*.

Kernel	C	γ	Mean test score (%)
Linear	1	-	48.9
	10	-	61.4
	100	-	86.4
	1000	-	*96.8*
	10000	-	95.8
RBF	1	0.01	48.9
		0.001	48.9
	10	0.01	48.9
		0.001	48.9
	100	0.01	48.9
		0.001	48.9
	1000	0.01	50.0
		0.001	48.9
	10000	0.01	72.9
		0.001	50.0

tion is only visualized for this type of classifier. The example in Table 5.3 shows that Linear *SVMs* with $C = 100$, $C = 1000$ and $C = 10000$ all have high test scores. This observation holds for every cross-validation fold. Therefore, one of these three parameter combinations was observed as best-performing in all cases. In Figure 5.1 the percentual distribution of the C parameter is shown for each variable, averaged over 10 iterations.

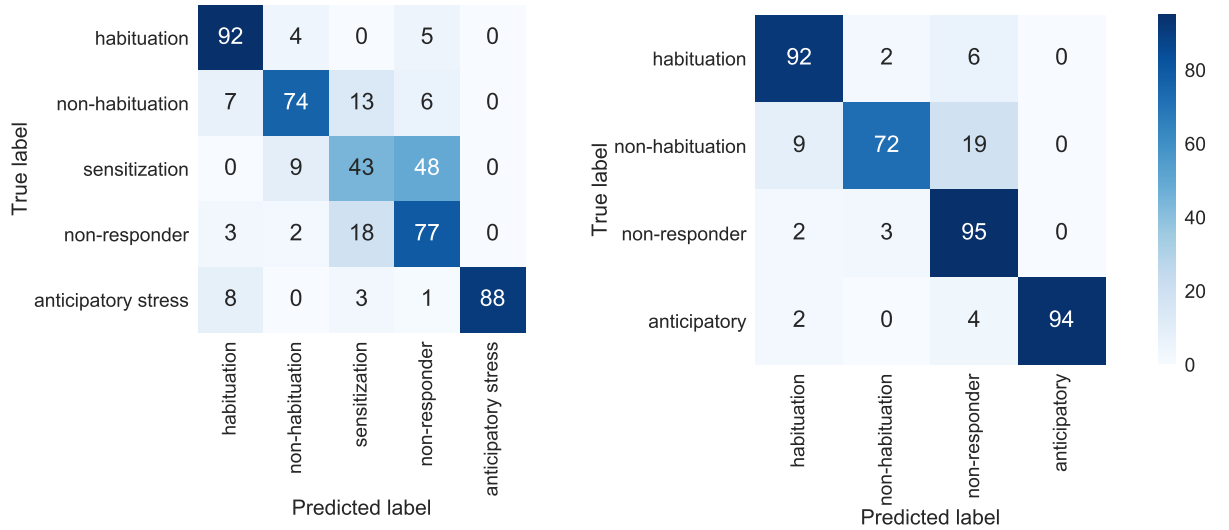


(a) Cortisol **(b)** Amylase **(c)** IL-6

Figure 5.1: *SVM* classification parameter distribution.

Figure 5.2 (a) shows the confusion matrix for all five cortisol classes. The mean accuracy for 5 classes is $78.1\ \% \pm 11.2\ \%$. It is visible that the 'habituation' and 'anticipatory stress' classes achieved the highest accuracies with 92 and 88 %, whereas the groups 'non-habituation' and

'sensitization' have an accuracy of 74 % and 77 %, respectively. For 'sensitization' the accuracy is at 43%, which means that the algorithm classified the test samples more often wrongly than correctly. Almost all of the wrongly labeled 'sensitization' samples (about 84 %) were assigned to the 'non-responder' group. If the classes 'sensitization' and 'non-responder' are combined into a joint 'non-responder' class, as displayed in Figure 5.2 (b), a mean accuracy of $92.2\ \% \pm 9.7\ \%$ is achieved. Why this is reasonable is explained in the next chapter (6.1).



(a) 5 classes, Mean Accuracy: $78.1\ \% \pm 11.2\ \%$

(b) 4 classes, Mean Accuracy: $92.2\ \% \pm 9.7\ \%$, 'non-responder' and 'sensitization' combined to 'non-responder' class

Figure 5.2: Confusion matrix for cortisol classification in %.

The overall mean accuracy for amylase classification was $61.2\ \% \pm 12.3\ \%$. As depicted in Figure 5.3, the mean accuracies for the amylase classification was 75 % for 'habituation', 55 % for 'sensitization', 60 % for 'non-responder', and 43 % for 'non-habituation'. The 'non-habituation' class shows a lot of scattering. Samples of this class were often misclassified (66 %) and samples of other classes were incorrectly assigned to this class multiple times. For example, this was the case for 30 % of the subjects from the 'non-responder' class.

For *IL-6* classification the confusion matrix (visualized in Figure 5.4) shows the accuracy for the individual classes with 99 % for 'habituation', 71 % for 'non-habituation', and 89 % for 'sensitization'. The mean accuracy for IL-6 responder type labeling is at $91.2\ \% \pm 6.3\ \%$.

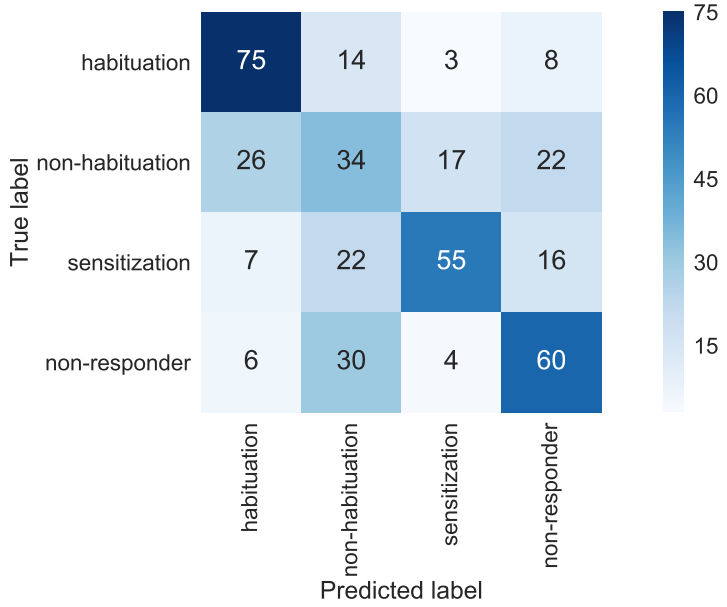


Figure 5.3: Confusion matrix for *sAA* classification in %. Mean accuracy: $61.2\,\% \pm 12.3\,\%$

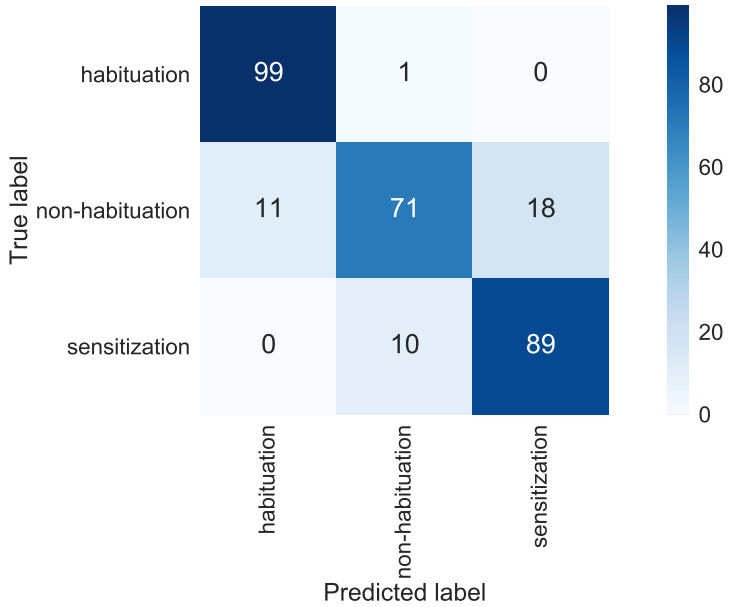


Figure 5.4: Confusion matrix for IL-6 classification in %. Mean accuracy: $91.2\,\% \pm 6.3\,\%$

5.2 Stress Pattern Correlation

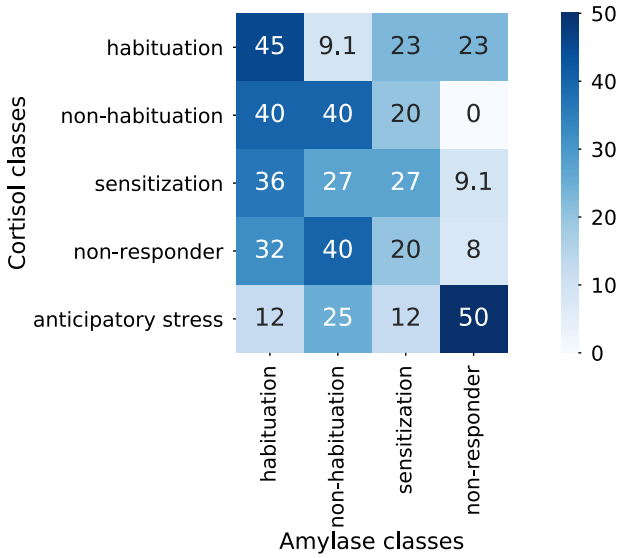
Results for finding connections within the patterns of different stress-reactive variables are shown with their correlation coefficients and p-values in Figures B.1 and B.2. For a significance level of $\alpha = 0.01$, results revealed the following five significant cominations:

- 'amylase non-habituation' – 'cortisol non-responder' ($r = 0.28$, $p = 0.004$)
- 'amylase non-responder' – 'cortisol anticipatory stress' ($r = 0.27$, $p = 0.006$)
- 'il6 non-habituation' – 'cortisol non-responder' ($r = 0.25$, $p = 0.01$)
- 'il6 sensitization' – 'cortisol anticipatory stress' ($r = 0.26$, $p = 0.007$)
- 'il6 sensitization' – 'amylase non-responder' ($r = 0.25$, $p = 0.008$)

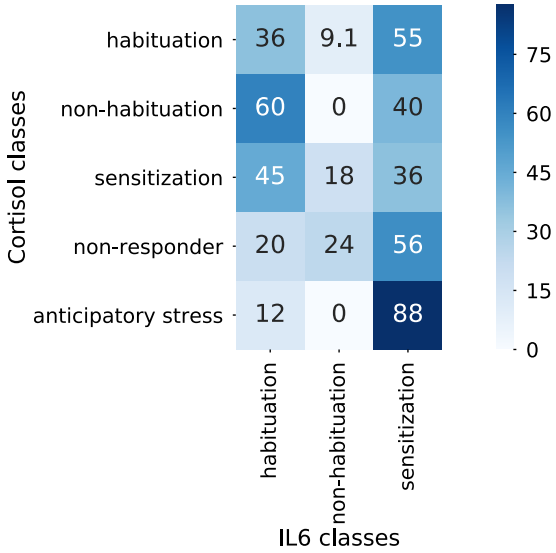
The remaining combinations with $p \leq 0.01$ have negative correlations. The combinations were all found where both classes are within one biological variable, for example 'cortisol non-responder' – 'cortisol habituation' ($r = -0.34$, $p < 0.001$). These are not relevant because every subject only has exactly one class for each variable. Hence, the combinations within one variable will always show strong negative correlations.

Similar connections can be found again in the prevalence of class combinations:

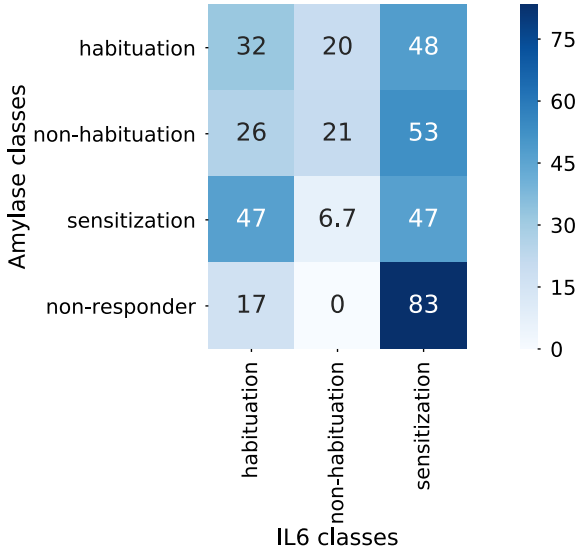
- 40 % of cortisol non-responders were also in the 'amylase non-habituation class' (Figure 5.5 (a))
- 50 % of the subjects which show anticipatory stress in cortisol are amylase non-responders (Figure 5.5 (a))
- 88 % of the cortisol anticipatory stress group are also in the IL-6 'sensitization' class (Figure 5.5 (b))
- 83 % of amylase non-responders show IL-6 sensitization (Figure 5.5 (c))



(a) Cortisol and amylase. Normalized on cortisol.



(b) Cortisol and IL-6. Normalized on cortisol.



(c) Amylase and IL-6. Normalized on amylase.

Figure 5.5: Prevalence of class combinations (%).

Chapter 6

Discussion

6.1 Stress Pattern Recognition

Results showed that the classification based on the template matching approach proved to be less precise than expected with accuracies for cortisol and amylase of under 60 %. For IL-6 the accuracy is even worse with about 30 %. For three classes, this result is not better than random guessing. A possible explanation why template matching performs relatively bad is the high variation within the classes. Therefore, misclassification due to a wrong template having the highest correlation with the test samples seems to happen relatively often. A more refined template generation approach – instead of only computing the mean values over all training samples as class templates – might lead to better results. Furthermore, using weighted samples for template matching could further increase classification accuracy.

The feature selection revealed that all additionally computed features were not useful for the aspired classification. This suggests that the calculated features do not separate the classes well. Additionally, some of the features are dependent. For example the Minimum-Maximum difference is an approximation of the standard deviation. The conclusion that the additionally computed features do not represent the aspects an experts uses for labeling could be drawn.

As described in Section 5.1, the best classification results were achieved when only considering raw samples. When feature selection was performed on the raw samples, the reduction of samples lead to a decrease in classification performance. Therefore, a higher density of sampling points might improve further improve classification performance. However, this would cause additional costs for analysis of saliva probes and might disrupt the whole study procedure, leading to altered stress responses.

In related work, subjects with high cortisol baseline often have to be excluded since subjects with high baseline are sometimes not able to show a stress reaction [Kud04]. In this thesis these high baseline samples did not need to be excluded because no considerable changes in accuracy were observed. On the one hand this compensated by the 'anticipatory stress' class which includes many of the high baseline subjects, and on the other hand leads to the possible conclusion that the classification process is robust enough to classify these correctly.

As Figure 5.2 reveals the mean accuracy considerably increased (78 % vs. 92 %) when combining two cortisol response classes into a joint class. Of course reducing the number of classes generally makes the classification task easier. Nevertheless, another reason is that data from subjects belonging to those two classes ('sensitization' and 'non-responder', respectively) appear to look very similar. During the first day, subjects of both classes showed a weak or no cortisol response, which is correct for both classes according to their definition (Table 4.3). However, on the second day many subjects from the 'sensitization' group, which should have a considerably higher response, showed a rather weak response (Figure 6.1). In consultation with the expert, it became clear that the definition for differentiating both classes was not clear enough for reliable labeling. Hence, it is plausible to merge to two classes.

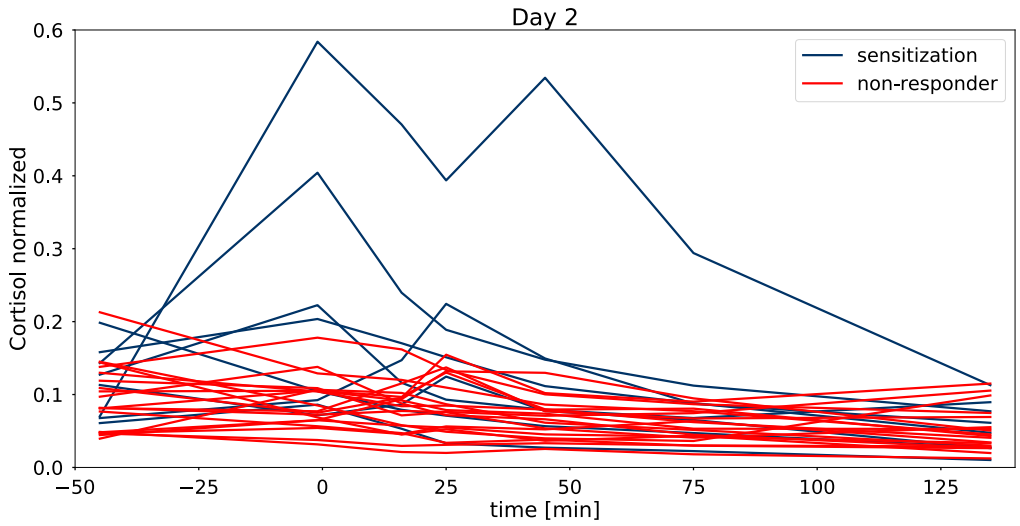


Figure 6.1: Comparison of all raw samples from day 2 of the cortisol classes 'sensitization' and 'non-responder'. Cortisol normalized on highest measured value.

Non-reliable labeling, for example due to unclear class definitions, leads to uncertainty in the ground truth data. In order to avoid this, a sharper definition of the classes is suggested. Additionally, multiple experts that independently label the same dataset would be helpful, since results obtained from machine learning can never be more accurate than the ground truth data.

For the amylase response classification, the confusion matrix (Figure 5.3) shows that the 'non-habituation' class was wrongly classified several times. Also, subjects from other classes were often wrongly assigned to this class. For example, 37 % of the subjects from the amylase non-responders were misclassified to 'non-habituation'. The are various possible reasons for this. For instance, unclear class definitions can, analogous to cortisol classification, lead to misclassifications. Besides errors caused by imprecise *ground truth* data, there are huge variations in the stress response values within the same class. This is proven by the high intra-class standard error (Figures 4.2 and B.3).

Additionally, the number of subjects per class is relatively small. For example, there are only seven subjects from the cortisol 'anticipatory stress' class (Figure **??**). For a more robust classifier more data would need to be recorded, because learning tasks are highly depending on large databases. Oversampling was used instead of undersampling, so that the amount of data is not further reduced.

In contrast, IL-6 pattern classification showed a high accuracy for all classes. This suggests that the patterns were more precisely defined by the expert and therefore, a clearer class differentiation was possible. Furthermore, results also show that three samples per day and subject are enough for automatic classification. From the three different patterns, non-habituation performed worst, which is the minority class with only about 14 % of the samples (10 of 71 subjects). Therefore, it is possible that the classifier tends to learn the majority classes better.

Another problem, which occurred in this work was that the results varied if the classification process is was repeated multiple times on the same dataset. The mean accuracy differed within a range of $\pm$ 2 %. Furthermore, the standard deviation is relatively high with about 10 % for cortisol classfication and even over 12 % for amylase. This can be caused by 'outliers' in the test or the training set, which will often be classified wrong. For instance, if one subject shows typical signs of two or more classes, it could either cause the classifier to learn signs which are not typical for this class in the training process, or it could be misclassified because the rules learned from other samples do not apply for this 'outlier'. Detecting those outliers would improve classification performance. However, this is only feasible with a bigger dataset. Another possible reason for these variations, could be the small size of the dataset. Therefore, the cross-validation splits vary a lot between each repetition.

The parameter optimization showed that increasing the C parameter also increased accuracy. However, as described in Section 4.2 *SVMs* with a higher C parameter tend to generalize less. Therefore, it is possible that the classifier is only trained well for this dataset, and that classification on another dataset will not be very accurate. However, it was attempted to avoid such overfitting with cross-validation. Additionally, samples of different classes can be found very close to each other in the parameter space. Hence, a smaller margin hyperplane(corresponding to a higher C value) managed to separate the different classes better.

Figure B.3 reveals some interesting insights how decision trees performed the 'best splits' in order to separate classes as well as possible. In the first node, the decision tree is capable of separating the set into two ('sensitization' and 'non-responder') and the remaining three classes quite well. This first decision is only based on sample number 3, which corresponds to the sample 15 minutes after the *TSST* on day 1. Cortisol reaches the peak values at this point for the normally expected 'habituation' response. In the next decision tree layer, 12 of 16 'anticipatory stress' samples are classified based on a high first sample, which, based on the class description, would be considered as "intuitive decision". This leads to the conclusion that the decision tree classifier performs a quite "human-like" classification of stress response patterns.

However, there are some issues in the medical context, which need to be considered. For instance, the IL-6 concentration in the serum is quite unspecific. IL-6 is only an indirect indicator of stress, and primarily in indicator of inflammation [McI14]. Therefore, non-neglectable percentage of the measured IL-6 concentration can also originate from other inflammation related processes. Furthermore, the cortisol and amylase samples collected from saliva highly depend on the used analysis device and method, and that a change in the analysis pipeline would lead to baseline shifts or changes in amplitude scaling. Therefore, that the classifier trained in this work would need to be retrained using additional data in order to increase generalizability of the underlying classification model.

For future work, and for application in a real-world scenario, it is proposed to remove the outer cross-validation step. After oversampling the minority classes with *SMOTE*, the learning algorithm can be trained on the whole dataset for the best results. Because Linear *SVMs* performed best in all scenarios, the optimization process can be limited to the parameter C. The results presented in this work suggest to vary C in the range $[10^2, 10^3, 10^4]$.

Another limitation of the approach used in this thesis is that the classification will probably not generalize well on other datasets using different time points relative to the stressor for collecting saliva and blood samples or a different number of samples. In order to still allow classification, samples from the new dataset would need to be mapped to the required form, for instance

by interpolating sample values. Another possibility would be retraining the classifier using information from the new dataset.

6.2 Stress Pattern Correlation

Some of the results gained from the analysis of pattern correlation could lead to new findings from a psychophysiological point of view. They suggest a connection between the three classes 'cortisol anticipatory stress', 'amylase non-responder' and 'IL-6 sensitization'. A possible interpretation is that the combination of a malfunctioning *SNS* stress reaction together with a high *HPA-Axis* response before stress exposure results in a higher inflammatory response. This could extend the results from previous work that a stronger *HPA-Axis* habituation is inversely related to inflammatory sensitization [Tho17].

The correlation between 'amylase non-habituation' and 'cortisol non-responder' subjects could indicate that a non-adaptive *SNS* reaction inhibits the cortisol response, or that, in contrast, the missing *HPA-Axis* response might lead to non-habituation in the *SNS*.

Furthermore, the correlation of 'cortisol non-responder' and 'IL-6 non-habituation' subjects could reveal that a missing *HPA-Axis* response leads to worse adaption of inflammatory processes to repeated stress.

The statistical methods applied in this work were relatively simple. Therefore, further statistical analysis could bring up even more inter-pattern linkages.

Additionally, the problem of the rather small study population holds for to this task as well. Hence, results should be only seen as first hints how the stress-reactive parameters might be related. The connections have to be further investigated in a larger study to verify these first findings and potentially draw more definitive conclusions. In order to achieve this goal, the automatic classification approach developed in this thesis could contribute.

Chapter 7

Conclusion and Outlook

The two research goals of this thesis were the classification and prediction of different stress responder types, and the analysis of correlations between these patterns. For the first task, a method capable of classifying four different cortisol responder types with high accuracy ($92.2\ \% \pm 9.7\ \%$) was introduced. Additionally, three different IL-6 classes can be discriminated with high accuracy of $91.2\ \% \pm 6.3\%$. Unfortunately, the achieved accuracy for amylase is a bit lower with about 60 %. In conclusion, the cortisol and IL-6 classification could be easily adapted to a recommender system for faster and easier labeling of new data.

Based on further data, which is currently being collected by Rohleder et al., the classifier can be evaluated on a completely new dataset. This can prove, whether the classifier overfits to the dataset used in this thesis or generalizes well.

An interesting approach for further analysis of the data would be the prediction of the stress responder types based on other variables. Especially the prediction based on psychological measures (Section 4.1.3), might have a great potential, because there a various influences on the stress reaction as already described previous work.

Another interesting topic for future work could be the exploration of relations of cortisol response patterns to acute stress with the diurnal variations of cortisol.

The results obtained for the second research question revealed a few new findings of how the three stress-reactive parameters explored in this thesis could be connected. Although some of these correlations have already been found in related studies, it is proposed to further investigate these, in order to understand the interactions better.

Appendix A

Patent

Computerized identification of normal and abnormal diurnal cortisol secretion patterns and levels of testosterone from human saliva samples

Publication Number	US7805396B2
Date of Publication	May 30, 2003
Inventors	Peter Wagner, Lene Martin
Assignee	Diagno-International BV. Schiphol(NL)
Abstract	Normal and abnormal diurnal cortisol secretion patterns are identified from human saliva samples by an in vitro method. First and second saliva samples are taken from one human individual at first and second predetermined times during the same day. A first cortisol concentration is determined in the first saliva sample, and a second cortisol concentration is determined in the second saliva sample. An abnormal secretion pattern is then compared to a normal secretion pattern with the help of a fuzzy logic algorithm. A function of the hypothalamic-pituitary-adrenal (HPA) axis is then determined for the human individual. Optionally, a testosterone level is determined from one of the samples and is used in combination with the cortisol concentrations to provide a redefined determination.

Appendix B

Additional Figures

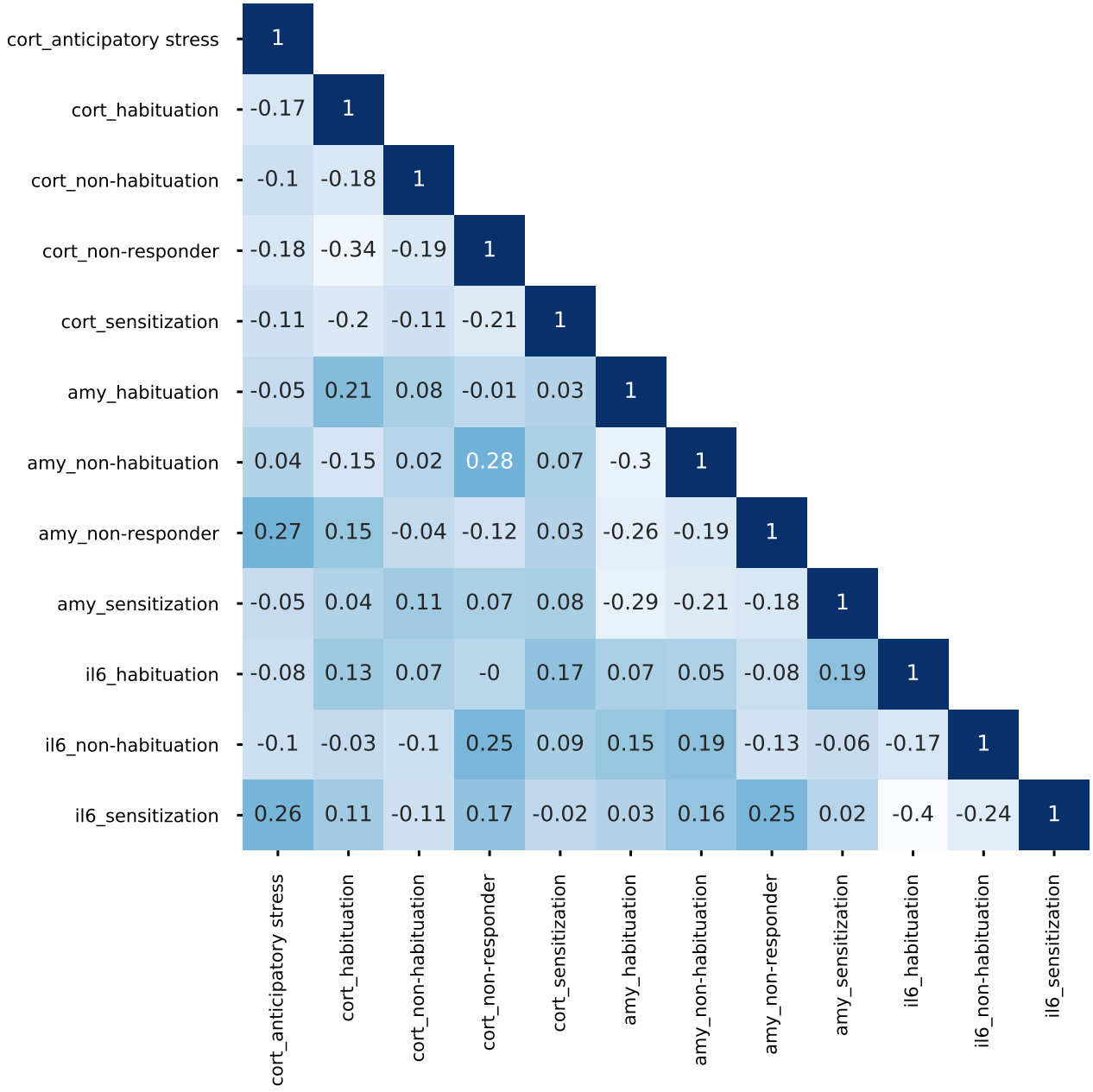


Figure B.1: Correlation coefficients for all class combinations.

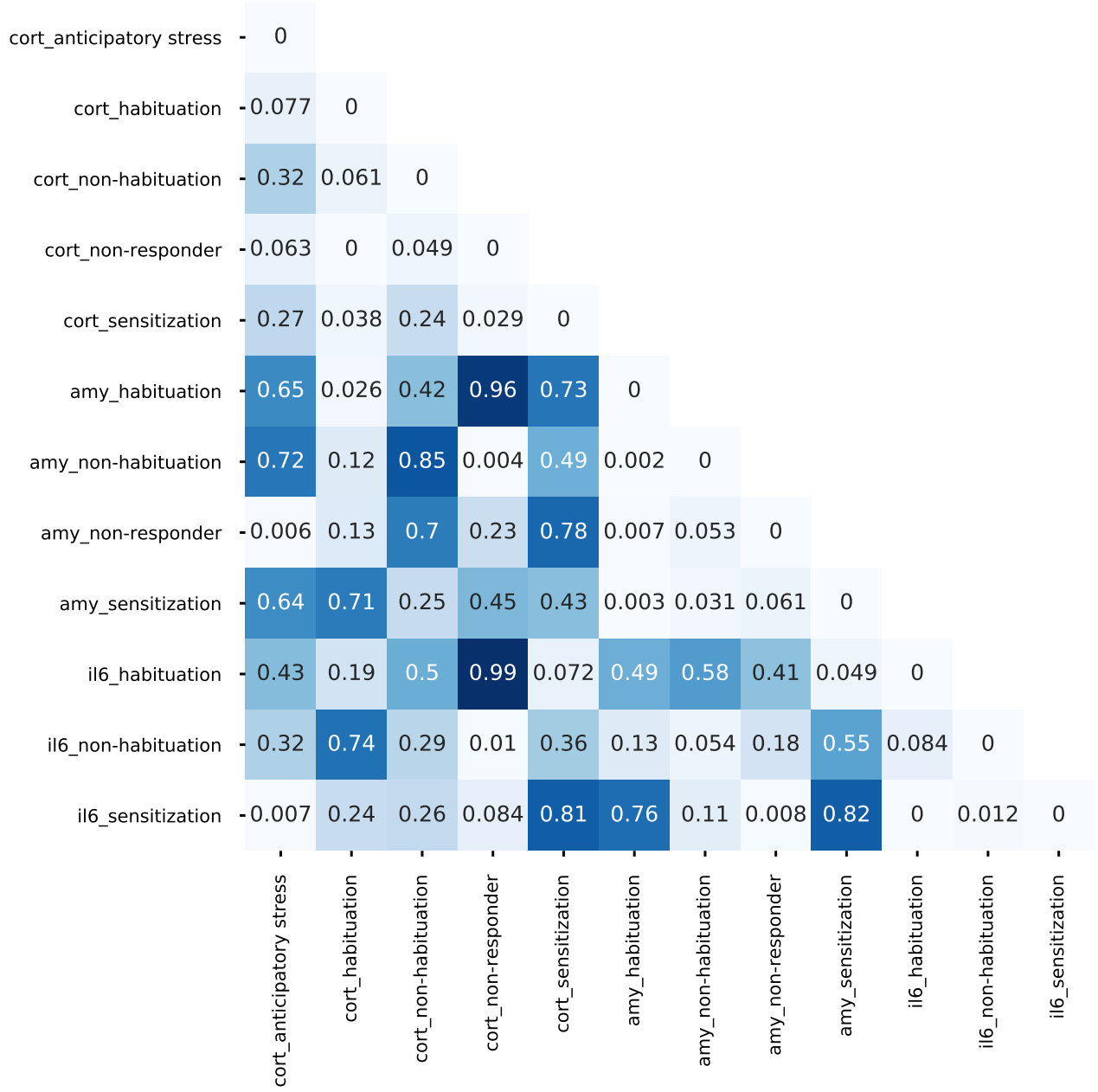


Figure B.2: p-Values for all class combinations.

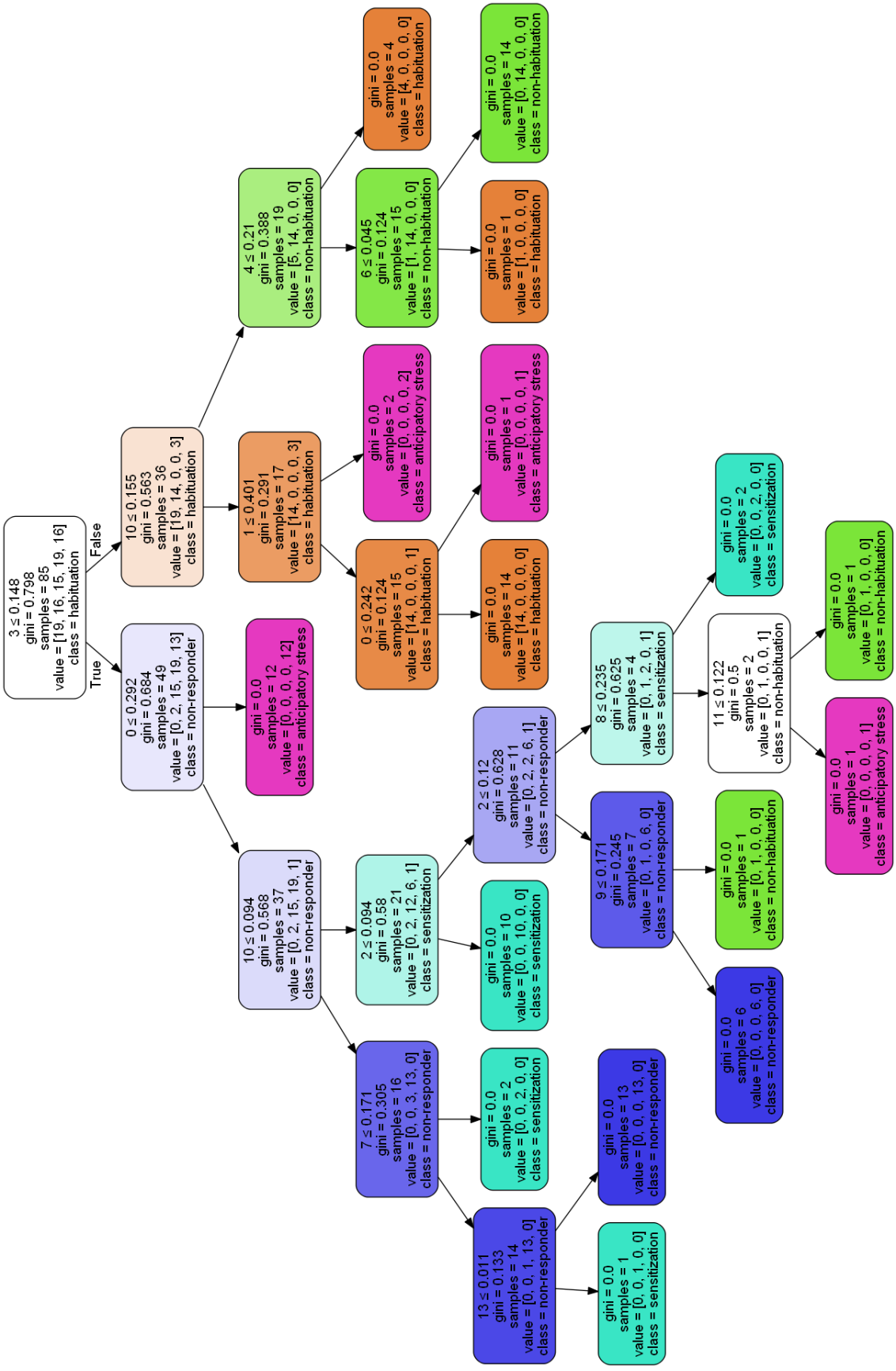
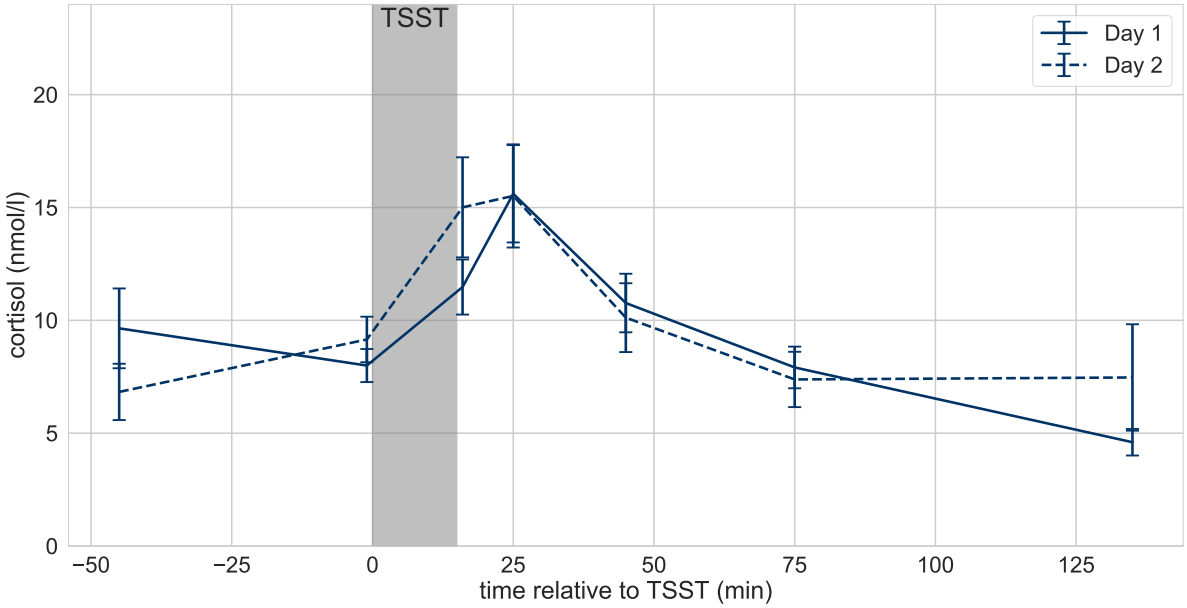
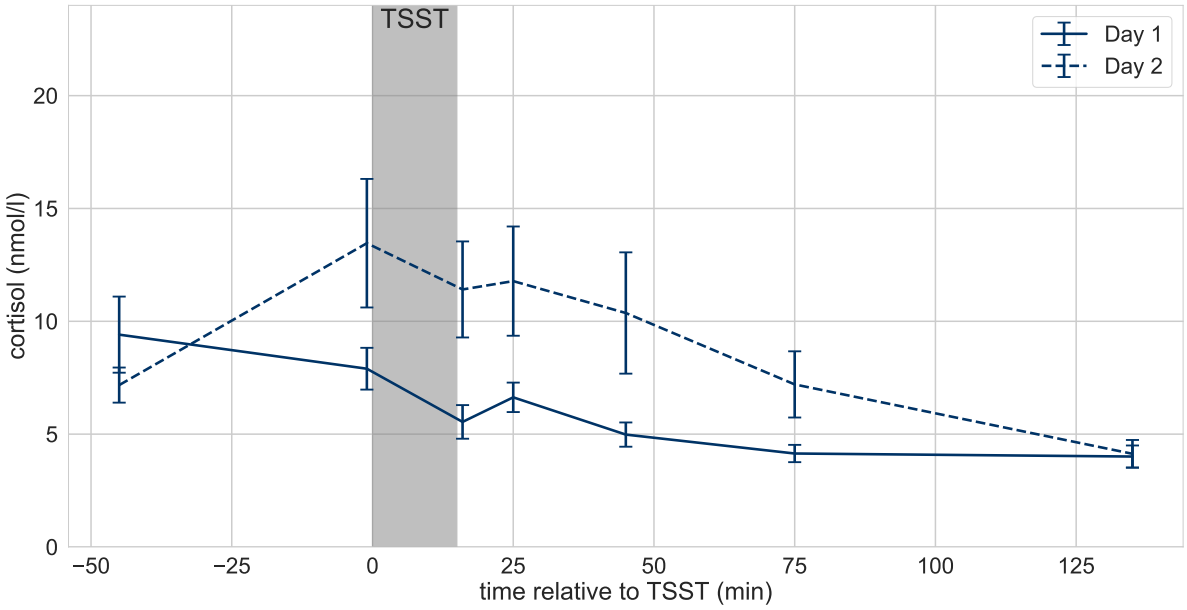


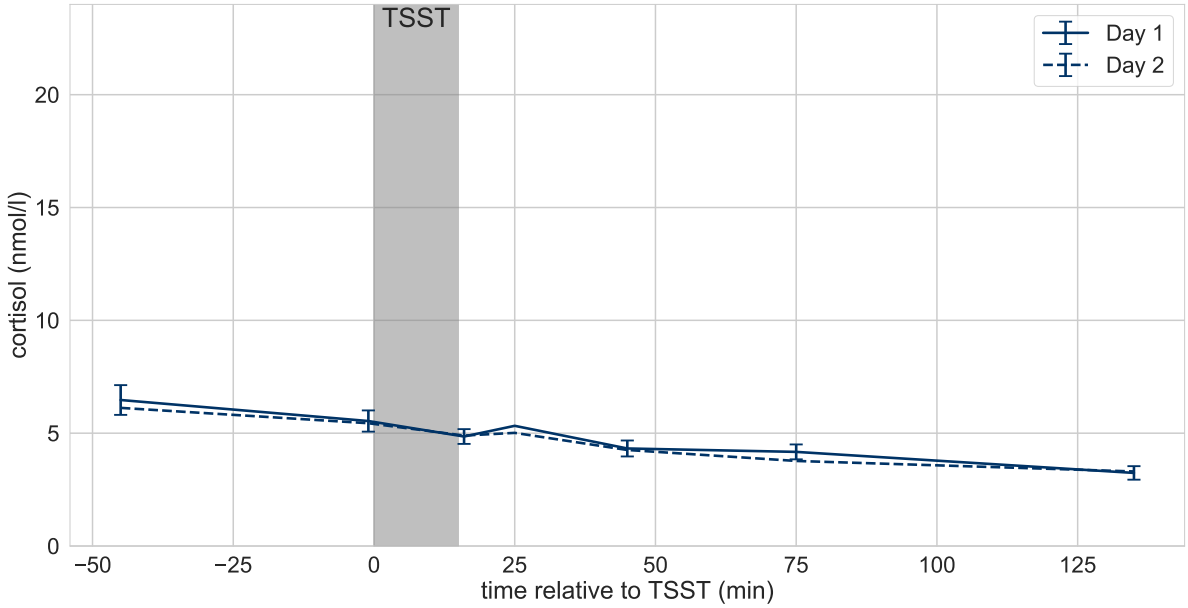
Figure B.3: Decision Tree for cortisol classification. 0-7 refers to the normalized cortisol values of the first day, 8-13 represent the values on the second day.



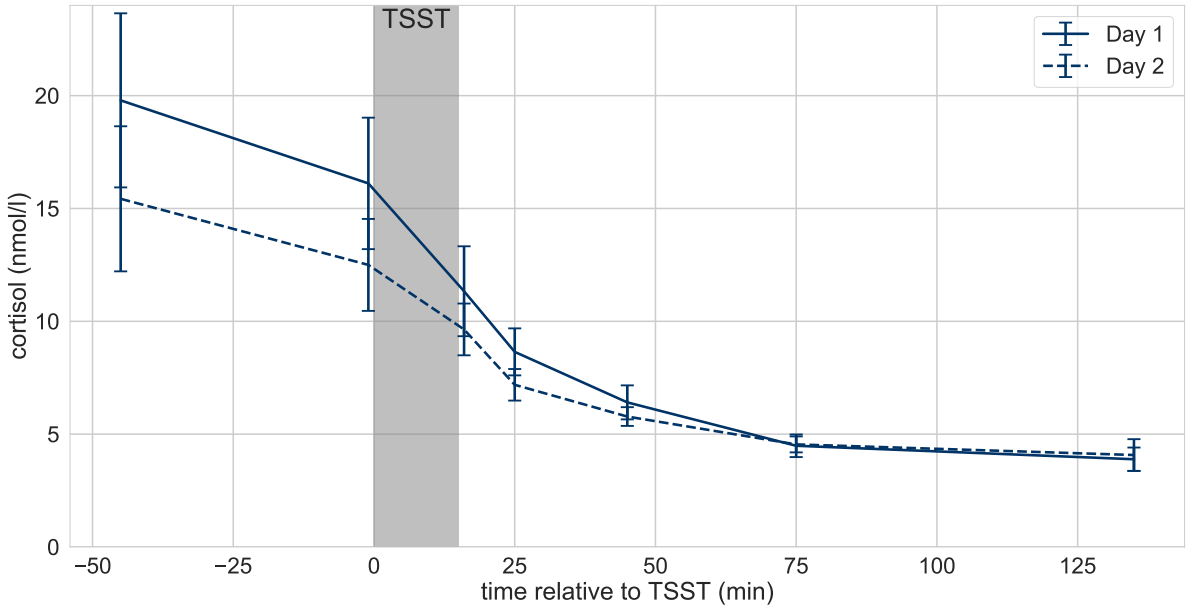
(a) Non-Habituation



(b) Sensitization



(c) Non-Responder



(d) Anticipatory

Figure B.3: Cortisol Classes. Mean and standard error.

Glossary

ACTH Adrenocorticotrophic Hormone

ANOVA Analysis of Variance

ANS Autonomous Nervous System

Automatic Activation The involuntary processing of stimuli in preparation for associated responses. This activation tends to occur more rapidly than that resulting from an intention. [Nug13]

Comorbidity Describes a further illness or syndrom, which can be distinguished by diagnosis

Cortisol A hormone from the adrenal cortex, the principal glucocorticoid [Lan11c]

CRH Corticotrophin Releasing Hormone

CV Cross-validation

Homeostasis Maintenance of a dynamically stable state within a system by means of internal regulatory processes that counteract external disturbances of the equilibrium [Bro93]

HPA-Axis Hypothalamic–Pituitary–Adrenal Axis

IL-6 Interleukin-6

kNN k-Nearest-Neighbors

ML Machine Learning

PTSD Posttraumatic Stress Disorder

RBF Radial Basis Function

sAA Salivary Alpha-Amylase

SMOTE Synthetic Minority Over-sampling Technique [Cha02]

SNS Sympathetic Nervous System

SVM Support Vector Machine

TSST Trier Social Stress Test

List of Figures

List of Tables

Bibliography

[Ana15] *Hypothalamic–pituitary–adrenal axis*, *Anatomography, http://lifesciencedb.jp/bp3d, accessed 06.01.2019*, 2015.

[App87] A. Appels, P. Höppener, P. Mulder: *A questionnaire to assess premonitory symptoms of myocardial infarction*, *International Journal of Cardiology*, Vol. 17, Nr. 1, 1987, pp. 15–24.

[Bre14] J. G. Breines, M. V. Thoma, D. Gianferante, L. Hanlin, X. Chen, N. Rohleder: *Self-compassion as a predictor of interleukin-6 response to acute psychosocial stress*, *Brain, Behavior, and Immunity*, Vol. 37, 2014, pp. 109–114.

[Bro93] L. Brown: *The new shorter Oxford English dictionary on historical principles*, Clarendon Press, 1993.

[Cha02] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer: *SMOTE: Synthetic Minority Over-sampling Technique*, *Journal of Artificial Intelligence Research*, Vol. 16, 2002, pp. 321–357.

[Chi02] Chih-Wei Hsu, Chih-Jen Lin: *A comparison of methods for multiclass support vector machines*, *IEEE Transactions on Neural Networks*, Vol. 13, Nr. 2, 2002, pp. 415–425.

[Coh83] S. Cohen, T. Kamarck, R. Mermelstein: *A Global Measure of Perceived Stress*, *Journal of Health and Social Behavior*, Vol. 24, Nr. 4, 1983, pp. 385.

[D'A11] A. D'Ambrosio, V. A. Tutore: *Conditional Classification Trees by Weighting the Gini Impurity Measure*, Springer, Berlin, Heidelberg, 2011, pp. 273–280.

[Dic04] S. S. Dickerson, M. E. Kemeny: *Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research*, *Psychological Bulletin*, Vol. 130, Nr. 3, 2004, pp. 355–391.

[Dun07] A. J. Dunn: *The HPA Axis and the Immune System: A Perspective*, *NeuroImmune Biology*, Vol. 7, 2007, pp. 3–15.

[Fik19] A. Fiksdal, L. Hanlin, Y. Kuras, D. Gianferante, X. Chen, M. V. Thoma, N. Rohleder: *Associations between symptoms of depression and anxiety and cortisol responses to and recovery from acute stress*, *Psychoneuroendocrinology*, Vol. 102, apr 2019, pp. 44–52.

[Fri15] J. U. Frisch, J. A. Häusser, A. Mojzisch: *The Trier Social Stress Test as a paradigm to study how people respond to threat in social interactions*, *Frontiers in Psychology*, Vol. 6, 2015.

[GB02] E. Gonzalez-Bono, N. Rohleder, D. H. Hellhammer, A. Salvador, C. Kirschbaum: *Glucose but not protein or fat load amplifies the cortisol response to psychosocial stress*, *Hormones and Behavior*, Vol. 41, Nr. 3, 2002, pp. 328–333.

[Gia14] D. Gianferante, M. V. Thoma, L. Hanlin, X. Chen, J. G. Breines, P. M. Zoccola, N. Rohleder: *Post-stress rumination predicts HPA axis responses to repeated acute stress*, *Psychoneuroendocrinology*, Vol. 49, Nr. 1, 2014, pp. 244–252.

[GL13] I. R. Galatzer-Levy, A. Nickerson, B. T. Litz, C. R. Marmar: *Patterns of lifetime PTSD comorbidity: A latent class analysis*, *Depression and Anxiety*, Vol. 30, Nr. 5, 2013, pp. 489–496.

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

[GL17] I. R. Galatzer-Levy, S. Ma, A. Statnikov, R. Yehuda, 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*, Vol. 7, Nr. 3, 2017.

[Kal02] M. Kalia: *Assessing the economic impact of stress-the modern day hidden epidemic.*, *Metabolism: clinical and experimental*, Vol. 51, Nr. 6, 2002, pp. 49–53.

[Kir92a] C. Kirschbaum, S. Wüst, D. Hellhammer: *Consistent sex differences in cortisol responses to psychological stress*, *Psychosomatic Medicine*, Vol. 54, Nr. 6, 1992, pp. 648–657.

[Kir92b] C. Kirschbaum, S. Wüst, C. J. Strasburger: *'Normal' cigarette smoking increases free cortisol in habitual smokers*, *Life Sciences*, 1992, pp. 435–442.

[Kir93] C. Kirschbaum, K. M. Pirke, D. H. Hellhammer: *The 'Trier social stress test' - A tool for investigating psychobiological stress responses in a laboratory setting*, in *Neuropsychobiology*, 1993.

[Kir94] C. Kirschbaum, D. Hellhammer: *Salivary cortisol in psychoneuroendocrine research: recent developments and applications*, *Psychoneuroendocrinology*, Vol. 19, Nr. 4, 1994, pp. 313–333.

[Kud04] B. M. Kudielka, N. C. Schommer, D. H. Hellhammer, C. Kirschbaum: *Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day*, *Psychoneuroendocrinology*, Vol. 29, Nr. 8, 2004, pp. 983–992.

[Kur17] Y. I. Kuras, C. M. McInnis, M. V. Thoma, X. Chen, L. Hanlin, D. Gianferante, N. Rohleder: *Increased alpha-amylase response to an acute psychosocial stress challenge in healthy adults with childhood adversity*, *Developmental Psychobiology*, Vol. 59, Nr. 1, 2017, pp. 91–98.

[Lan11a] F. Lang, P. Lang: *Blut und Immunsystem*, in *Basiswissen Physiologie*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 17–40.

[Lan11b] F. Lang, P. Lang: *Ernährung, Verdauungstrakt, Leber*, in *Basiswissen Physiologie*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 149–179.

[Lan11c] F. Lang, P. Lang: *Hormonale Regulation*, in *Basiswissen Physiologie*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 235–268.

[Lan11d] F. Lang, P. Lang: *Vegetatives Nervensystem*, in *Basiswissen Physiologie*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 327–340.

[Lem17] G. Lemaitre, F. Nogueira, C. K. Aridas: *Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*, *Journal of Machine Learning Research*, Vol. 18, Nr. 17, 2017, pp. 1–5.

[Mar11] R. L. Marca, P. Waldvogel, H. Thörn, M. Tripod, P. H. Wirtz, J. C. Pruessner, U. Ehlert: *Association between Cold Face Test-induced vagal inhibition and cortisol response to acute stress*, *Psychophysiology*, Vol. 48, Nr. 3, 2011, pp. 420–429.

[Mas93] G. Mastorakos, G. P. Chrousos, J. S. Weber: *Recombinant interleukin-6 activates the hypothalamic-pituitary-adrenal axis in humans*, *The Journal of Clinical Endocrinology & Metabolism*, Vol. 77, Nr. 6, 1993, pp. 1690–1694.

[McE98] B. S. McEwen: *Protective and Damaging Effects of Stress Mediators*, *New England Journal of Medicine*, Vol. 338, Nr. 3, 1998, pp. 171–179.

[McI14] C. M. McInnis, M. V. Thoma, D. Gianferante, L. Hanlin, X. Chen, J. G. Breines, S. Hong, N. Rohleder: *Measures of adiposity predict interleukin-6 responses to repeated psychosocial stress*, *Brain, Behavior, and Immunity*, Vol. 42, 2014, pp. 33–40.

[Mil05] G. E. Miller, N. Rohleder, C. Stetler, C. Kirschbaum: *Clinical depression and regulation of the inflammatory response during acute stress*, *Psychosomatic Medicine*, Vol. 67, Nr. 5, 2005, pp. 679–687.

[Nas07] N. M. Nasrabadi: *Pattern Recognition and Machine Learning*, *Journal of Electronic Imaging*, Vol. 16, Nr. 4, 2007, pp. 049901.

[Nat09] U. Nater, N. Rohleder: *Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research*, *Psychoneuroendocrinology*, Vol. 34, Nr. 4, 2009, pp. 486–496.

[Nie83] H. Niemann: *Klassifikation von Mustern*, Springer, 1983.

[Nug13] P. Nugent: *Automatic Activation*, *PsychologyDictionary.org*, 2013.

[Orb13] S. Orbell, H. Schneider, S. Esbitt: *Hypothalamic-Pituitary-Adrenal Axis*, in *Encyclopedia of Behavioral Medicine*, Springer New York, New York, NY, 2013, pp. 1017–1018.

[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, É. Duchesnay: *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825–2830.

[Rai03] C. L. Raison, A. H. Miller: *When Not Enough Is Too Much: The Role of Insufficient Glucocorticoid Signaling in the Pathophysiology of Stress-Related Disorders*, *American Journal of Psychiatry*, Vol. 160, Nr. 9, 2003, pp. 1554–1565.

[Roh02] N. Rohleder, B. M. Kudielka, D. H. Hellhammer, J. M. Wolf, C. Kirschbaum: *Age and sex steroid-related changes in glucocorticoid sensitivity of pro-inflammatory cytokine*

production after psychosocial stress, *Journal of Neuroimmunology*, Vol. 126, Nr. 1-2, 2002, pp. 69–77.

[Roh09] N. Rohleder: *Determinants and consequences of the acute stress-induced inflammatory response in human aging. Research Grant Proposal to the American Federation for Aging Research (AFAR)*, 2009.

[Sap00] R. M. Sapolsky, L. M. Romero, A. U. Munck: *How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions*, *Endocrine Reviews*, Vol. 21, Nr. 1, 2000, pp. 55–89.

[Sch98] P. Schulz, W. Schlotz, P. Becker: *Trierer Inventar zum Chronischen Stress (TICS) [Trier Inventory for Chronic Stress (TICS)]*, *Diagnostica*, Vol. 45, 1998, pp. 8–19.

[Sch03] N. C. Schommer, D. H. Hellhammer, C. Kirschbaum: *Dissociation Between Reactivity of the Hypothalamus-Pituitary-Adrenal Axis and the Sympathetic-Adrenal-Medullary System to Repeated Psychosocial Stress*, *Psychosomatic Medicine*, Vol. 65, Nr. 3, 2003, pp. 450–460.

[Tho17] M. V. Thoma, D. Gianferante, L. Hanlin, A. Fiksdal, X. Chen, N. Rohleder: *Stronger hypothalamus-pituitary-adrenal axis habituation predicts lesser sensitization of inflammatory response to repeated acute stress exposures in healthy young adults*, *Brain, Behavior, and Immunity*, Vol. 61, 2017, pp. 228–235.

[Udr88] I. Udris, M. Frese, D. Frey: *Belastung, Stress, Beanspruchung und ihre Folgen*, in *Angewandte Psychologie*, 1988, pp. 427–429.

[vK06] R. von Känel, B. M. Kudielka, D. Preckel, D. Hanebuth, J. E. Fischer: *Delayed response and lack of habituation in plasma interleukin-6 to acute mental stress in men*, *Brain, Behavior, and Immunity*, Vol. 20, Nr. 1, 2006, pp. 40–48.

[Wea02] J. D. Weaver, M.-H. Huang, M. Albert, T. Harris, J. W. Rowe, T. E. Seeman: *Interleukin-6 and risk of cognitive decline: MacArthur studies of successful aging.*, *Neurology*, Vol. 59, Nr. 3, 2002, pp. 371–8.

[Wüs04] S. Wüst, E. F. C. Van Rossum, I. S. Federenko, J. W. Koper, R. Kumsta, D. H. Hellhammer: *Common Polymorphisms in the Glucocorticoid Receptor Gene Are Associated with Adrenocortical Responses to Psychosocial Stress*, *Journal of Clinical Endocrinology and Metabolism*, Vol. 89, Nr. 2, 2004, pp. 565–573.