CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

Machine Learning for Suicide Risk Assessment on Depressed Patients

The use of Electro Dermal Orienting Reactivity to Identify Hyporeactive Patients

Master's thesis in Computer science and engineering

ARNAUD MOULIS

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2019

Master's thesis 2019

Machine Learning for Suicide Risk Assessment on Depressed Patients

The use of Electro Dermal Orienting Reactivity to Identify Hyporeactive Patients

ARNAUD MOULIS

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Telephone +46 31 772 1000
SE-412 96 Gothenburg
Gothenburg, Sweden 2019

Machine learning for Suicide Risk Assessment on Depressed patients
The use of Electro Dermal Orienting Reactivity to identify hyporeactive patients
Arnaud Moulis

Supervisor: Alexander Schliep, Department of Computer Science and Engineering
Industrial Supervisors: Claes Holmberg & Daniel Poté, Emotra AB
Examiner: Graham Kemp, Department of Computer Science and Engineering

Master's Thesis 2019
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2019

Machine learning for Suicide Risk Assessment on Depressed patients
The use of Electro Dermal Orienting Reactivity to identify hyporeactive patients
Arnaud Moulis
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Suicide is a global phenomena and the leading cause of death in some countries and age groups, accounting for nearly 1 million deaths and 10 million attempts per year. It costs society a substantial amount of money both directly and indirectly not to mention the tremendous amount of emotional distress and pain for families and friends. This thesis is the first of its kind that tries to automate a manual process of suicide risk assessment with machine learning techniques, successfully doing so with support vector machine models. A precision of 82% and an accuracy of 89% is reached and proves the potential to further develop this method for assessing suicide risk among depressed patients.

Keywords: Suicide Risk Assessment, Suicide Detection, Machine Learning, Support Vector Machine.

Acknowledgements

I would like to thank my supervisor Alexander Schliep for his support and guidance during this thesis as well as my examiner Graham Kemp for constructive feedback and ideas. Special thanks goes to Claes Holmberg and Daniel Poté for encouragement and allowance to work with their start-up company.

Gothenburg, June 2019
Arnaud Moulis

List of Figures

Notation

Matematical expressions

Notation	Meaning
$\bullet$	Dot product
$\|\|v\|\|$	Norm of a vector
v^T	Transposed vector
$\hat{v}$	Transposed vector

Abbreviations

Abbreviation	Meaning
ASP	Affective Signal Processing
COA	Critical orientation area
DC	Direct Current
EDA	Electrodermal activity
ER	Event related
FIR	Finite impulse response
GOA	General orientation area
HMM	Hidden Markov Model
ML	Machine learning
NS	Non-specific
SCL	Skin conductance level

1
Introduction

1.1 Background

Suicide accounts for approximately 1 million deaths globally with an additional 10 million suicide attempts globally per year and is the leading cause of death in some countries and age groups[1]. Suicide both directly and indirectly costs society a substantial amount of money and inflicts enormous emotional distress for relatives and friends. Over the years, suicide prevention methods and strategies have been researched to become more effective, but the diagnostic tools have remained limited to interviews with patients. The interviews with the patients are a challenge for clinicians to standardize and interpret in terms of suicide risk due to subjectivity. A more objective approach would be preferable.
Researchers noted already 100 years ago that depressed patients had lower electrodermal activity (EDA) than the typical population [38] and has been confirmed by similar studies since then. Furthermore, Edman and co-workers [7] found that there also was a faster habituation of repeated sound stimuli, called hyporeactivity, among depressed patients with a record of violent suicide attempts versus patients with non violent suicidal attempts . Violent suicidal attempts is regarded as cuts in arms, jumping from heights or any other attempts which inflict severe pain. Overdosing on medication is for example not regarded as a violent suicide attempt [7].

Until quite recently, Thorell [30] showed a significant difference between suicide attempts and hyporeactivity which was inspired from Sokolov's theory on general and specific orienting reactions for repeated stimuli [25, 26]. Figure 2.3 shows a colour coded graph that idicates the different types of orientation reactions which Sokolov regarded as relevant in determining hyporeactivity.

The tests conducted by Thorell and others to detect hyporeactivity were similar in that after three or five minutes of silence a number of identical audio stimuli were presented in varying intervals. The stimuli were sinus tones in the frequency of 1 kHz, 85 or 90 dB over general hearing threshold, one second duration, given in 15 to 80 second intervals [9].

With regard to Thorell's work of establishing a significant correlation between hypoactivity as theoretically depicted in Figure 1.1 and suicide risk, this thesis aims to automatically detect and classify hyporeactivity by analyzing EDA signals.

[1] World Health Organization. Suicide facts, 2018

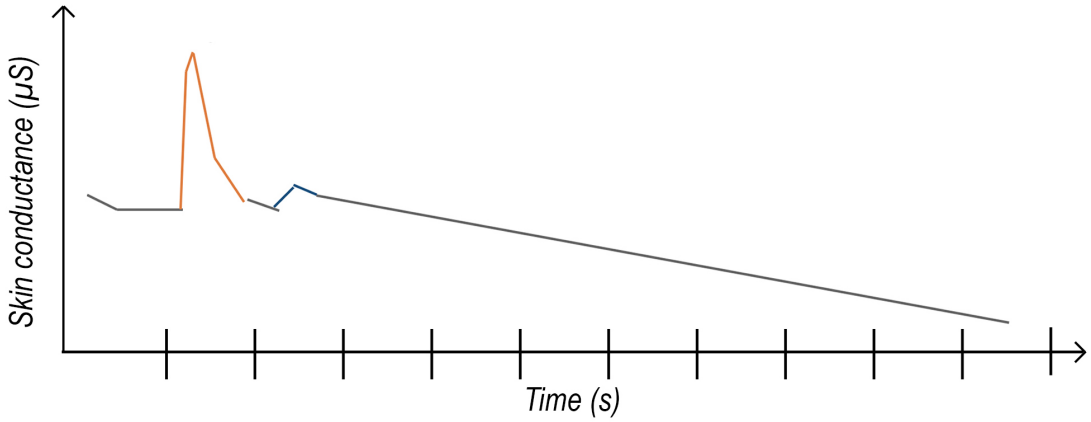


Figure 1.1: *Y-axis is the skin conductance while X-axis is time. Along the X-axis vertical lines are depicted as sound stimulus. A reaction after the first sound stimulus is expressed as an orange colored curve followed by declining values in gray indicating the individual ignores the consecutive sound stimuli across time.*

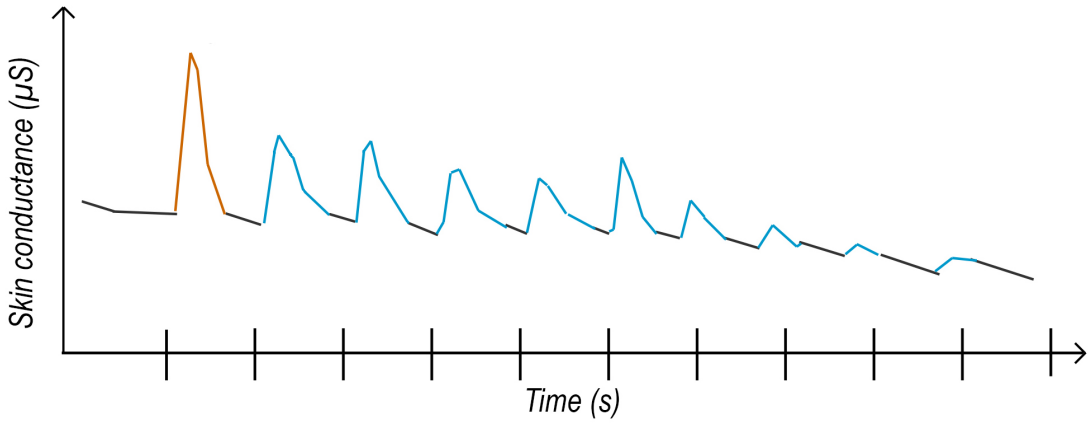


Figure 1.2: *A reaction after the first sound stimulus is expressed as an orange colored curve, but unlike Figure 1.1 is followed by consecutive smaller reactions in parallel with sound stimuli, indicating that the individual responds to the sound stimuli several times after the initial sound stimulus.*

1.2 Related Work

To the best of our knowledge, this is the first attempt to classify suicidal risk among diagnostically depressed patients using machine learning on EDA signals. However, research papers on preprocessing of times series such as EDA exists. Notably in research papers, EDA also goes under the name of skin conductance (SC), galvanic skin response (GSR) or electrodermal response (EDR) but is today gathered under one name, EDA [2].

Nevertheless, many articles have worked with EDA signals, specially with the problem of distinguishing calmness and distress from the signal. Varying results ranging from 75% to 95% in accuracy [21, 15, 11, 24] have been reached. It is however important to highlight that these papers including other approaches have used a significant number of features from more than one physiological sensor. This thesis, due to heavily relying on the EDA signal, is greatly influenced by a relatively recent paper from 2017 [39], where a team of researchers successfully classified conditions of calmness and distress only using EDA analysis.

They applied a low-pass Finite Impulse Response (FIR) filter onto the signal to reduce the noise produced from different electrical stages. Next, they applied a cubic spline interpolation algorithm to separate the baseline from the relevant signal. Finally, they extracted the relevant features and used a decision tree based model to classify their data. Another research team used a more sophisticated algorithm to separate the baseline from the relevant signal [10]. They claim a baseline elimination accuracy of 96% on their own artificially created EDA signals. The latter algorithm is applied in this thesis but is discarded since it is not applicable on our data; a detailed explanation can be found under section 3.2.1.

The mentioned papers above all deal with binary classification between calmness and distress. This thesis deals with a classification problem of EDA signals which has never been done before and must therefore consider other features and time intervals than previous research papers. For this reason, the recognized and copious book about EDA [2] has been used heavily for inspiration in this thesis.

1.3 Motivation

Observational studies have shown a correlation of suicide risk and hyporeactivity which can be measured with EDA by placing electrodes on the skin, see Figure A.1 in Appendix. EDA has been proven to be significantly different between depressed patients and depressed patients who are in risk of suicide [31], Figures 1.1 and 1.2, modified from Emotra's website[1] , shows an ideal example of this difference. However, the EDA between the groups are not as clearly defined as these theoretical images depicts since experts with vast knowledge of EDA manually has to incorporate mathematical tools as support to interpret the data correctly [31].

[1]Emotra. The edor method, June 2019. URL: http://emotra.se/sv/product/

1.4 Goal

The goal of this project is to automate the interpretation of available data for suicide risk assessment by using machine learning. The automation consists of a binary classifier which after input of biometric data categorizes the data as either hyporeactive or reactive. The project consists of following steps:

- Review different machine learning methods.
- Given the quality, structure and volume of data, select the most suitable method or build a model.
- Test and verify functionality of chosen model.

1.5 Scope

This project only presents the model which yielded the best results after trial and error between two considered models to solve this problem, Hidden Markov modeling (HMM) and Support Vector machine modeling (SVM). HMM yielded positive results and is a rational method to use for this classification problem since it builds upon the assumption that EDA signals consists of and can be simulated with emission and transition probabilities between states, a main criteria for HMM to work. This assumption is confirmed by a research paper [10] that demonstrated that EDA can be simulated with three components; a phasic signal, a tonic signal and noise which are described more in detail in Section 2.1. However, since the noise component is influenced by many confounding factors such as age, genetics and mental state [5], hence generating a high variance in the data and given the small number of hyporeactive examples in our data set, it was not surprising that the HMM did not yield any reliable results. To counter the excessive noise, homing in on specific segments of the EDA signal was preferable. Thus, focus shifted to the SVM method since it does not need to consider the whole sequence of the EDA signal, thereby excluding most of the noise and limiting the confounding factors within the data.

1.6 Problem specification

From a sequence of EDA data and heart rate, suicide risk should be assessed by classifying it as either hyporeactive or reactive data. Upcoming paragraphs summarizes the essential steps needed in solving the problem using SVM. A more detailed description of the tasks can be found in Chapter 3.

Task 1: Pre-processing The available EDA-signals and the manual labeling of them are in different files and formats, which has to be extracted and concatenated into one single database. The EDA-data is noisy (contains spikes of unwanted muscle contractions from the patients), irregular (has different scales from person to person) and is sometimes of different sizes or partly lacks a signal. Therefore a need for removing outliers, cleaning and standardizing the data is necessary.

Task 2: Feature extraction The next step is to extract features from the cleaned data which allows the SVM to differentiate between hyporeactive and reactive patients.

Task 3: Training & tuning 80% of the labeled data is used for training. Training is done over all combinations of extracted features, along with parameter tuning of a 5-folded cross-validated grid-search for each combination.

Task 4: Classification & evaluation The remaining 20% of the labeled data is used for evaluation. All trained models which are optimized for a unique set of features and corresponding tuned parameters predicts each EDA sequence from the remaining 20% as either hyporeactive or reactive. The top 10 models which yielded best scores are presented.

1.7 Thesis Outline

This chapter describes the aim and problem specification of this master thesis. Chapter 2 gives an insight into the theory of SVM and its algorithm in addition to the theory of hyporeactivity classification in which this thesis builds upon. Chapter 3 describes the pipeline of the method and the implementation of the theory used. The results from the algorithm are presented in Chapter 4 and a discussion about the method and the generated results can be found in Chapter 5 along with ideas for future work. Finally, Chapter 6 consist of a conclusion of the master thesis.

2

Theory

2.1 Electrodermal activity

EDA can be largely said to consist of skin conductance response (SCR) and skin conductance level (SCL). SCRs are short bursts of increased amplitude in Micro-Siemens (μS), event related (ER-SCR) if bound to a stimulus, non-specific (NS-SCR) otherwise. The amplitude of a SCR has before been arbitrarily defined as 0.05 μS due to visual inspection [2], however, this definition has been changing since the introduction of computer scoring algorithms. The other component of EDA is a slower changing level of μS. The difference between the two is clearly depicted in Figure 2.1, reproduced from an article online [8].

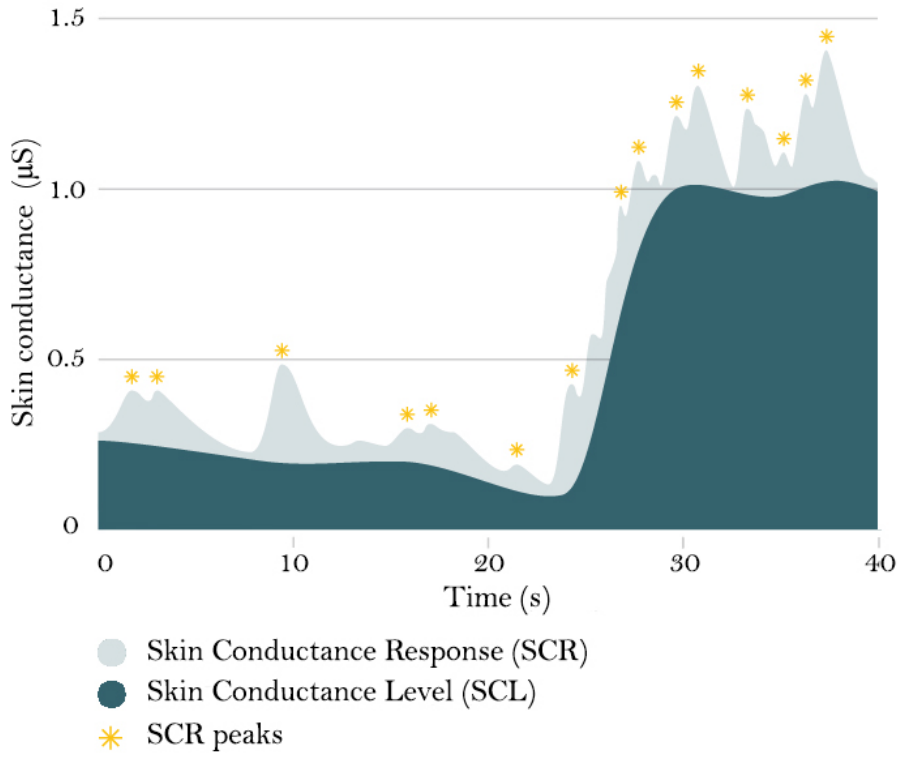


Figure 2.1: *A theoretical example of an EDA graph with its components (noise excluded)*

2.1.1 SCR Response reaction

The exact and true definition of a SCR is unclear and is therefore difficult to filter out from the high presence of noise within the data. Nevertheless, one significant assumption has been considered in this work; the SCR is only considered as such if its rising time starts within the time interval of 0.4 to 4 seconds after sound stimuli onset. It cannot occur 0.4s before stimuli onset due to limitations of the sympathetic system [2]. The intervals can be observed as vertical lines in Figure 2.3.
Regarding the SCR itself, there are some relevant definitions of SCR measures in Table 2.1 along with corresponding Figure 2.2 borrowed from *the* standard reference source on EDA written by Wolfram Boucsein [2]. It is important to note that it has been shown that SCR gestalts varies depending on skin type and genetic aspects between individuals [32].

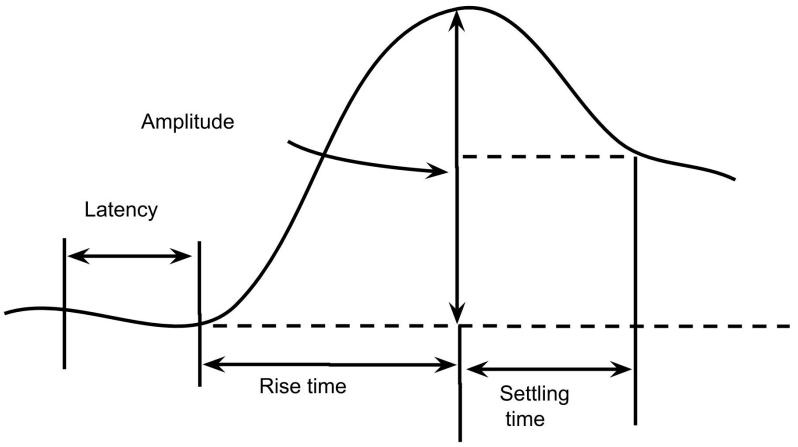


Figure 2.2: *An example of an SCR and some measurements. Clarification of the measurements can be seen in Table 2.1*

Table 2.1

Measure	Defenition	Typical Values
SCL	Tonic level of electrical conductivity of skin	2-20 μS
Change in SCL	Gradual changes in SCL measured at two or more points in time	1-3 μS
Frequency of NS-SCRs	Number of SCRs in absence of identifiable electing stimulus	1-3 per min
ER-SCR amplitude	Phasic increase in conductance shortly following stimulus onset	0.2-1.0 μS
ER-SCR latency	Interval between stimulus onset and SCR initiation	1-3 sec
ER-SCR raise time	Interval between SCR initiation onset and SCR peak	1-3 sec
ER-SCR settling time	Interval between SCR peak and 50% recovery point of SCR amplitude	2-10 sec
ER-SCR (trials to habituation)	Number of stimulus presentations before two or three trials with no response	2-8 stimulus presentations
ER-SCR habituation (slope)	Rate of change of ER-SCR amplitude	0.01-0.05 μS per trial

2.1.2 Time window of SCR

Algorithms for the detection of EDA gestalts are not easily obtainable for computational analysis. To extract certain event-related SCRs, the time window to consider on the EDA is of utter importance. Unfortunately, there is no clear definition or guideline for these time windows. For example, in one paper [36] the time window is set between 1 and 3s , in another 1.2 and 4s [16] or even 1 and 2.4s [6]. Due to the lack of an exact criteria of appropriate time windows or their inadequate application, some misinterpretation of the SCR and its features is inevitable.

2.1.3 Orientation stimuli

Orientation stimuli is a group name for stimuli which are segmented into different categories due to the insights they reveal. The categories and segments can be found in Figure 2.3.

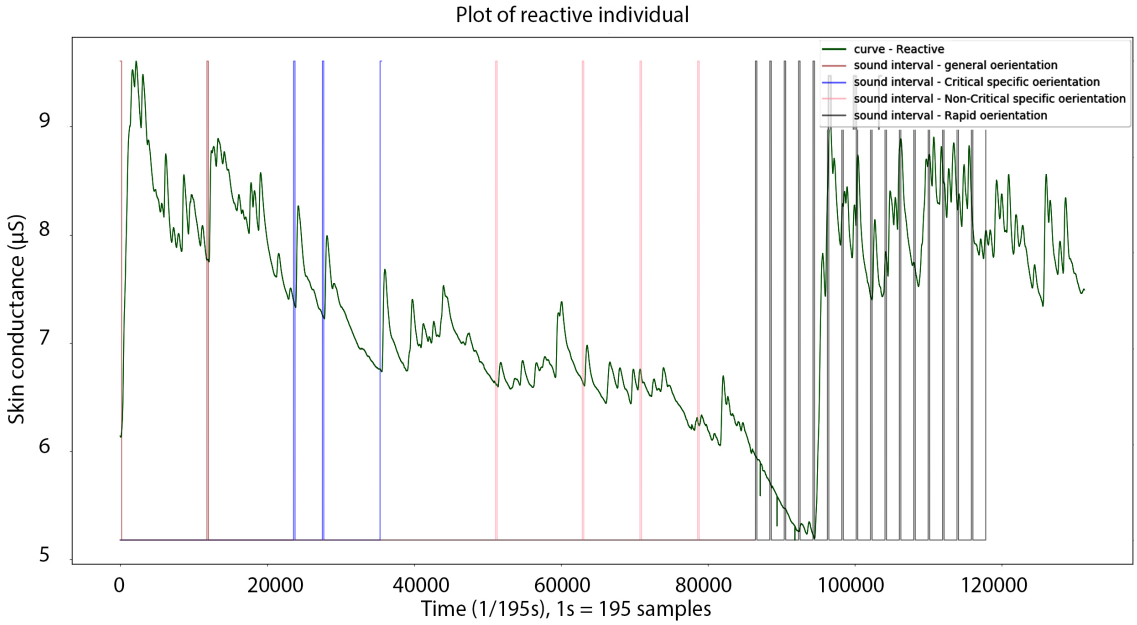


Figure 2.3: *An example of a reactive individual with colour coded orientation segments and stimuli onsets.*

General orientation area (GOA) or the segment consisting of the two first sound stimuli functions as attention grabbing stimuli, "waking up" the brain and making it susceptible and ready for upcoming stimuli. It works in the same way as one would turn around and redirect all ones attention to an unexpected loud sound coming from behind. GOA consists of two signals since some people do not redirect all their attention after solely the first stimulus. GOA therefore serves as trying to objectively standardize the attention level among all participants for the upcoming stimuli.

The following three sound stimuli (3,4,5), marked in blue, are part of the critical orientation area (COA) which is considered to be of most importance when it comes

to classifying participants in the experiments mentioned by Thorell. It is important since it usually contains the habituation pattern necessary for classification. The habituation pattern is defined as three consecutive non-SCRs [31]. However, it is hard to differentiate a SCR from a non-SCR with noise created by participants since they at times are seemingly indistinguishable. Nevertheless, if a reaction occurs after sound stimuli 3,4 and 5, it is hence probable to be classified as reactive since the pattern of habituation usually occurs quickly unless interfered with noise. A theoretical example of a habituation pattern occurring after onset of the first stimulus can be seen in Figure 1.1. The next 4 sound stimuli (6,7,8,9) belonging to the non-critical orientation area (NCOA) are less relevant, again, since hyporeactive patients tend to have revealed a habituation pattern this far into the experiment. However, NCAO can still help to infer hyporeactivity if habituation is prolonged by any means. Lastly, the consecutive fast and rhythmic sound stimuli in the end of Figure 2.3 are there for convention due to similar studies including this part. It provides some information on other psychological aspects but is nothing this thesis is taking into account. In summary, this thesis will only consider the most relevant orientation areas, COA and NCOA to detect hyporeactivity. In other words, only the time windows after sound stimuli 3 to 9 are to be considered.

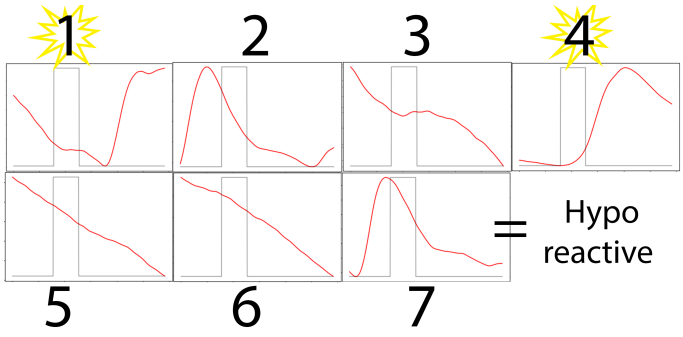


Figure 2.4: *Segments taken from COA and NCOA, ordered 1 to 7. Habituation pattern detected after stimulus 4. Image 2 and 7 have the shape of a reaction curve but occurs before the sound stimulus itself, hence regarded as noise.*

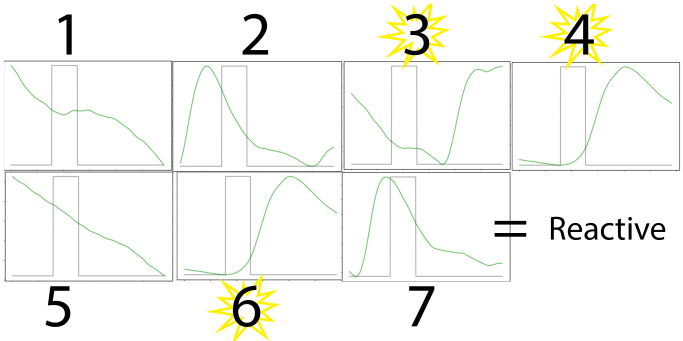


Figure 2.5: *No habituation pattern found since there is no 3 consecutive non-reactions present.*

2.2 Support Vector machine

Support Vector Machines (SVMs) build on the principle of separating classes of data points with a linear decision boundary on a 2D plane, or hyperplane if there are more than 2 dimensions. For further details refer to [35] and [34]. A line is conventionally described as $y = mx + b$, however, dealing with hyperplanes m gets exchanged with ω^T which represents the parameters of the plane and y describes if the point is on either side of the decision boundary, in a way functioning as a label. If y is negative, it belongs to the triangle class, if it is positive it belongs to the circle class as can be seen on 2.6. The line or decision boundary separating the plane can be described as 2.1, where x is a point on the plane and b is the distance from the origin. For upper boundary of one class, that equation will equal to 1 and for the other it will equal to -1, both representing the margin of our SVM classifier

$$\omega^T \bullet x + b = 0. \tag{2.1}$$

The goal is to determine the best ω which separates the classes, that is, a separation which has the the largest margin between the classes. Many approaches can be used, the simplest is linear SVM.

2.2.1 Linear SVM

Given a set of N samples, $X = x_1...x_N$ and the goal to separate the two classes with the largest margin, one wants to find the normal vector $\omega = \omega_1...\omega_K$ using a set of equations 2.2. Start by defining a support vector, the vector which points to the closes sample x_s. Let Z_p be a point on the hyperplane, ϵ the perpendicular distance between the hyperplane and x_s whilst b determines the offset of the hyperplane from the origin along ω

$$\begin{cases} \omega^T \bullet x_s + b = 1 \\ x_s = z_p + \epsilon\hat{\omega}, \end{cases} \tag{2.2}$$

$$\omega^T \bullet (z_p + \epsilon\hat{\omega}) + b = 1 \Longleftrightarrow \epsilon\omega^T \bullet \hat{\omega} + \omega^T \bullet z_p + b = 1 \Longleftrightarrow \epsilon\|\omega\| = 1. \tag{2.3}$$

Rearranging equation 2.2 leads to equation 2.3 which gives the solution to the best margin by minimizing $|\omega|$ or $|\omega|^2$. Next, to find the optimal ϵ no sample may reside within the margin. An equation describing this criterion is 2.4 where y_i is the class for each sample. Following this criteria results in a hyperplane in which the classifier best can linearly separate the classes Fig. 2.2,

$$y_i \left(\omega^T \bullet x_i + b\right) \geq 1. \tag{2.4}$$

2.2.2 Soft margin SVM

In many hyperplanes the expression of equation 2.4, that no sample may reside within the margin does not hold true. Some samples are bound to reside on the wrong side of the separation line. To solve this, a penalty ζ is applied on each

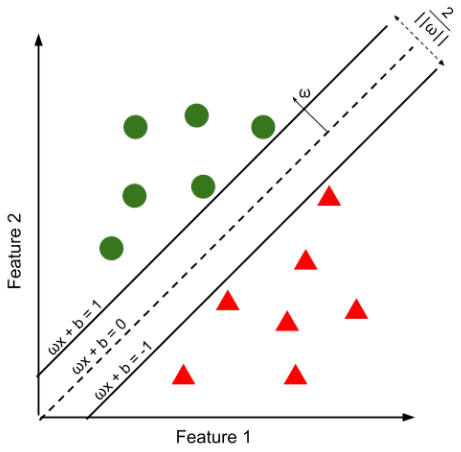


Figure 2.6: *Separation between classes using linear SVM*

sample on the wrong side of the separation line turning the SVM model into a so called soft margin model. With this model, one searches for the optimal C to solve equation 2.5 before finding the optimal separation line as

$$\arg\min_{\omega,b,\zeta} \left(\|\omega\|^2 + C \sum_i \zeta_i \right), \tag{2.5}$$

following criteria of equation 2.6

$$y_i \left(\omega^T \bullet x_i + b \right) \geq 1 - \zeta \quad \zeta \geq 0. \tag{2.6}$$

The equation makes a trade-off between penalizing wrongly classified samples and changing the size of the margin. The larger C becomes , the greater the penalization.

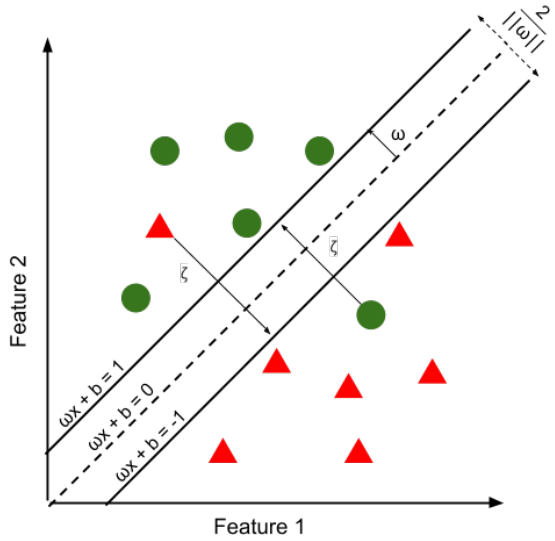


Figure 2.7: *Separation between classes using soft margin SVM*

2.2.3 Non-linear SVM

In reality, both linear and soft margin SVMs are inadequate to solve many classification problems due to not being linearly separable. A well know example of a non-linear classification problem is the *XOR problem* seen on Figure 2.8. A solution to this problem is proposed by Boster et al. [1] where kernels $k(x_i, x_j)$ and a third dimension are added to the hyperplane. The kernels below are three commonly used kernels:

- Linear: $k(x_i, x_j) = x_i^T * x_j$
- Polynomial: $k(x_i, x_j) = (x_i * x_j + 1)^d, \gamma > 0$
- Radial Basis Function (RBF): $k(x_i, x_j) = exp(-\gamma||x_i - x_j||^2), \gamma > 0$

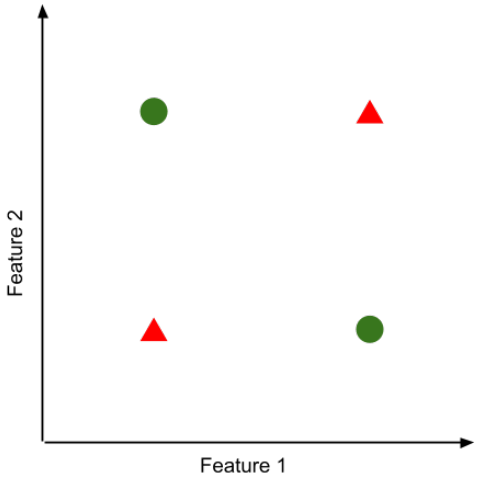


Figure 2.8: *XOR classification problem*

Here, γ, d, k and c are parameters set by the user. Due to the limited time of this project only the two simplest kernels were used: the RBF and linear kernel. The polynomial kernel has many more parameters, augmenting the complexity of the model [12] and was therefore excluded. For all feature combinations the RBF kernel outperformed the linear kernel due to it being more complex in its shape. An extensive description of kernels and how they work can be found in a paper from MIT [23].

2.2.4 Overfitting

When fitting boundary algorithms to the sample data it is easy to think that the best decision boundary is the one which achieves to separate the classes perfectly. This is not the case since the sample data on which the algorithm has trained simply does not represent the real world. Once new data are generated or observed, the trained algorithm's perfect boundary no longer fits the newly acquired data rendering it faulty by definition. This phenomenon is called overfitting. In contrast, if the boundary does not fit the sample or training data at all, it most likely will not

perform better on new data neither; this is called underfitting. An appropriate fit to the data is one which separates most of the data whilst not having a too complex boundary line giving it robustness against unseen future data. In Figure 2.9 one can see mentioned examples.

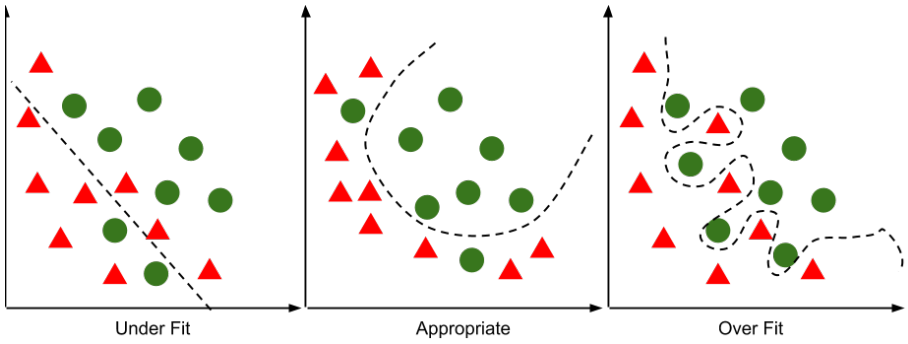


Figure 2.9: *Example of over,under and appropriate separation of classes in a hyperplane*

3
Methods

3.1 Overview

The data acquired from Emotra was gathered using the EDOR System[1], and consists of several columns from which the most relevant for this thesis are:

- **The sound stimuli**. This is used to indicate when a patient receives stimuli and the start of the stimuli is used to calculate response times of the other signals.
- **The skin conductance**, this is the primary input of the method. How this signal reacts to a sound stimulus is the central part of the method.

Less relevant data for this thesis are:

- **Event Markers** – Inputs from the test leader to indicate that the patient did something that deviated from the normal state of waiting and listening. E.g. if the patient sneezes or starts to cough.
- **Heart Rate (DC)** - Gives an approximation of the heart rate signal from the patient.

The numerous Event markers throughout the data were plotted and considered in the pre-processing step but were not presented since the attached comments of event markers all claim no significant influence on the EDA signal. Furthermore, heart rate approximation was highly inaccurate and faulty but was included in the model anyway.

Table 3.1: Quantitative measures of the groups can be found in this table. Part of the acquired data is of medium quality due to the contemporary development of data collection of patients.

	Hypo-Reactive	Reactive	Total
Medium-Quality	173	500	673
High-Quality	88	256	344
Total	261	756	1017

[1]Emotra. Edor system and products, June 2019. URL: http://emotra.se/sv/product/

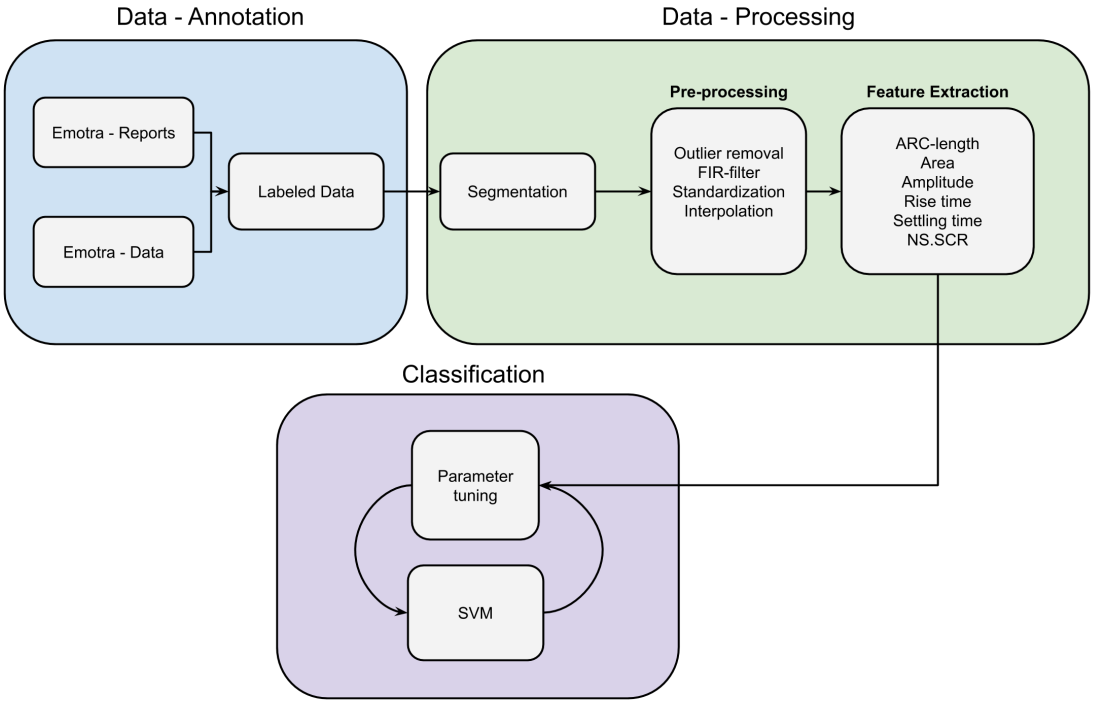


Figure 3.1: *Flowchart of process*

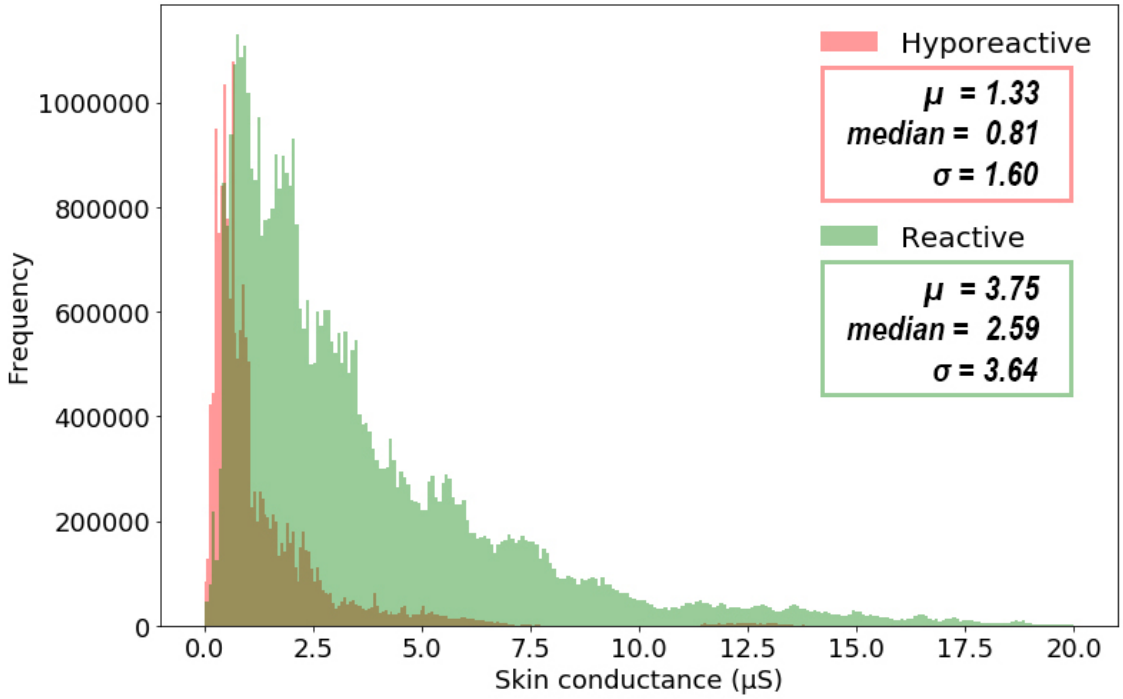


Figure 3.2: Density plots of hyporeactive and reactive sequence values. The mean, median and standard deviation for each group is also presented and will serve as a guide for the baseline classifier

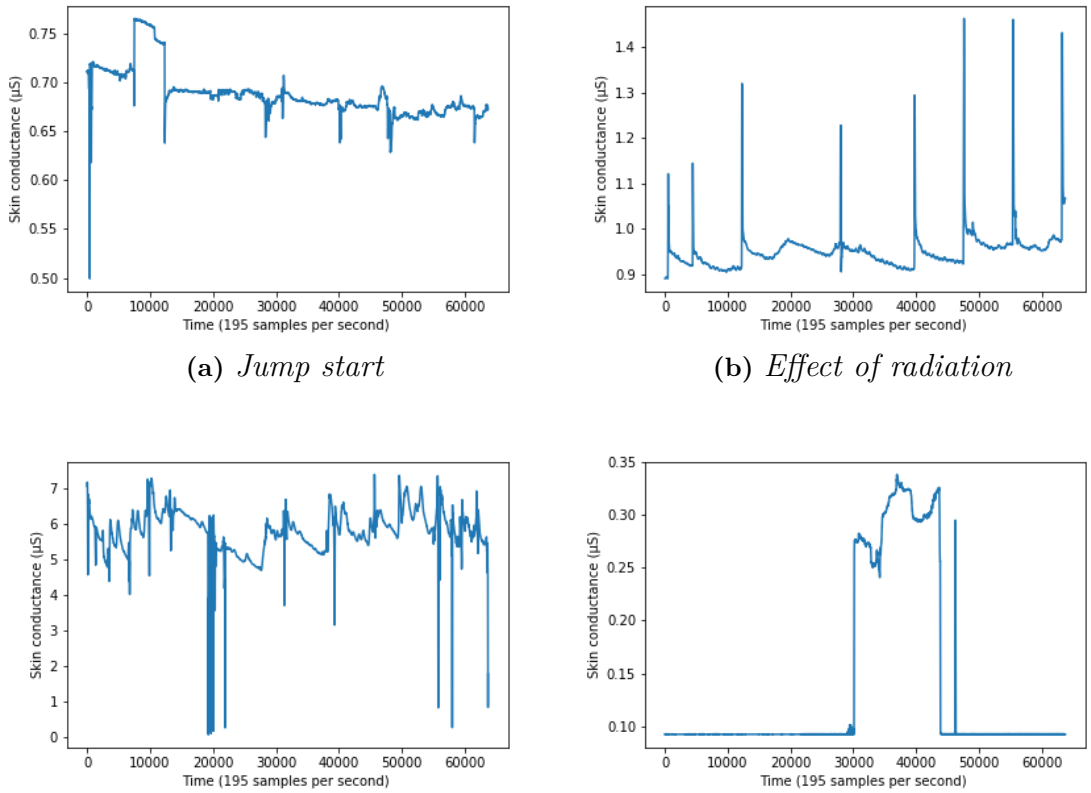
3.2 Pre-processing

Merging of data

The provided data from Emotra consisted of two types of data, firstly the signal data (EDA) in tab-separated values format, secondly the manual labeling in a Microsoft Word document. The manual labeling of EDA signal was carried out by various assistants and was lastly confirmed by Thorell. The EDA data and the labels were imported into Python and merged into a single database using Natural language processing tools. This process resulted in the loss of a third of the EDA data due to missing labels. Furthermore, the complete data set is a concatenation of recordings which had slightly different measurement procedures. The ratio between them can be seen in Table 3.1.

Noise in data

The recording of skin conductance was done with Direct Current (DC). Due to this, observed data has generated artifacts such as very high peaks or very low valleys at the beginning of measurement. This phenomenon is due to the body of the individual acting as a conductor charging the body with electrodes until a stable level for measurements can be reached as seen here in Figure 3.3a.



(a) *Jump start*

(b) *Effect of radiation*

(c) *Signal drop*

(d) *Outlier*

Figure 3.3: Pre-processing issues

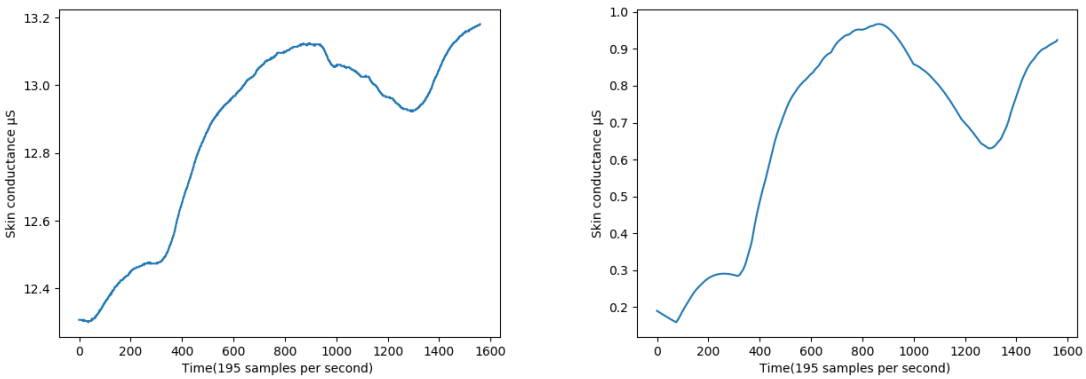
Another type of artifact causing deep valleys within the signal data might be due to the individual moving their fingers or in any way disconnecting the electrodes from skin causing a severe dip in the signal data as can be seen in Figure 3.3c.

On the contrary, a steep increase might be due to a cough or a sneeze which shortly increases muscle tension and awareness in the individual. Another explanation is that the measurement tools are sensitive to radiation of some sort, also causing the steep spikes upward. This could be due to nearby electrical devices.

A very important aspect to consider is that some people perspire more from their hands than others, resulting in better conductivity and thus a more explicit output. The opposite is true for people with very dry hands. To counter this difference an amplifier is adapting the sensitivity of the measurement instrument at the beginning of the experiment for each individual, finding a level of measurement which can register the EDA signal properly. This adaptation does not come without cost since this very process can set the recording sensitivity to such a high level it becomes hard to tell the difference between SCR or noise overall. In some cases the amplification is so high the very bits of measurement are visible in the data.

3.2.1 Baseline, SCL removal

The extraction of SCR usually needs a subtraction of SCL in the EDA signal, as described in paper [39]. Since this thesis only is considering short time windows of the EDA signal, the phasic SCL signal does not have enough time to influence the SCR signal. To confirm this, the appliance of the cvxEDA algorithm was done on each EDA signal. The extraction and removal of SCL can be seen in Figure 3.4 which followed the recommended parameter settings used by Greco et al.(2016) [10].



(a) *Unprocessed EDA signal* **(b)** *SCL removal & normalization*

Figure 3.4: Automatic baseline elimination using the cvxEDA algorithm

Even though the paper [10] shows how the algorithm achieves 96.7% sensitivity regarding SCL elimination in artificially generated EDA data, this thesis chooses

not to use it considering the short time intervals and not having a perfect baseline elimination. It is also questionable if artificially generated EDA signals presented in the research paper resembles the real ones extracted by Emotra.
However, disregarding the SCL elimination, the algorithm provided this project with NS.SCR frequency between COA and NCO. This frequency was chosen to be used as a feature in training process.

3.2.2 Standardization

Once the cleaning of data was complete the measurements of different individuals have to become comparable with each other. Again, since some of the individuals perspire more than others, the recorded EDA differs tremendously in values. To make these recordings comparable, standardization is necessary. Standardization is commonly used and sometimes fundamental for certain machine learning algorithms to work properly.

The goal of the standardization process is to put different variables on the same scale. It achieves this by calculating the mean and standard deviation for a variable. Then, for each observed value of the variable, subtracting the mean and dividing it by the standard deviation, essentially turning the mean of the variable to 0. The values within the variable has now become a score which reveals how far they are from the mean in terms of standard deviations. A score of 1 for one specific observation means that it falls 1 standard deviation away from the mean. This interpretation is true regardless of the type of variable one standardizes [20].

3.2.3 Segmenting

Certain parts of the EDA signal are more interesting than others, as described in Section 2.1.3. The section mentions that the onset of a reaction has to start withing an interval of 0.4 to 4 seconds after each sound stimulus.
To catch the whole curve, including the settling period of reactions, an interval of 0.4 to 8 seconds after each sound stimulus was cut out within the specific and non-specific orientation area resulting in 7 segments of 1560 data points each, as seen in Figure 2.5.
To get a better idea of general types of reactions among reactive and hyporeactive individuals, K-means clustering of the Euclidean distance for sequential data called Tslearn [29] was applied to all segments. In Figure 3.5 one can observe the 6 most typical hyporeactive segments and in Figure 3.6 the 6 most typical reactive segments.

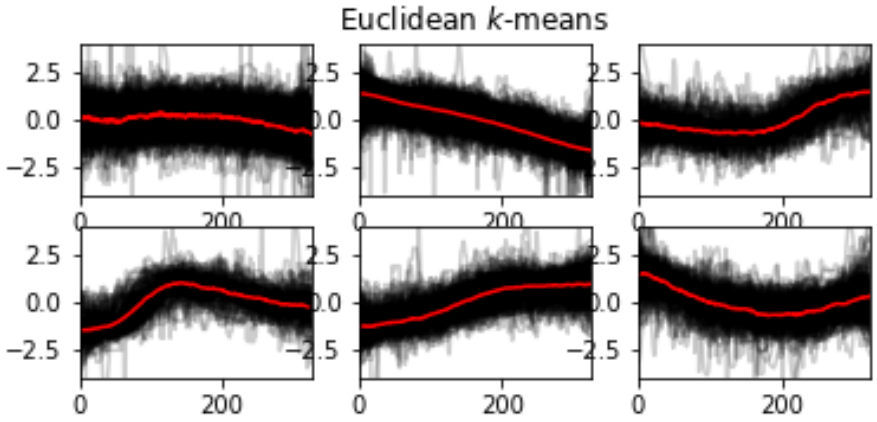


Figure 3.5: The clustering of hyporeactive segments. The X-axis representing time has been down-sampled form 1600 to 300 data points. The Y-axis represent the log values of skin-conductance.

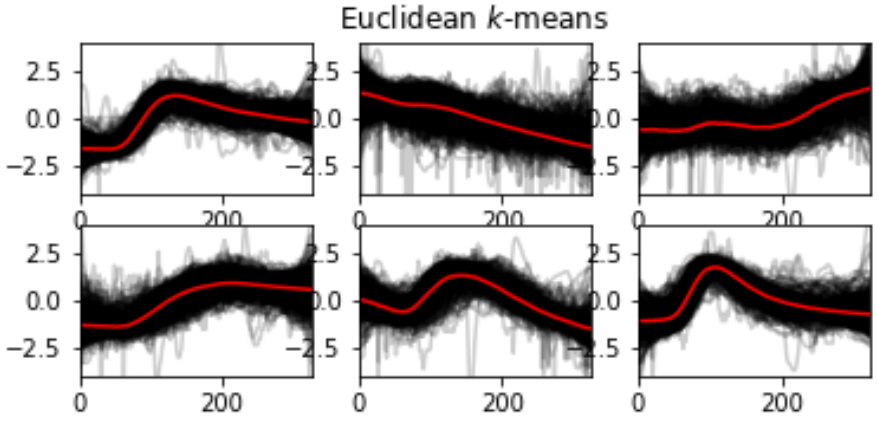
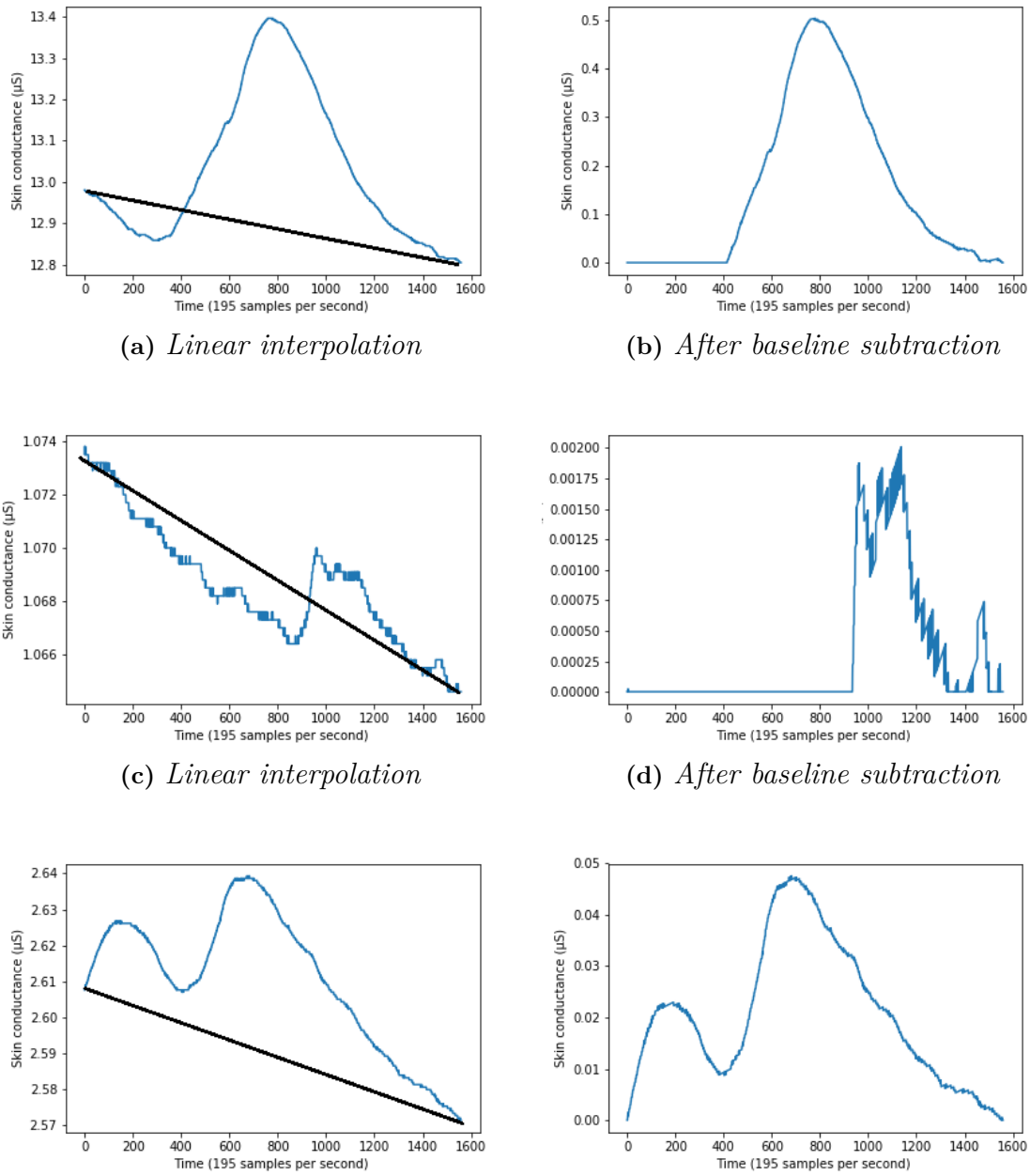


Figure 3.6: The clustering of reactive segments. The X-axis representing time has been down-sampled form 1600 to 300 data points. The Y-axis represent the log values of skin-conductance.

3.2.4 Feature extraction

An SCR consists of area, arc length, amplitude, raising time and a settling time as depicted on Figure 2.2. To extract the features of the SCR, the SCR itself has to be extracted. Since the cvxEDA algorithm explained in Section 3.2.1 did not succeed in extracting SCR, a simpler SCR extraction is needed. In the spirit of Ockhams razor, a simple interpolation between the first and last point of the segment is made to find the baseline. The SCR is extracted by removing the Y-values on the line from the curve as seen Figure 3.7.
Once the SCR curves were extracted, area was obtained from the curves using trapezoidal rule instead of Simpson's rule because the curves in Figure 3.7 many times are non-parabolic which causes problems for Simpson's rule. Amplitude was obtained by simply picking the highest value in the sequence. Length of the curve was obtained with equation 3.1 where δX was set to 1 and N = 1560, the data



(a) *Linear interpolation*

(b) *After baseline subtraction*

(c) *Linear interpolation*

(d) *After baseline subtraction*

(e) *Linear interpolation*

(f) *After baseline subtraction*

Figure 3.7

points for the whole segment. Settling time was attained of δX from the amplitude point to the point in which the amplitude had decreased 50%. In a similar fashion, raising time was attained of δX from the amplitude point to the point in which 5% of the amplitude was reached. If the points of interest were not reached, the δX simply becomes the length from the amplitude point to x = 0 whilst settling time

became the length from amplitude point to x = 1560.

$$ARC = \sum_{n=1}^{N} \sqrt{(\delta Y)^2 + (\delta X)^2} \tag{3.1}$$

To see if there is a difference between the features box plots were made. Due to the large variance in the data, only log values were presented in the box plots. Box plots of each segment position for each group feature were made, for example (Hypo Area1, Hypo Area2...Hypo Area7) against (Reactive Area1, Reactive Area2...Reactive Area7) were made but did not show any interesting deviation in values between segment positions and are therefore not presented. Only a general box plot of all area values between the groups, i.e (Hypo Area) against (Reactive Area) are presented in Figure 3.8.

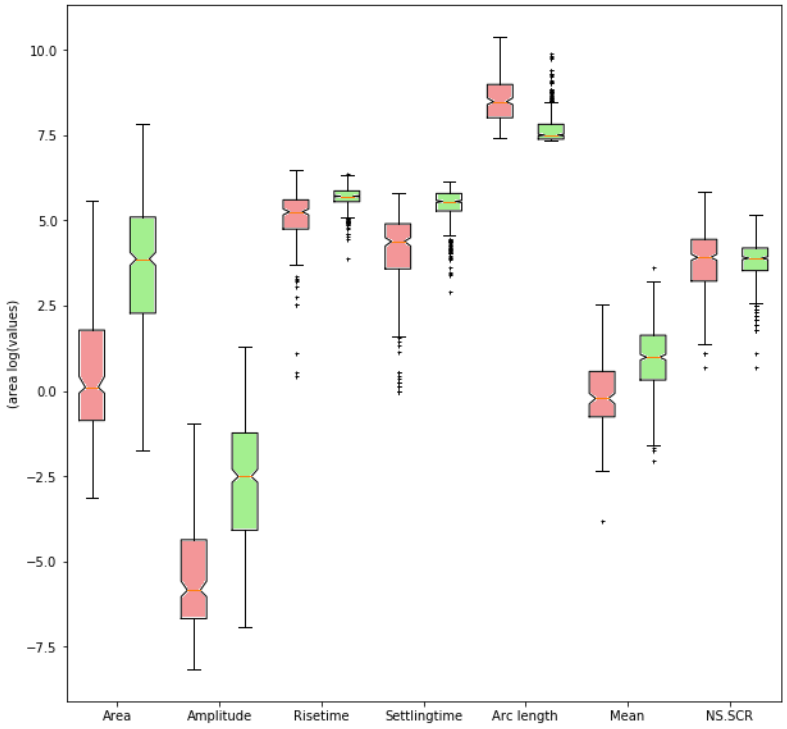


Figure 3.8: *Boxplots revealing how the values differ between the groups, visually giving an idea of which features could become linearly separable. Heart rate is excluded in this plot due to insignificant difference between the groups*

3.3 Classification

3.3.1 Baseline

A baseline classifier is useful as a point of reference when evaluating how well a trained machine learning model performed. For classification problems "ZeroR" is a common baseline classifier, which corresponds to assigning every unknown data point to the mode of the data set, in other words the class with most observations,

see Equation 3.4. In addition to this, two additional baseline classifiers are presented, one regarding the lowest mean between the classes, see Equation 3.2 and another between the lowest median of the classes, see Equation 3.3. The lowest median and mean are based on the descriptive data of the two classes, as can be seen in Figure 3.2. A sequence is denoted S, group of hyporeactive samples denoted G_H, group of reactive samples denoted G_R, *1* for hyporeactivity classification and *0* for reactivity classification.

$$S(p) = \begin{cases} 1 & if\ \ S(p) \leq mean(G_H) \\ 0 & else \end{cases}, \tag{3.2}$$

$$S(p) = \begin{cases} 1 & if\ \ S(p) \leq median(G_H) \\ 0 & else \end{cases}, \tag{3.3}$$

$$S(p) = \begin{cases} 1 & if\ \ mode(G_H + G_R) = 1 \\ 0 & else \end{cases}. \tag{3.4}$$

3.3.2 SVM

The library used for SVM classification is from SKlearn [3]. As described in Section 2.2, a set of features for hyporeactive and reactive individuals are placed on a hyper plane in which the goal is to find a linear separation between them. The set of features can be found in Table 3.2.

Table 3.2

Features	
AE	Area
AM	Ampitude
ARC	ARC length
HR	Heart rate
NS.SCR	Number of detected NS.SCRs in COA and NCOA
RT	Raise time
ST	Settling time

When a separation-line is found, individuals having a greater value than the threshold of the separation-line is assigned to hyporeactive and vice versa.

3.3.3 Tuning of SVM parameters

The purpose of tuning model parameters is to circumvent the algorithm from getting stuck in a local maxima or minima [12]. For Non-linear SVM, two parameters are important to consider, γ and C. γ was initialized with values between 0.1 and 100 whilst C was initialized between 0.1 and 1000. The higher the γ value, the more it tries to fit the training data. C is a penalty parameter of the error term and controls the trade off between a smooth decision boundary and classifying the training points

correctly. If C or γ is set too high, the training will result in overfitting as depicted in Figure 2.9. The γ and C parameters were chosen by highest F1-score and used as a guide to do a finer grid search of the parameters.

3.4 Implementation details

3.4.1 Software

The SVM classifier was implemented using sklearn [3] and was implemented in the Python environment named Spyder[19], part of the Anaconda platform[18]. Other important libraries used were Numpy, pandas, matplotlib and scipy and neurokit [17] in which the cvxEDA algorithm was found.

4
Results & Evaluation

This Chapter first describes the evaluation process of the models and then presents the results.

4.1 Evaluation measures

In order to evaluate our model and between models confusion matrix is used. The confusion matrix is frequently used for evaluating binary classifiers accuracy [27], by displaying its result consisting of four different values:

- True positive (TP) Samples which are correctly classified as Hyporeactive.
- True negative (TN) Samples which are correctly classified as Reactive.
- False positive (FP) Samples which are incorrectly classified as Hyporeactive.
- False negative (FN) Samples which are incorrectly classified as Reactive.

	Predicted True	Predicted False
Labeled Positive	True positive (TP)	False negative (FN)
Labeled Negative	False positive (FP)	True negative (TN)

A number of performance measurements can be derived from the confusion matrix. Accuracy in Equation 4.1, shows the overall performance of the classifier, given a symmetric data set (as many false positive as false negative).
Precision in Equation 4.2, answers the question, of all patients predicted as hyporeactive, how many actually are hyporeactive?
Recall in Equation 4.3, answers the question, of all patients that are hyporeactive, how many did we detect?
Finally, the F1 score is a measurement of the balance or harmonic average between precision and recall.A score of 1 means a perfect precision and recall 4.4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4.4}$$

4.2 Experiment Setup

Since our model's true performance only can be measured empirically on new unseen data which has not been a part of the training, we separate the total data into two categories; Training data and test data. The training data will be used as input to the models while the test data will be used for evaluation with the evaluation measures above. Using the evaluation measures above is an excellent way to evaluate binary classifiers [13]. However, evaluating it solely on one part of our data can be misleading, specially if the data is irregular and has a limited sample size.
To avoid an unfair evaluation due to biasing the model with one specific division of the data, the training and testing of the model will iterate over different divisions of the sampled data as can be seen in figure 4.1. This method is called cross validation or k-fold cross validation, k referring to the number of iterations through the sample. There is no formal rule to the value of k but usually it is set to 5 or 10 since they have been shown to empirically yield the least non-biased and the least variance error estimates [13]. "As k gets larger, the difference in size between the training set and the resampling subsets gets smaller. As this difference decreases, the bias of the technique becomes smaller" [14]. Therefore, choosing k to equal 5 is to be considered a valid choice in our thesis. Due to cross-validating our model the sample size for hyporeactive models is 52 and for reactive models each sample is of size 151.

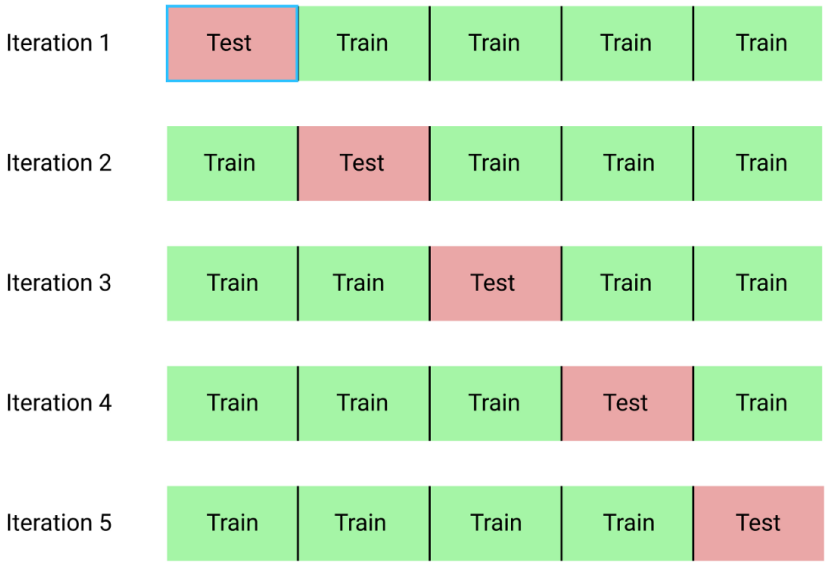


Figure 4.1: Cross validation

4.3 Results

The results are sorted with descending F1 score. The scores from the SVM only contains the top 10 results, all of them except one found the best score when C = 1000 and gamma = 0.001 with RBF-kernel.

	Accuracy	Precision	Recall	F1-score
Baseline				
Mean	0.737	0.669	0.490	0.566
Median	0.780	0.488	0.583	0.532
Frequency	0.744	0.500	0	0
SVM				
AM + RT + ST	0,887	0,778	0,729	0,753
AM + RT + ST + M	0,887	0,778	0,729	0,753
AE + ST + M + NS.SCR	0,872	0,822	0,673	0,740
AE + RT + ST + M	0,877	0,778	0,700	0,737
AE + ST + NS.SCR	0,872	0,800	0,679	0,735
AM + ST + NS.SCR	0,872	0,800	0,679	0,735
AM + ST + M + NS.SCR	0,872	0,800	0,679	0,735
AE + AM + RT + ST	0,877	0,756	0,708	0,731
AM + RT + ST + ARC	0,872	0,778	0,686	0,729

5
Discussion

5.1 Results

The best model in terms of F1-score achieved 78% precision and 73% recall. With regard to the scope and goal of this thesis, and outperforming the baseline classifier, these results did prove the functionality of machine learning on automating the classification process of hyporeactivity. Noteworthy to point out, the highest precision score achieved is 82%. This means that when the classifier finds a hyporeactive patient it is 82% likely to be correct, announcing to the expert that this patient needs further investigation as quickly as possible. Since suicidal individuals are rather unpredictable and might be gone the next day, waiting time for treatment is crucial and the precision score relevant. On the other hand, recall reveals how many of all hyporeactive patients the classifier finds, which means 27% of patients are not detected by the classifier. This means that those patients get the usual waiting time for treatment. With this in mind, this thesis allows faster treatment for almost $\frac{3}{4}$ of the patients in immediate need.
By observing Figure 3.8 it is not surprising that the SVM model found Area, Raising time and Settling time to be the features yielding highest scores since their values differ between the groups. The reason why they differ can be explained by the theory in which this thesis builds upon, that hyporeactive patients have a habituation pattern, hence less manifestation of SCRs, hence less area. Less SCRs also means that hyporeactive patients will have shorter raising and settling times. This can be visually confirmed from the clustering of segments in Figure 3.5 and Figure 3.6 and with further confirmation from the box plots of 3.8.

Overall an F1 score of 75% for most ML algorithms is not great, and there are no other results to compare with regarding classification of hyporeactive or reactive patients. The only comparable references are of other EDA related studies, such as classification between calmness and distress. In one study, using a very complex mixture of models such as k-means, Gaussian mixture models, SVM and decision trees, a precision of 75% was reached [15]. In another study, using heart rate variability (HRV) to supplement the EDA signal, 78% precision was reached using SVM classifiers [22]. To the author's knowledge, the best results in classifying calmness and distress only using EDA signal is from a research paper paper published 2017 [39]. They reached a recall of 93.9% and precision of 85.36% using a tree based model. They used 17 features which all were significantly distinguishable between the calm and distressed group, sometimes with values differing by a factor

of three. Indeed, this thesis does not reach as high precision or recall scores as the 2017 paper [39], but again they are not classifying the same problem and face different challenges and feature considerations. Two examples are given given below.

Firstly, this thesis seeks out to find a habituation pattern concealed in segments of short time windows after stimuli onsets. This means an appropriate time window needs to be found to capture the ER-SCRs. To add to the complexity, as described in Section 2.1.2, the time window for detecting ER-SCR is bound to individual characteristics.

Secondly, the amount of features to be extracted from these segments limits the model of this thesis compared to other research papers classifying calmness and distress. This does not mean that features extracted from longer sequences, such as NS.SCR, does not help. However, judging from the literature this thesis is building upon, NS.SCR or any other feature outside the time windows of sound stimuli onset should not differ between hyporeactive or reactive patients.

5.1.1 Future Work

For future work, considering other features outside the ER-SCR might be worth investigating for the classification of hyporeactivity. Additionally, a more sophisticated prepossessing technique should be applied to ensure that the feature extraction is more accurate. A refined interpolation method of ER-SCR extraction within the small segments would surely increase the precision score of the model. How this is to be done needs to be researched since there does not seem to be any simple rule for it as seen in Figure 3.7. One idea could be to make use of the clustering of segment signals in Figure 3.5 to pick an appropriate interpolation technique for ER-SCR extraction. It would also be of interest to see which sequences that are harder to classify than others such as between A and B in Figure 5.1. However, this exceeds the scope of this thesis.
Lastly, regarding the time window estimate for segmentation, Stern and Walrath proposed to standardize the time windows differently for every individual using the individual modal value [28]. If applied, this could make an big impact on the accuracy of the model.

5.1.2 Data quality & Co founder

As with any machine learning problem, the results one gets is largely due to the quality of the data and the preprocessing method of it. The problems of the preprocessing method used in this thesis gives one explanation of the results, but the discussion of the noisy data and why it seems to differ from other research papers deserves attention as well.

A very crucial part of Affective Signal Processing (ASP) is that emotional responses, such as the reaction to a certain image or sudden sound, is embedded in layers of psychological and physiological components which is manifesting itself in different

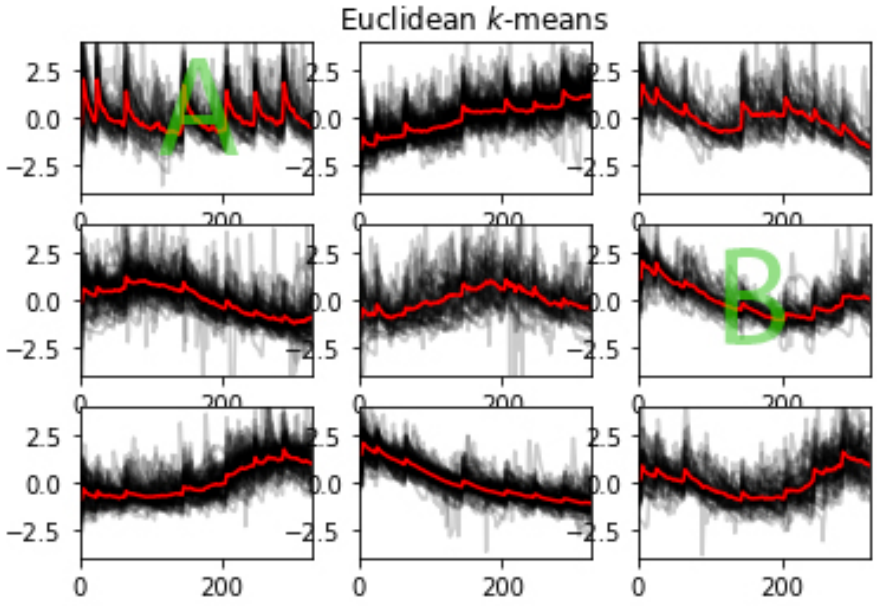


Figure 5.1: Clustering of sequences

time periods. For example, psychological aspects might be disposition (long term - years), circumstance (days), mood (hours) and emotion (seconds)[4]. To create an accurate model of effective responses and regard these multitude of influences on a medium such as Skin conductance is not an easy task. Especially, this could be true for a sample of depressed patients since disposition, circumstances and moods are naturally expected to deviate from the normal population. It most certainly would explain the high variance in the data which in turn could explain the differences in results compared with other papers.

Secondly, the type of the measurement tool and its location in a room with other electrical devices plays an important role in getting accurate measurement [33]. Since Emotra uses their own measurement tool for recording EDA signals and the location of nearby electronic devices might differ from other studies, comparing the data quality between studies is perplexing.

Lastly, the paper of 2017 [39] only has young participants in their sample whilst this thesis sample consists of people from all age groups. Age has been shown to bias the recordings of EDA signals in a study where the purpose was to study age and gender effects [37]. However, due to conflicting research, there is no consensus in the scientific community about this aspect [2].

5.2 Ethical perspective

As with all machine learning applications, ethical questions arise. Since this paper sets out to assess people in the risk-zone of committing suicide, which allows for earlier treatment in form of heavy medication, the need of correct classification is crucial. Relying too much on results given by machine learning classifications might be catastrophic for patients which are misclassified. Hence it is important that the classifier is updated and tested regularly to minimize the models being biased. In this thesis, the balance of gender, age and ethnicity among patients are not specified, hence selection bias should be assumed. If this kind of classifier is to be introduced into hospitals; elimination of these biases is necessary.
Recommendations for decreasing bias in general include; acknowledgement that biases most likely are to occur anyway due to the difficulty of completely representing the world population and perfectly collect balanced data. Secondly one can adhere to diagnostic criteria and use statistical rules to predict behavior and detect unbalanced data before data collection is finished [9].
Furthermore, the model should be considered as an aid in the classification process rather than an absolute truth. Even though a high recall score might be reached, every single case of detected hyporeactivty should be evaluated and confirmed by an expert to avoid misclassification and mistreatment. Therefore, the main usage of the model created from this thesis should be considered as a detector allowing for earlier investigations from experts.

6
Conclusion

In conclusion, reactivity and hyporeactivity can be detected with machine learning accurately. The obtained results were better than the baseline classifier and are surely a sign of the potential of machine learning. A classifier of this kind should only be considered as a guide and not an absolute judge.
Since the classifier's goal is to find hyporeactive patients for the purpose of earlier treatment by automating a complex manual analysis and thereby saving valuable time of experts, the classifier seeks to warn the expert that the patient is hyporeactive and needs immediate further attention. A precision of 82% was achieved, meaning that the classifier is correct 82% of the time, with regard to the recall score, allowing faster treatment for almost $\frac{3}{4}$ of suicidal patients in immediate need. However, no matter the success rate of the classifier, prescribing heavy drugs to patients is a serious issue and manual confirmation of the detected suicidal patients is necessary in any case.

Bibliography

[1] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[2] Wolfram Boucsein. *Electrodermal Activity*. Springer, second edition, 2012.

[3] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[4] Richard J Davidson, Klaus R Sherer, and H Hill Goldsmith. *Handbook of affective sciences*. Oxford University Press, 2009.

[5] Michael Dawson, Anne Schell, and Diane Filion. The electrodermal system. pages 200–223. Cambridge University Press, second edition, 2000.

[6] Robert Edelberg. Electrodermal recovery rate, goal-orientation, and aversion. *Psychophysiology*, 9(5):512–520, 1972.

[7] Gunnar Edman, Marie Åsberg, Sten Levander, and Daisy Schalling. Skin Conductance Habituation and Cerebrospinal Fluid 5-Hydroxyindoleacetic Acid in Suicidal Patients. *Archives of General Psychiatry*, 43(6):586–592, 06 1986.

[8] Bryn Farnsworth. *Skin Conductance Response – What it is and How to Measure it*, 2019 (accessed June 04, 2019). `https://imotions.com/blog/skin-conductance-response/`.

[9] Howard N. Garb. Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice*, 4(2):99–120.

[10] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi. cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 63(4):797–804, April 2016.

[11] Rui Guo, Shuangjiang Li, Li He, Wei Gao, Hairong Qi, and Gina Owens. Pervasive and unobtrusive emotion sensing for human mental health. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 436–439. IEEE, 2013.

[12] Chih-Wei Hsu, Chih-Chung Chang, and CJ Lin. A practical guide to support vector classification technical report department of computer science and information engineering. *National Taiwan University, Taipei*, 2003.

[13] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014.

[14] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*, volume 26. Springer, New York, 01 2013.

[15] Hindra Kurniawan, Alexandr V Maslov, and Mykola Pechenizkiy. Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 209–214. IEEE, 2013.

[16] Douglas F Levinson and Robert Edelberg. Scoring criteria for response latency and habituation in electrodermal research: A critique. *Psychophysiology*, 22(4):417–426, 1985.

[17] D Makowski. Neurokit: A python toolbox for statistics and neurophysiological signal processing (eeg, eda, ecg, emg...). *Memory and Cognition Lab, Paris, France*, nov 2016.

[18] Python Software Foundation. *Anaconda Software Distribution - Version: 5.3.0*, 2019.

[19] Python Software Foundation. *Spyder - Version: 3.3.1*, 2019.

[20] John O. Rawlings. *Applied regression analysis : a research tool*. Springer, 1998.

[21] Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, and Oscar Martinez Mozos. Stress detection using wearable physiological sensors. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 526–532. Springer, 2015.

[22] Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, and Oscar Martinez Mozos. Stress detection using wearable physiological sensors. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 526–532. Springer, 2015.

[23] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[24] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2009.

[25] E N Sokolov. Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, 25(1):545–580, 1963. PMID: 13977960.

[26] Evgeny N Sokolov, John A Spinks, Risto Näätänen, and Heikki Lyytinen. *The orienting response in information processing.* Lawrence Erlbaum Associates Publishers, 2002.

[27] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manage.*, 45(4):427–437, July 2009.

[28] John A Stern and Larry C Walrath. Orienting responses and conditioning of electrodermal responses. *Psychophysiology*, 14(4):334–342, 1977.

[29] Romain Tavenard. tslearn: A machine learning toolkit dedicated to time-series data, 2017. `https://github.com/rtavenar/tslearn`.

[30] L. H. Thorell. Electrodermal activity in suicidal and nonsuicidal depressive patients and in matched healthy subjects. *Acta Psychiatr Scand*, 76(4):420–430, Oct 1987.

[31] L.-H. Thorell. Valid electrodermal hyporeactivity for depressive suicidal propensity offers links to cognitive theory. *Acta Psychiatrica Scandinavica*, 119(5):338–349, 2009.

[32] Gaetano Valenza, Antonio Lanata, and Enzo Pasquale Scilingo. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE transactions on affective computing*, 3(2):237–249, 2012.

[33] Egon L van den Broek, Joris H Janssen, Marjolein D van der Zwaag, and Jennifer A Healey. Prerequisites for affective signal processing (asp)-part iv. In *Biosignals*, pages 188–193, 2010.

[34] Vladimir Vapnik. The support vector method of function estimation. In *Nonlinear Modeling*, pages 55–85. Springer, 1998.

[35] Vladimir Vapnik. *The nature of statistical learning theory.* Springer science & business media, 2013.

[36] Peter H Venables and Margaret J Christie. Electrodermal activity. *Techniques in psychophysiology*, 54(3), 1980.

[37] PH Venables and DA Mitchell. The effects of age, sex and time of testing on skin conductance activity. *Biological psychology*, 43(2):87–101, 1996.

[38] Auguste Vigouroux. *Etude sur la résistance électrique chez les mélancoliques.* PhD thesis, Faculté de médecine de Paris, 1890.

[39] Roberto Zangróniz, Arturo Martínez-Rodrigo, José Manuel Pastor, María T. López, and Antonio Fernández-Caballero. Electrodermal activity sensor for classification of calm/distress condition. *Sensors*, 17(10), 2017.

A

Appendix

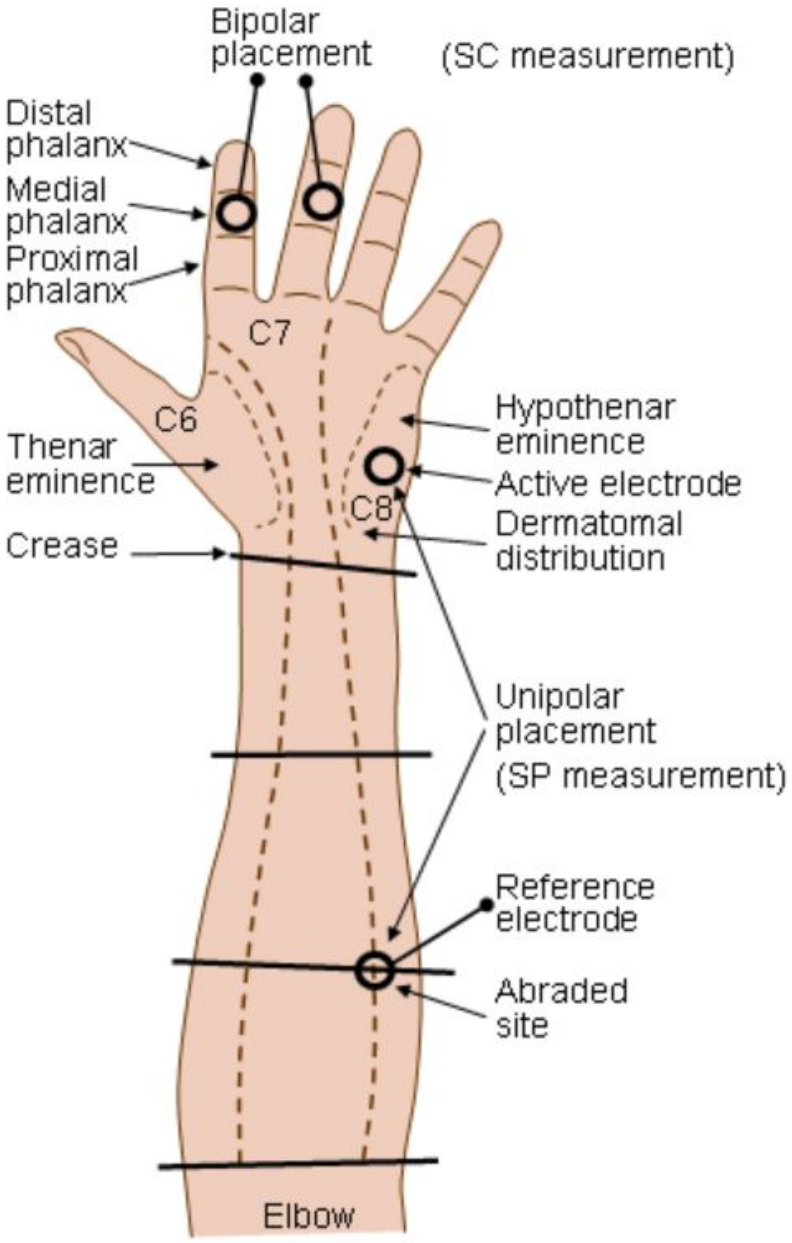


Figure A.1: *The placement of electrodes to record EDA, image borrowed from Electrodermal Activity 2nd edition, Wolfram Boucsein Springer [2].*