Karlsruhe Institute of Technology - KIT Department of Economics and Management

Institute for Operations Research (IOR)

Analytics and Statistics

Prof. Dr. Oliver Grothe

Classification of Schizophrenia based on Activity Time Series using Hidden Markov Models

Master Thesis

of

Matthias Boeker
Matr. Nr.: 1675534
Industrial Engineering and Management

Date of delivery
21.12.2020

Supervision by:
Simula Metropolitan Center for Digital Engineering
Dr. Michael Riegler
Dr. Paal Halvorsen
Dr. Hugo Lewi Hammer

Contents

List of Figures

List of Tables

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, 21.12.20

....
(Matthias Boeker)

1 Introduction

Worldwide, 11 to 20 persons are diagnosed with schizophrenia per 100.000 people every year - independent of socioeconomic status. In Germany, 19 new cases are diagnosed per 100.000 inhabitants, each year Gaebel and Wölwer, 2010. Schizophrenia is characterised by a pattern of disturbances in the areas of attention, perception, thinking, ego function, drive, affectivity, and psychomotor function (Gaebel and Wölwer, 2010).

The diagnosis of schizophrenia is based on criteria that are summarised in internationally used diagnostic systems. The Diagnostic and Statistical Manual of Mental Disorders is one of the diagnostic systems (Gaebel and Wölwer, 2010). The diagnosis of schizophrenia is made, if a patient shows, for one month or longer, at least one symptom from the areas ego disorders, delusions, and hearing voices, or two symptoms from the domains of other hallucinations, formal thought disorders, psychomotor symptoms, or negative symptoms (Saß, 2007).

Social disabilities of schizophrenics exist in particular in the areas of communicative behavior and social competence Saß, 2007. The attitude and behavior of the social environment towards schizophrenic persons has also a significant influence on their quality of life. Schizophrenic persons are often stigmatized and discriminated against by the public and the media (Gaebel et al., 2005).

According to a study by (Palmer et al., 2005), schizophrenic patients are at a higher risk for suicides. The study estimates that about 4.9% of schizophrenic patients will commit suicide in the course of their lives. Extrapolated to a country level, that would be 5000 suicides per 100,000 inhabitants. For comparison, in Germany 2019, a total of 10 suicides were committed per 100.000 inhabitants (Destatis, 2019).

This work presents classification variables derived from the parameters of a Hidden Markov model to improve the identification of the disease of schizophrenia by a persons' activity.

1.1 Problem Statement

Mental diseases such as depression or schizophrenia are diagnosed based on self-reported experiences, questionnaires, and psychiatric assessments.
The criteria for a diagnosed mental disease are stated in the Diagnostic and Statistical Manual of Mental Disorders (Carr et al., 2020). Many of these criteria include catatonic or disorganized behavior and behavioral abnormalities (Sano et al., 2012). These criteria are embedded in the activity record of a person. Especially the disruption of the circadian rhythm is a common indicator for psychiatric disorders (Wulff et al., 2010).
The circadian rhythm or circadian cycle refers to the bodies inner clock. The circadian rhythm refers to rhythmic biological cycles that follow an approximately 24 hours course (of Health and Service, n.d.). The rhythm is heavily influenced by light and the production of melatonin, a hormone that causes drowsiness (of Health and Service, n.d.).
A quantitative study of a person's activity over time can provide insight into the diagnosis of mental diseases.
Actigraphy is a method of analyzing human activity, applied to quantify activity. The method is non-invasive and involves wearing some sort of portable advice over a certain amount of time (Morgenthaler et al., 2007). The usage of actigraphy increased in the research literature on sleep and circadian rhythm. It can be used as a diagnostic instrument in the assessment of sleep disorders and circadian rhythm disorders (Morgenthaler et al., 2007).

1.2 Approach

The measured activity by actigraphy is the data basis for this work. Based on the quantified activity, the goal of this work is the classification of schizophrenic and non-schizophrenic persons. The overall goal of this work is a classification of the activity time series, to predict whether a person is schizophrenic or not.

Mental diseases cause disruptions of the circadian cycle and other behavioral abnormalities. Accordingly, the first goal is to model the resting and active periods of a person. Different Hidden Markov models are used to model the two states of activity.By modeling resting and active periods, it is assumed to find valuable information for predicting the mental condition of a subject. The difference between schizophrenic and non-schizophrenic persons is assumed to be comprised within the model parameters of the Hidden

Markov model. The Hidden Markov model is chosen as it has already been successfully applied to model states of activity on actigraphy time series, acceleration data or physiological data (Carr et al., 2020), (Xu et al., 2020), (Witowski et al., 2014), (Witowski, 2018), (Huang et al., 2018).

Hidden Markov models are applied to identify whether a person is asleep or active according to their activity measurements. Thus, each observation of activity belongs to a certain state of activity.
Other mathematical models like in (Skeldon et al., 2014) describe the dynamics of active and resting periods as a harmonic oscillator. The harmonic oscillator, which follows a 24 hours course, can be described by a trigonometric function. The trigonometric function can be used to model the underlying circadian cycle of a person. The trigonometric function is continuous and cannot be used to directly infer the status of activity. However, the oscillating behavior can certainly describe the probability of the activity states.

The idea is to extend a Hidden Markov Model by modeling its probabilities of being in a certain state as time-dependent and describing it by a trigonometric function. The likelihood of being in the state of rest increases during the night and decreases during the day. Assuming that schizophrenic patients tend to show circadian rhythm disturbances and behavioral abnormalities, it is interesting to investigate the time-dependent transition probabilities between the two different states, active and resting.

To obtain a time changing transition probability and to incorporate a trigonometric function in the transition probabilities, an extension of the Expectation-Maximization Algorithm, called Baum-Welch, for Hidden Markov models is presented. The Baum-Welch algorithm makes use of the time-invariance property of Markov Chains to optimize the likelihood (Banachewicz et al., 2008). However, it is assumed that the time structure contains valuable information for the classification. Thus, the transition probabilities are modeled as a function of time. A link function connects a covariate X_t and the elements of the transition probability matrix Q_t so that a trigonometric function can be included.

Four different hypotheses are presented and examined in this paper. Each hypothesis addresses the extent to which the difference between schizophrenic and non-schizophrenic individuals is reflected in the model parameters of a Hidden Markov model.

Hypothesis 1 (H1) *The distribution of activity during the resting and active period are different between schizophrenic and non-schizophrenic subjects. The parameters of the state probability distribution represent a powerful predictor.*

The active and resting-state are modeled as hidden variables within the Hidden Markov model. Each hidden variable is described by a different probability distribution from the same distribution family. The probability distribution parameters are assumed to be a potentially good feature for classification.

Hypothesis 2 (H2) *The transition probabilities of a Hidden Markov model are different between schizophrenic and non-schizophrenic subjects and can be used as a significant feature.*

Hidden Markov models provide estimates of the probability to change between different states or stay in a certain state. These probabilities are called transition probabilities. Hypothesis 2 states that the transition probabilities might differ between schizophrenic and non-schizophrenic subjects in a way that these probabilities can predict the mental state of a person. An indication of hypothesis 2 is the fact, that the circadian cycle is disrupted for persons with schizophrenia.

Hypothesis 3 (H3) *Fluctuations in the time depending transition probabilities will give valuable information about the difference between schizophrenic and non-schizophrenic subjects and can be used as a significant feature.*

An extension of the Hidden Markov model is introduced within this work. The extension allows the Hidden Markov model to have time depending transition probabilities. In the classic Hidden Markov model, the transition probabilities are modeled as constant over time. Since a time dependence is allowed in the extension, the analysis of the fluctuation of the transition probabilities captured by their variance is analyzed.

Hypothesis 4 (H4) *The circadian rhythm can be approximated by trigonometric function. Thus, a trigonometric function is integrated into the time-dependent transition probabilities by link function and corresponding link coefficients. It is assumed that the link coefficients contain the disruptions from the circadian cycle of schizophrenic compared to non-schizophrenic subjects. contain*

The extension of the Hidden Markov model allows time depending transition probabilities. Moreover, a covariate can be integrated into the time course of the transition probabilities. A trigonometric function is integrated as a covariate to include a model for the circadian cycle. The covariate is integrated with the help of a link function. The link parameters of the link function are assumed to comprise valuable information for classification.

The model parameters are suggested features for the classification of the time series between schizophrenic and non-schizophrenic subjects. Logistic regression is applied as a classification method. The question is whether the proposed features improve the classification compared to the features already used in the literature. To answer this question, features from literature are used as a baseline comparison. A comparative study on the different classification features, derived from the literature and the parameters of the Hidden Markov model, is conducted. Through this study, the validity of the four hypotheses put forward will be assessed.

In a nutshell, three different Hidden Markov models, the time-independent, time-dependent, and time-dependent model with an integrated covariate, to model the rest/active pattern of a person. The time-dependent models and the according to Expectation-Maximization (EM) algorithm are implemented within the scope of this work, as there are not provided by any Python package. The performance of different model parameters of the Hidden Markov models, stated within the hypotheses 1 to 4, are evaluated as features for the time series classification by logistic regression.

1.3 State of the Art

The analysis of sleep/wake patterns and the quantification of the circadian rhythm has been based on actigraphy. Studies have shown that actigraphy is useful in detecting sleep patterns, sleep disorders or even neurobehavioral disorders (Sadeh, 2011), (Ancoli-Israel et al., 2003). However, the relationship between sleep/wake patterns and mental disorders, especially severe mental disorders such as schizophrenia, based on actigraphy has not been sufficiently studied. (Tahmasian et al., 2013).

The general assessment of sleep bases on quantitative and qualitative methods. Actigraphy is one of the quantitative methods. Throughout the literature, different assessment tools but actigraphy are used. A common qualitative assessment tool is the Pittsburgh sleep quality index questionnaire (Tahmasian et al., 2013). The questionnaire helps on a subjective base to assess for instance the sleep quality, sleep latency, and sleep efficiency (Buysse et al., 1989).

A great part of research focused on a combined quantitative and qualitative analysis of the sleep/wake pattern. Due to the additional qualitative analysis, mostly basic quantitative methods are applied. Commonly the total activity, the mean activity or the standard deviation of the daily and nightly activity are measured (Afonso et al., 2014), (Apiquian et al., 2008), (Docx et al., 2013), (Kume et al., 2015), (Lindamer et al., 2008), (J. L. Martin et al., 2005), (Robillard et al., 2015).

Besides mean and standard deviation, more elaborate methods for quantitative analysis have been used in the literature. J. L. Martin et al., 2005 fits a cosine function to the activity time series and extracts the parameters. Characteristics of wake/sleep patterns like the intraday variance (IV) and the interday stability (IS), which are introduced by Witting et al., 1990 are applied by Berle et al., 2010 to differentiate between schizophrenic and non-schizophrenic participants of a study. Witting et al., 1990 also introduced the measure of the most active 10 hours (M10) and the least active 5 hours (L5). These measures are applied by Wulff et al., 2010 in a study analyzing sleep/wake patterns of schizophrenic patients. Sano et al., 2012 evaluated the cumulative distribution of the wake and rest periods, which are defined by some threshold. Hauge et al., 2011 applied Fourier Analysis to analyze the variance between schizophrenic and non-schizophrenic persons in low and high-frequency ranges.

Hidden Markov models as an unsupervised learning algorithm have already been used in the literature to model the sleep/wake pattern based on activity time series like in (Domingues et al., 2013) and (Li et al., 2020).

Albert et al., 2020 applied an Hidden Markov model for toddler activity classification. Besides sleep/wake patterns, Albert et al., 2020 distinguished different movement patterns like crawling, walking, being carried, etc. based on the toddlers' recorded activity. Liu et al., 2019 applied a Hidden Markov model on activity time series in combination with the heart rate measured by commercial devices.

Huang et al., 2018 introduced a Hidden Markov model to improve the classification between wake and sleep by incorporating the cosine structure of the circadian rhythm (Huang et al., 2018). Huang et al., 2018 identified three states of activity, resting, active, and highly active, to study the effect of multidrug chemotherapy on patients at home (Huang et al., 2018). This work bases on the idea by Huang et al., 2018. Carr et al., 2020 based their analysis on the same ground as Huang et al., 2018. Carr et al., 2020 used the classification of the wake/sleep pattern to predict the depression score of patients with bipolar disorder.

The here presented approach continues on the application of the Hidden Markov model to identify wake/sleep states by using a similar approach as Huang et al., 2018 and Carr et al., 2020.
Other than Carr et al., 2020, this work classifies schizophrenic and non-schizophrenic persons. Carr et al., 2020 used the different identified wake/sleep states to extract directly interpretable features, like estimated sleep time, average inactive periods, or longest period inactive. In this work, the features for the classification are model parameters from the Hidden Markov model like the transition probabilities. Additionally, this work tries to give an interpretation of the model parameters.

The PSYKOSE - Motor Activity Database of Patients with Schizophrenia is a data set provided by Jakobsen, Ceja, et al., 2020 and was introduced in February 2020. The work by cites jakobsen2020psykose compared different machine learning classifiers to predict schizophrenia. The classification is performed on three basic statistical features Jakobsen, Ceja, et al., 2020. Jakobsen, Ceja, et al., 2020 provides the only other analysis with the same overall goal on the same data, as this thesis. Thus, the results of Jakobsen, Ceja, et al., 2020 and the results of this work will be discussed in chapter 7.

1.4 Data Description

The data set contains actigraphy data collected from patients with schizophrenia. There are 54 participants in total. 22 recorded persons are diagnosed with schizophrenia and 32 persons are in the control group. Sensor data collected over several days are available for each person. In addition to the sensor data, demographic data and medical assessments during the observation period are available. The activity is measured by an actigraphy device called Actiwatch model AW4 provided by Cambridge Neurotechnology Ltd, England. The device records the intensity of its acceleration along the x,y, and z axes (Jakobsen, Ceja, et al., 2020). The activity is recorded at a frequency of 32 Hz. The data contains recorded acceleration for every minute Jakobsen, Ceja, et al., 2020. The recording duration varied from person to person and ranged from 9 to 20 days per person Jakobsen, Ceja, et al., 2020. The data set is publicly available.

2 Theory

The activity time series are modeled with three proposed Hidden Markov models. The concept of a stochastic process is briefly introduced to comprehensively explain the Hidden Markov Models based on it. The fitted parameters of each model are used for the least absolute shrinkage and selection operator logistic regression in order to classify schizophrenic and non-schizophrenic persons. According to the two steps, the theory on the different applied models is introduced.

To evaluate the classification performance of the model parameters as a classification feature, a baseline classification is performed. The features of the baseline classification are derived from the literature. A small-scale literature review on classification features is presented at the end of this chapter.

The theory chapter first gives a brief introduction to stochastic processes. The aim is to create a common basis for the introduction of the Markov process and the advanced Hidden Markov Model. After the basic concept of the models is demonstrated, the algorithm for parameter estimation of a Hidden Markov model is explained. This includes the most important equation on which the algorithm was implemented for this thesis. Finally, the least absolute shrinkage and selection operator logistic regression is briefly introduced.

2.1 Brief Introduction to Stochastic Processes

A Hidden Markov model is an extension of the Markov Chain which describes a random process with certain properties. A random process is a set of random variables that are assigned to an index set T.
The set of random variables is defined on a common probability space $(\Omega, \mathcal{F}, P)$. The probability space is a triple of the sample space Ω, the event space $\mathcal{F}$ and the probability function P. Considering a stochastic experiment of flipping a coin, the sample space Ω is the set of every possible outcome - heads or tails. The event space $\mathcal{F}$ is the actual observed outcomes of the experiment. The probability distribution P assigns a probability to each event.

The index set T is usually linear ordered and can be interpreted as time. But other interpretations e.g. the sequence of text are also possible.
The random process can be defined as $\{X(t)\}_{t \in T}$. The short notation is

X_n where n substitutes t. The index set T can be continuous or discrete. Which values the random numbers will take depends on the probability distribution to which they are following. In the context of the given data, the sample space Ω is every possible activity time series generated by a person. The event space $\mathcal{F}$ is a set of the observed activity time series, which are here considered as a database. The process is usually defined by its underlying probability distribution. In this work, it is assumed that the stochastic process is Gaussian, thus follows a Gaussian distribution. However, other distributions are possible. Some examples are introduced.

The Poisson process for instance can be interpreted as a counting process. A counting process would answer the question of how many instances are counted at a time. The number of points in a set has the Poisson distribution, see Last and Penrose, 2017, pp. 273.
The Wiener process or Brownian motion is a well-known example of a Gaussian Process, see Brémaud, 2020, pp. 206. The Wiener process is a continuous Gaussian process $\{X(t)\}_{t\in\mathbb{R}_+}$ with independent and standard normal distributed increments. This leads to the definition of the Gaussian process itself which is assumed to govern the given observations in this work.
The definition of a Gaussian process is based on the definition by Brémaud, 2020.
A real-valued stochastic process $\{X(t)\}_{t\in\mathbb{R}}$ assigned to an arbitrary index set T is called a Gaussian process if for all $n \geq 1$ and for all $t_1, ..., t_n \in T$, the random vector $(X(t_1), ..., X(t_n))$ is Gaussian, see Brémaud, 2020, pp. 206. Every finite collection of random variables within that random variable have a multivariate normal distribution.

In a nutshell, a random process $\{x_1, ..., x_t, ...x_n\}$ is a sequence of independent and identically distributed random variables taking values of some set S at time t, see Billingsley, 1961, pp.12. Based on the general stochastic processes, the Markov process is introduced in the following chapter.

2.2 Markov Process

A stochastic process is called a Markov process when it fulfills the Markov properties. The Markov property states that the future is independent of the past, given the present. The information from the past is already embedded in the present, see Bengio, 1999. According to that, the most information there is about the future is within the present. This statement can be formalized as an equation and a graph. The definition varies but in this work, a discrete stochastic process fulfilling the Markov property is referred to as Markov Chain. A continuous stochastic process fulfilling the Markov property will be referred to as the Markov process. It is assumed that all observations are independent of previous ones but the most recent, see Rabiner, 1989. This statement can be applied to equation (1) to derive the formal Markov property.

Let $\{X\}_n$ be a time-discrete stochastic process of random numbers taking values of a finite set S. The finite set S will be referred to as the state space and the elements in S are referred to as states i, j, see Brémaud, 2020, pp. 221. If the process is in state i at time t, it is noted as $X_t = i$.

The joint distribution of a sequence of observations can be described by Bishop, 2006, pp.607

$$P(X_1, ..., X_N) = \prod_{n=1}^{N} p(x_n | x_1, ..., x_{n-1}) \tag{1}$$

The Markov property is given if for all indices in T and all states of $i, i_{n-1}, ..., i_0, j, j_{n-1}, ..., j_0$ in S the equation (2) counts.

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = j | X_n = i) \tag{2}$$

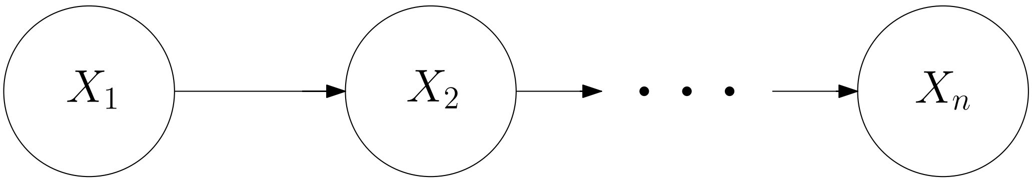


Figure 1: Visualization of the probability dependence in a first order Markov Chain. Source: own illustration

The Markov property of equation (2) is defined as following. The conditional probability of being in state j at time $n+1$, given the information on all previous states, is equal to the conditional probability of being in state j at time $n+1$, X_{n+1} given only the previous state $X_n = i$. The probabilistic dependence on the past states is only connected to the future through the present state, see Brémaud, 2020, pp. 221. Consequently, all the necessary information about the future is assumed to be in the present state. Figure 1 shows the dependence structure visualised as a chain.

Figure 1 displays a first-order Markov Chain. It is possible to have m order Markov Chains, which means the future state conditions on the m past states. The more advanced Hidden Markov process is introduced in the subsequent chapter.

2.3 Hidden Markov Process

Markov Chains describe changing systems over time. The system generates observations and can be e.g. a patient, the weather, a sensor, etc. The Markov Chain are limited. It expects that the system is fully observable, see Rabiner, 1989 and Keselj, 2009. This limitation is useful but very restrictive, too.

To overcome this restriction the observations are modeled as a probabilistic function of the state (Rabiner, 1989). A system is assumed to generate two different stochastic processes, which are related. There is the observed stochastic process, which generates our data. BesidesAn, there is a latent stochastic process, which can only be deduced from the observable process. Figure 2 represents the graphical structure of the two stochastic processes. For each observable random variable X_n a latent variable Z_n is introduced. Only the the latent or hidden stochastic process fulfills the Markov property (Bengio, 1999). Hence, the distribution of the observable variables $p(x_n|x_1, ..., x_{n-1})$ does not underlie any conditional independence property see Bishop, 2006, pp.607.

In the following the joint distribution of the introduced Hidden Markov model is derived, see Bilmes et al., 1998, Bishop, 2006 and Bengio, 1999.
Let $Z_1, ...Z_N$ be the discrete multinomial random variables of the stochastic process that satisfies the Markov property. Let $X_1, ..., X_N$ be the observed random variables. Notice, when the small notation of random variables is used, the work refers to the observation of the random variable.

$$P(X_1, ..., X_n, Z_1, ..., Z_N) = P(Z_1) \prod_{n=2}^{N} P(Z_n|z_{n-1}) \prod_{n=1}^{N} P(X_n|Z_n) \tag{3}$$

From the joint probability the elements of an Hidden Markov Model can be derived.

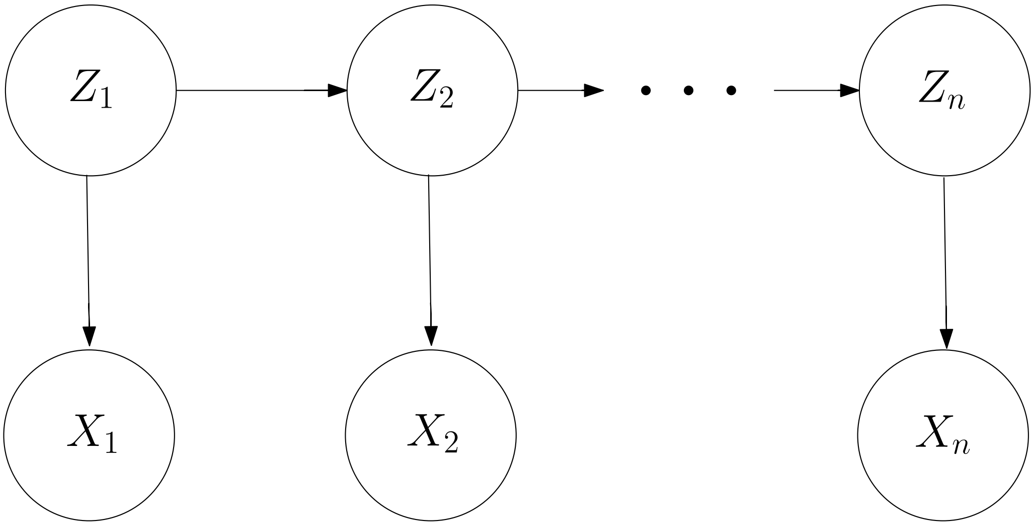


Figure 2: Visualization of the Hidden Markov Chain where X_t represents the observed random variables. Z_t is the hidden/latent variable of the state of the process. Source: own illustration

The conditional probability $P(\mathbf{X}|\mathbf{Z})$ is called emission probability. The emission probability is a probability distribution for each time step on X given the state Z and are referred to as

$$b_i(n) \coloneqq P(\mathbf{X}|Z_n = i) \tag{4}$$

Consider the case $b_i(n) = P(\mathbf{X}|Z_n = i)$ where the emission probabilities represent the probability that the observation at time t was generated by state i.
Consequently, an Hidden Markov model can be interpreted as a mixture model for a certain time step, see Bishop, 2006, pp. 607. The probabilities might be given as Gaussian, hence a Gaussian Mixture model.

The initial state has a special status, since it does not have a previous state. The initial state can be graphically is derived in figure 2. A initial state distribution has to be given Bengio, 1999. The initial state probability $\pi = \{\pi_i\}$ is given by Rabiner, 1989

$$\pi(i) \coloneqq P(Z_1 = i) \tag{5}$$

The initial state probability is considered as prior knowledge (Bengio, 1999). The probabilities are subject to $\sum_i \pi_i = 1$.

The term $P(Z_n|Z_{n-1})$ refers to the probability to transmit from one state into another in consecutive time steps. The transmission probabilities can be expressed in a matrix $A = \{a_{ij}\}$. Depending on the number of states k, the transition matrix becomes the size $A \in \mathbb{R}^{k \times k}$. The equation is given in the following,

$$a_{ij} = P(Z_n = i|Z_{n-1} = j) \tag{6}$$

The transition probabilities express the likelihood of switching from state j to state i, for a next time step.

The above-described elements represent the structure of the Hidden Markov model. Thus, the Hidden Markov model can be described by the triple $\lambda = (\pi, A, b)$, see Rabiner, 1989. Given the model structure, there are three main problems, see Rabiner, 1989, Bishop, 2006, pp.607 and Keselj, 2009.

1. **Likelihood**
   Given the observations $x_1, ..., x_N$ and a model λ determine the probability $p(X|\lambda)$ that a sequence of observations is generated by the model.

2. **Decoding**
   Given a model λ and a sequence of observations X find the corresponding state sequence.

3. **Learning**
   Given the observations $x_1, ..., x_N$ how can the model parameters adjusted $\lambda = (\pi, A, b)$ in order to maximize $p(X|\lambda)$, the likelihood of the observation given a Hidden Markov model. The problem of model training is faced with this problem.

The problem of the likelihood is solved through the Forward algorithm, see Rabiner, 1989. This algorithm is also applied to the problem of learning. There will be a brief introduction to the Forward algorithm. The problem of decoding is solved with the Viterbi algorithm, see Viterbi, 1967 and see Forney, 1973. The Viterbi algorithm is implemented and applied in this work but there is a special focus on the solution of the learning problem through the Baum-Welch algorithm, see Baum et al., 1970 see Bilmes et al., 1998. Hence, the Viterbi algorithm is not introduced in depth.

The counterpart to the Forward algorithm is the Backward algorithm. Both are often introduced together as the Forward-Backward algorithm. The next chapter will introduce the Forward-Backward algorithm.

2.4 The Forward-Backward Algorithm

As the name indicates, the Forward-Backward Algorithm consists of two parts, the forward and the backward part. The goal of both parts is to calculate the probability of the occurrence of a certain observation sequence given a certain model $P(\mathbf{X}|\lambda)$. This conditional probability can be reformulated. To do so, two other probabilities are introduced.
$P(\mathbf{X}|\mathbf{Z},\lambda)$ is the probability of a observation sequence for a given state sequence. This can be defined as the product over the time steps of the emission probabilities, see Rabiner, 1989.
Let's recall, the emission probabilities $b_{Z_n}(X_n)$ represent the probability that the observation at time n is generated by the state Z_n at time n Consequently, the probability of an observation sequence for a given state sequence is given by

$$P(\mathbf{X}|\mathbf{Z},\lambda) = b_{Z_1}(X_1)b_{Z_2}(X_2)...b_{Z_N}(X_N) \tag{7}$$

Let $P(\mathbf{Z}|\lambda)$ be the probability of a certain state sequence conditioned on a certain model. The probability of a state sequence can be written as the product over all time steps n of the transition probabilities and the initial state probability Rabiner, 1989.

$$P(\mathbf{Z}|\lambda) = \pi_{Z_1}a_{Z_1Z_2}a_{Z_2Z_3}...a_{Z_{T-1}Z_T} \tag{8}$$

The sum over all time steps of the product of these just introduced probabilities leads to the targeted $P(\mathbf{X}|\lambda)$.

$$P(\mathbf{X}|\lambda) = \sum_{\mathbf{Z}} P(\mathbf{X}|\mathbf{Z},\lambda)P(\mathbf{Z}|\lambda) \tag{9}$$

The computation of equation (9) takes $2N \cdot m^N$ calculations, where m is the number of states, see Rabiner, 1989. The term grows exponentially with the length of the stochastic process N, which makes it for a certain length computational too expensive.
To overcome this problem, the Forward-Backward algorithm introduces a recursion and in the manner of dynamic programming, breaks the problem down into sub-problems. The forward part will be introduced first.

The forward variable is introduced as

$$\alpha_n(i) \coloneqq P(X_{1:n}, Z_n = i|\lambda) \tag{10}$$

Equation (10) can be interpreted as the joint probability of the observation sequence up to observation n and the hidden variable Z_n being in state i, conditional on the given model. The probability can be also formulated as

$$P(X_{1:n}, Z_n = i|\lambda) = \sum_{Z_{n-1}=1}^{m} P(X_n|Z_n)P(Z_n|Z_{n-1})P(Z_{n-1}, X_{1:t-1}). \tag{11}$$

The conditional probabilities introduced in equation (11) are all known through the model λ.

- $P(X_n|Z_n)$ is the emission probability at time n.
- $P(Z_n|Z_{n-1})$ is the transition probability.
- $P(Z_{n-1}, X_{1:t-1})$ is the forward variable $\alpha_{n-1}(i)$ for the previous time step, see Rabiner, 1989.

$\alpha_n(i)$ can be solved recursively. The Forward algorithm consists of three parts, the initialization, the recursion and the termination.

1. **Initialization**

$$\alpha_i(1) = \pi_i b_i(1), \qquad 1 \leq i \leq m \tag{12}$$

2. **Recursion**

$$\alpha_i(n+1) = \left[\sum_{j=1}^{m}, \alpha_i(n)a_{ij}\right] b_j(X_{n+1}), \qquad 1 \leq n \leq N, \quad 1 \leq i \leq m \tag{13}$$

3. **Termination**

$$P(\mathbf{X}|\lambda) = \sum_{i=1}^{m} \alpha_N(i) \tag{14}$$

In step 3. of the Forward algorithm, the desired probability of the observations given the model are calculated. It is defined as the sum over all possible states of the last iteration step in N.

The introduced Forward algorithm only uses $m^2 \cdot N$ calculations instead of $2N \cdot m^N$ (Rabiner, 1989). Since the Forward algorithm introduces the forward variable as joint probability of the observation sequence up to observation n, the Backward algorithm follows the same procedure but backwards.
The backwards variable is defined as

$$\beta_n(i) \coloneqq P(X_{n+1:N} | Z_n = i, \lambda). \tag{15}$$

The backward variable indicates the probability of the partial observation sequence from $n+1$ to N given the state i at time n. The Backwards algorithm is solved recursively, too. The initialization step sets $\beta_i(N) = 1$, Rabiner, 1989.

1. **Initialization**

$$\beta_i(N) = 1, \qquad 1 \leq i \leq m \tag{16}$$

2. **Recursion**

$$\beta_i(n) = \sum_{j=1}^{m} a_{ij} b_j(X_{n+1}) \beta_j(n+1), \qquad n = N-1, N-2, ..., 1, \quad 1 \leq i \leq m \tag{17}$$

3. **Termination**

$$P(\mathbf{X}|\lambda) = \sum_{i=1}^{m} \beta_i(1) \pi_i b_i(X_1) \tag{18}$$

Both algorithms calculate the desired probability $P(\mathbf{X}|\lambda)$ in different ways.

The forward and backward variable will be useful for the Baum-Welch algorithm, which couples the Forward-Backward algorithm with the Expectation-Maximization algorithm. The Baum-Welch algorithm is introduced in the subsequent chapter.

2.5 The Baum-Welch Algorithm

The Baum-Welch algorithm approximately derives the maximum likelihood estimate of the model parameters of the Hidden Markov model. The Baum-Welch algorithm follows the structure and idea of the Expectation-Maximization algorithm (EM) (Bilmes et al., 1998). The EM is applied, when the maximum likelihood estimate cannot be derived analytically. This is the case, when the data is incomplete, see Bilmes et al., 1998 and Baum et al., 1970. The case of incomplete data is also given when the model includes unobserved variables, called latent or hidden variables. This applies to the Hidden Markov model.

Basic EM Algorithm

The EM consists of two steps, the expectation step, and the maximization step. With the two steps, the maximum likelihood is derived iteratively. These two steps will be introduced in more detail later. The EM guarantees that the likelihood function will increase and converge to a local maximum Bilmes et al., 1998.

Let $\mathbf{X}$ be the observed but incomplete data. It is assumed that there exists a complete data set $\mathbf{Y} = (\mathbf{X}, \mathbf{Z})$. $\mathbf{Z}$ is the set of latent or hidden variables, see Bilmes et al., 1998.
Given the complete set, let $p(\mathbf{Y}|\theta) = p(\mathbf{X}, \mathbf{Z}|\theta)$ be the joint probability distribution of observed $\mathbf{X}$ and unobserved variables $\mathbf{Z}$, given a parameter set $\theta \in \Theta$. The joint probability distribution is formulated in equation (19) as the product of the marginal distribution of $\mathbf{X}$ and the conditional probability of the latent variables $\mathbf{Y}$ given the observed ones $\mathbf{X}$ Bilmes et al., 1998.

$$p(\mathbf{X}, \mathbf{Z}|\theta) = p(\mathbf{Z}|\mathbf{X}, \theta)p(\mathbf{X}|\theta) \tag{19}$$

From equation (19) the likelihood of the complete data $L(\theta|\mathbf{Y})$ set is derived, as well as the likelihood of the incomplete data $L(\theta|\mathbf{X})$.

The goal of the EM is to find the parameter set which maximizes the likelihood of the incomplete data set. Due to the incompleteness, this cannot be done analytically. The EM first finds the expectation of the likelihood of the complete set, given a parameter set. Then the EM derives new parameters which maximize the likelihood of the expected value, see Bilmes et al., 1998.

The EM algorithm starts with an initial guess of the parameter set. The guess can be arbitrary or include prior knowledge.
After initialization, the expectation step is conducted. The EM finds the expected value of the log-likelihood of the complete set concerning the unobserved $\mathbf{Z}$ given $\mathbf{X}$ and a parameter set θ^{i-1}, see Baum et al., 1970. The

notation of the parameter set already indicates its iterative form. The expected value is given by

$$Q(\theta, \theta^{i-1}) = E\left[\log p(\mathbf{X}, \mathbf{Z}|\theta)|\mathbf{X}, \theta^{i-1}\right]. \tag{20}$$

Usually the joint distribution of the compete data set is from the exponential family so that a differentiation is possible. A new set of parameters is derived by maximizing the equation 20.

$$\theta^i = \underset{\theta}{\text{argmax}} \quad Q(\theta, \theta^{i-1}) \tag{21}$$

The two steps are repeated until a local maximum is found or a certain number of iterations is reached. For a detailed explanation of the EM algorithm, see the article by Bilmes et al., 1998 on which this chapter is based.

After the basic idea of the EM algorithm is presented, the special case for the Hidden Markov model, namely the Baum-Welch algorithm is introduced.

Baum-Welch Algorithm

The introduction of the Baum-Welch algorithm will not be done too extensively. The focus will be on the calculation of the parameter update and the composition of the likelihood function $Q(\theta, \theta^{i-1})$.

The Hidden Markov model can be described by a triple of parameters $\lambda = (\pi, A, b)$. These are the parameters for which the likelihood function needs to be maximized. The notation of the previous chapter for a general parameter set θ is replaced by the notation for the specific parameter set of Hidden Markov models λ.

Let us recall the expected value of the log likelihood function in equation (20). The expected value can be calculated with

$$Q(\lambda, \lambda^{i-1}) = \sum_{z \in Z} \left[\log p(\mathbf{X}, \mathbf{z}|\lambda) p(\mathbf{X}, \mathbf{z}|\lambda^{i-1})\right]. \tag{22}$$

The expected value sums over all possible state sequences. The probability $p(\mathbf{X}, \mathbf{z}|\lambda)$ is the joint probability of the observed sequence and a specific state sequence, given the model parameters. This joint probability can be written in terms of the model parameters.

$$p(\mathbf{X}, \mathbf{z}|\lambda) = \pi_{z_0} \prod_{n=1}^{N} a_{z_{n-1}z_n} b_{z_n}(X_n) \tag{23}$$

Equation (23) can be inserted into equation (22) which leads to equation (24)

$$Q(\lambda, \lambda^{i-1}) = \sum_{z \in Z} \pi_{z_0} p(\mathbf{X}, \mathbf{z}|\lambda^{i-1}) + \sum_{z \in Z} \left(\sum_{n=1}^{N} \log a_{z_{n-1}z_n} \right) p(\mathbf{X}, \mathbf{z}|\lambda^{i-1}) + \sum_{z \in Z} \left(\sum_{n=1}^{N} \log b_{z_n}(X_n) \right) p(\mathbf{X}, \mathbf{z}|\lambda^{i-1}) \tag{24}$$

The obtained form of the likelihood function comes with an useful property. The equation can be split into independent terms for each parameter respectively, see Bilmes et al., 1998. The property is made use of in this work to modify the parameter estimate of the transition probabilities. The procedure will be explained in more detail in chapter 4.

From each term, a function of the parameter update can be derived. Before the parameter update function is presented, some additional variables are introduced. As mentioned in chapter 2.4, the Forward-Backward algorithm is applied in the frame of the Baum-Welch algorithm. The forward and backward variables are then used to calculate the parameter updates. Besides the forward and backward variables, two other variables are introduced.

Lets consider the probability of state i at time n given the observations X and the model parameters. This probability is referred to as $\gamma_i(n) = p(Z_n = i|X, \lambda)$ Bilmes et al., 1998. The probability $\gamma_i(n)$ is given by

$$p(Z_n = i|X, \lambda) = \frac{p(Z_n = i, X|\lambda)}{\sum_{j=1}^{m} p(Z_n = j, X|\lambda)} \tag{25}$$

Where $p(Z_n = i, X|\lambda)$ can be defined as the product of the forward and backward variable in state i at time n as following, $p(Z_n = i, X|\lambda) = \alpha_i(n)\beta_i(n)$ (Bilmes et al., 1998). Thus, equation (25) can be written in terms of the forward and backward variable. This leads to

$$\gamma_i(n) := \frac{\alpha_i(n)\beta_i(n)}{\sum_{j=1}^{m} \alpha_j(n)\beta_j(n)} \tag{26}$$

The sum of $\gamma_i(n)$ over time comes with an useful interpretation. The sum expresses the expected number of transitions away from state i, see Bilmes et al., 1998.

The probability of being in state i at time n and in state j at time $n+1$ can be expressed as $\xi_{ij}(n) = p(Z_n = i, Z_{n+1} = j|X, \lambda)$. This expression can be formulated through the transmission probabilities, the emission probabilities and the forward and backward variables, see Bilmes et al., 1998. The expression is presented in equation (27)

$$\xi_{ij}(n) := \frac{\alpha_i(n) a_{ij} b_j(X_{n+1}) \beta_i(n+1)}{\sum_{j=1}^{m} \alpha_i(n) a_{ij} b_j(X_{n+1}) \beta_j(n+1)} \tag{27}$$

The sum of $\xi_{ij}(n)$ over time provides another useful interpretation. The sum expresses the expected number of times the system changes from state i to state j.
Both the sum of $\gamma_i(n)$ and $\xi_{ij}(n)$ over time provide interpretations which are helpful in formulating the parameter updates.

The initial state probability describes the probability of being in state i at time $n = 1$. This probability is introduced as $\gamma_i(1)$, see Bilmes et al., 1998. Consequently, the update of the initial state parameter for each EM step is the following,

$$\tilde{\pi}_i = \gamma_i(1) \tag{28}$$

The transition probability states how likely it is to change from one state into another. For a given state i, the transition probability can be computed by the quotient of the expected number of transition from i to j and the expected number of transitions away from i. The update of the transition probability is given by

$$\tilde{a}_{ij} = \frac{\sum_{n=1}^{N-1} \xi_{ij}(n)}{\sum_{n=1}^{N-1} \gamma_i(n)} \tag{29}$$

The difference between discrete and continuous has not been introduced yet. The difference is important for updating the emission probabilities. Only the update function for the continuous case will be introduced.

The emission probabilities are probability distributions for each time step on X given the state Z, see Bishop, 2006, pp.618. For each time step, the observation is drawn from a Mixture Model, where each $z \in Z$ follows a mixture component. Often, just like in this work, a Gaussian Mixture model is considered. The parameters of a Gaussian distribution are the mean value μ and the covariance Σ. These parameters have to be updated for each mixture component. E.g. Let the observed variable be the temperature per day and let the hidden variable be summer or winter. The distribution of the temperature is different in summer and winter. While estimating the likely state of a given observation, the distribution parameters have to be adapted, too.

The component coefficient c_{il}, the component mean μ_{il} and the component covariance Σ_{il} have to be estimated and updated. The index l represents the l^{th} component of the Mixture model. The probability that the l^{th} component in state i generates the observation X_n is given by

$$\gamma_{il}(n) = \gamma_i(n) \frac{c_{il} b_{il}(X_n)}{b_{il}(X_n)} \tag{30}$$

The coefficient $b_{il}(X_n)$ is the emission probability of generating observation X_n by the l^{th} component in state i (Bilmes et al., 1998). Given equation (30) the update functions for the component coefficient, the component mean and the component covariance are presented here.

$$\tilde{\mu}_{il} = \frac{\sum^N \gamma_{il}(n) X_n}{\sum^N \gamma_{il}(n)} \tag{31}$$

$$\tilde{c}_{il} = \frac{\sum^N \gamma_{il}(n)}{\sum^N \gamma_i(n)} \tag{32}$$

$$\tilde{\Sigma}_{il} = \frac{\sum^N \gamma_{il}(n)(X_n - \mu_{il})(X_n - \mu_{il})^T}{\sum^N \gamma_{il}(n)} \tag{33}$$

The presented Baum-Welch algorithm is implemented in this thesis. In order to obtain a time dependent Hidden Markov model and to incorporate a covariate into the time dependent transmission probability, some extensions to the presented algorithm are conducted. The extensions of the algorithm only apply to the parameter update and likelhood optimization of the transmission probabilities, since the parameters in the likelihood function (19) can be split and optimized individually, see Bilmes et al., 1998 and Carr et al., 2020.

The time dependence of the transmission probability matrix and the incorporation of the covariate are introduced in the chapter 4.

2.6 hmmlearn

hmmlearn is a Python package that provides Hidden Markov models and according to algorithms (Sergey Lebedev, n.d.). The package is built on scikit-learn, NumPy, SciPy, and matplotlib Sergey Lebedev, n.d. The package provides Hidden Markov models with Gaussian emission, Hidden Markov model with Gaussian mixture emissions, and Hidden Markov Model with multinomial (discrete) emissions (Sergey Lebedev, n.d.). The package is based on the work of Rabiner, 1989 and Bilmes et al., 1998, just like the implementation of Hidden Markov models presented in this work. The hmmlearn package has been used in this work and the implementation of a customized Hidden Markov model uses parts of the hmmlearn structure.

2.7 LASSO Regression

Regression with $l1$ regularization was introduced by Tibshirani, 1996 as least absolute shrinkage and selection operator (LASSO) regression. The LASSO logistic regression is performed to classify the schizophrenic and non-schizophrenic group. Logistic regression is considered as a method that provides interpretability of the model fit of single variables. Moreover, $l1$ regularization can be applied to select features.

The $l1$ regularization penalizes the loss function, here the residual sum of squares, by adding the $l1$ norm of the coefficient parameters (Tibshirani, 1996). The loss function is presented in equation (34)

$$\min_{\beta} = ||y - X\beta||_2^2 + \lambda||\beta||_1 \tag{34}$$

Equation (34) shows that the penalty term is added by a scaling parameter λ. The penalty on the size of the parameters β depends on the scale of λ. The scaling parameter λ must be determined.

2.8 Classification Features

A brief overview of the classification features already used in the literature is presented. Qualitative features derived directly from the activity time series are considered.
Table 1 gives an overview of the time-series features used in the literature to analyze activity time series and active/rest patterns.

The mean activity and standard deviation are implemented the most. Recalling chapter 3, a difference between schizophrenic patients and the control group are already observed in the first two moments of the time series distribution. Hence, mean activity and standard deviation are good indicators for classification.

The interdaily stability (IS) and the intraday variability are introduced by Witting et al., 1990.
The interdaily stability (IS) gives an insight into the day-to-day variation. With an increasing day-to-day variation, the linkage between the rest-active rhythm and the Zeitgeber is loosing up (Witting et al., 1990). A strong linkage between the rest-active rhythm and the Zeitgebers would be a clear orientation of its activity to the 24h cycle day/night cycle.

The intraday variability (IV) captures circadian disturbances and a potential split-up of the circadian rhythm (Witting et al., 1990). A high IV could suggest day-time naps e.g.

M10 and L5 are the average activity during the most active 10 and the least active 5 hours a day. M10 can be seen as a good approximation of amplitude of the circadian rhythm, see Witting et al., 1990, while L5 captures activity during sleep.

The root mean square successive differences statistic was introduced by Von Neumann et al., 1941. The mean square successive differences are the standard deviation of the differences time series. Initially, this statistic is meant to unveil a shifting trend in the data. If the mean square successive differences would be significantly different from the standard deviation, there is the effect of a trend see Von Neumann et al., 1941. In the context of activity measuring, it captures the variability in the change of activity.

A common approximation of the circadian rhythm are trigonometric functions (Naitoh et al., 1985), (Bell-Pedersen et al., 2005), (De los Santos et al., 2017). Parameters of the trigonometric function have been used to describe the circadian cycle and applied as features for classification. These features are not considered for the baseline classification. The focus of the work does not lay on fitting trigonometric function and is considered as too complex for the baseline classification.

Autocorrelation is widely used in time series analysis. It quantifies the serial dependence of a time series (Shumway and Stoffer, 2000), (Box et al., 2015). Depending on the degree, the autocorrelation measures the similarity of the values up to the kth observations. The reviewed literature has taken the autocorrelation up to observation 1: $AR(1)$.

This work does not consider the analysis of frequencies by applying Fourier transformation nor calculate the sample entropy. These models are too complex for a baseline classification and it would exceed the extent of this work.

Table 1 summarises the features, which are considered for the baseline classification.

Features for Baseline Classification	Literature Source
Mean activity & Standard Deviation	Carr et al., 2020, Huang et al., 2018, Sano et al., 2012, Lauerma et al., 1994, Huang et al., 2018, Berle et al., 2010, Jakobsen, Ceja, et al., 2020, Scott et al., 2017, Witting et al., 1990, J. Martin et al., 2000, Jakobsen, Garcia-Ceja, et al., 2020
Interdaily Stability(IS) & Intradaily Variability (IV)	Witting et al., 1990, Huang et al., 2018, Gonçalves et al., 2014, Zuurbier et al., 2015
Least active 5 hours (L5) & Most active 10 hours (M10)	Huang et al., 2018, Gonçalves et al., 2014
Root mean square successive differences (RMSSD)	Hauge et al., 2011, Scott et al., 2017
Autocorrelation coefficient	Scott et al., 2017
Not considered Features	
Cosine amplitude & period	J. Martin et al., 2000, J. L. Martin et al., 2005, Naitoh et al., 1985, Bell-Pedersen et al., 2005, De los Santos et al., 2017
Variance of Fourier frequencies	Hauge et al., 2011, Scott et al., 2017
Sample Entropy	Hauge et al., 2011, Scott et al., 2017

Table 1: Overview of quantitative features of actigraphy time series in derived from the literature.

3 Data Analysis

The chapter Data Analysis will present an extensive view of the data structure and the underlying distributions of the given activity time series.
First, a general visual analysis of the data is conducted, followed by a first simple comparison between the patient and the control group.
Second, the distributions of the time series and some transformations are examined. The hypothesis arises, that the observations during periods of rest and activity come from different distributions.
After, the structure of the 24h average activity and the structure of the 12h daily and nightly average activity are analyzed. The visual analysis is conducted to substantiate the hypothesis that observations during periods of rest and activity are generated by different distributions.
Finally, the stationarity and the autocorrelation of the time series are examined. Especially the autocorrelation is supposed to give an insight into the seasonality of the data.

3.1 Data Structure

Figure 3 shows example patients' data on the left side and data of the control group on the right side. The data was recorded from 9 to 20 days per person. The data reveals the circadian rhythm of a person. The periods in which a person is active appears in blocks. The periods between the blocks are seen as resting periods. The resting periods seem to have a lower mean and lower variance and appear to be shorter in time than the active periods. These observations match with known human behavior since resting periods are in general shorter than active ones.
It is known that a person's rest/active pattern is affected by the day and night rhythm. This is related to social norms and the daylight patterns (of Health and Service, n.d.). Thus, the time series observed are likely to follow some sort of 24h seasonality.

The transition between the active and resting periods appear quite abrupt. As soon as a person lays down to rest or sleep, it reduces its movements. This leads to the observed block formation within the time series. Therefore abrupt changes of mean and variances over time can be observed.
The abrupt changes in time, and the different mean and variance, as the first two moments of the normal distribution, for active and resting periods let the hypothesis arise that the observations of the active and resting period

come from different distributions.

It is of interest to investigate if the observations of the active and resting periods come from different distributions. Unfortunately, it is not known at what time a person is resting or active. The state of active/resting is hidden. A sufficient model of the active and resting period might give us valuable information about the difference between the circadian cycle of a healthy person and a schizophrenic person.

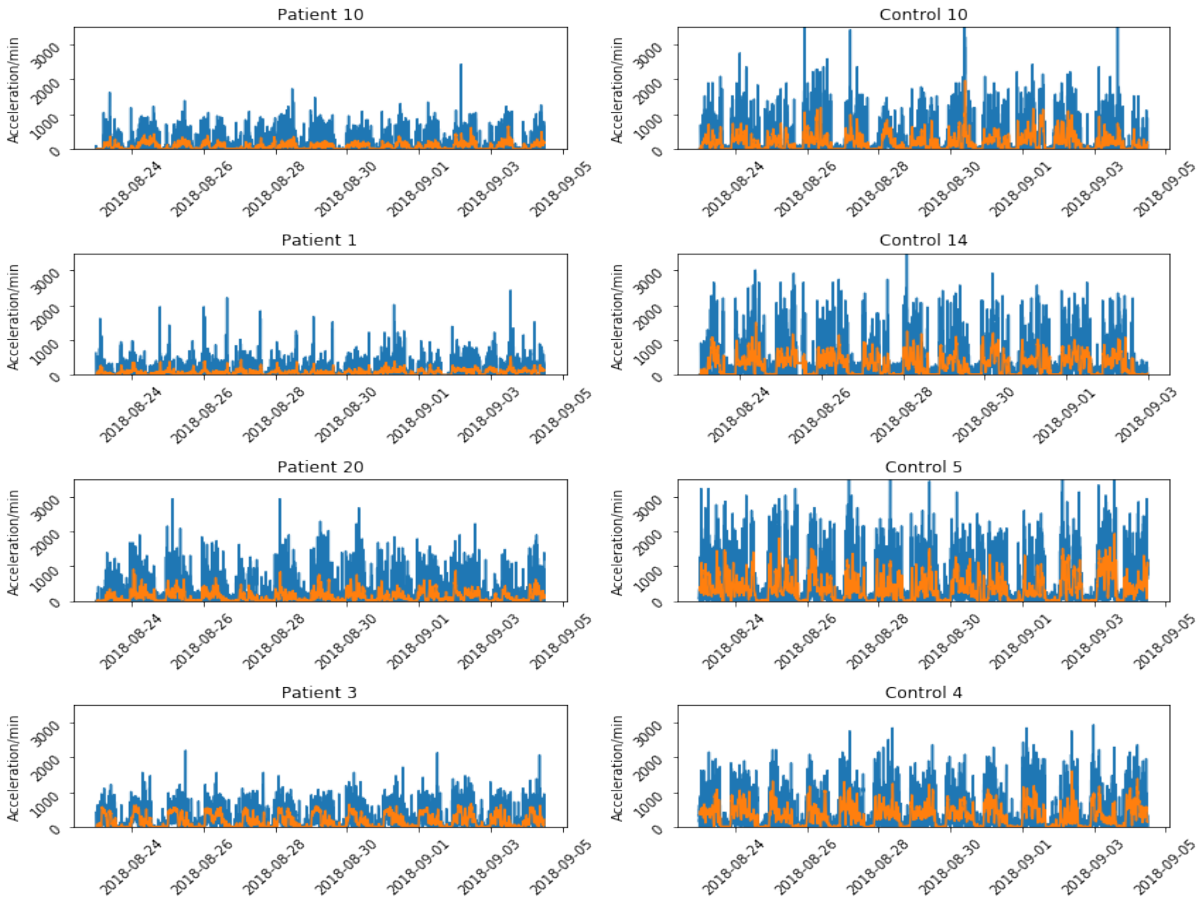


Figure 3: Sample time series of different patients and control group plotted next to each other. The blue data presents the original time series, the orange is a averaged time series by window size 30. Source: own illustration

Figure 3 shows us some differences between schizophrenic patients and the control group already. There seems to be a difference in the scale of time series between the two groups. The control group appears overall more active than the patients. Moreover, the overall variance of activity occurs to be higher for the control group than for the patients. This will be further investigated when comparing the moments of the time series distribution.

It can be observed, that the control groups seem to stick more to a rest/active pattern than the patients. The patient's time series don't show as strong changes between the two different states of activity as the control group does. The control group activity gives the impression of following more a periodic pattern.
The observations underline the diagnosed characteristics of patients with a schizophrenic condition. The first visual analysis gives the idea, that patients show a disturbed circadian cycle and behavioral abnormalities.

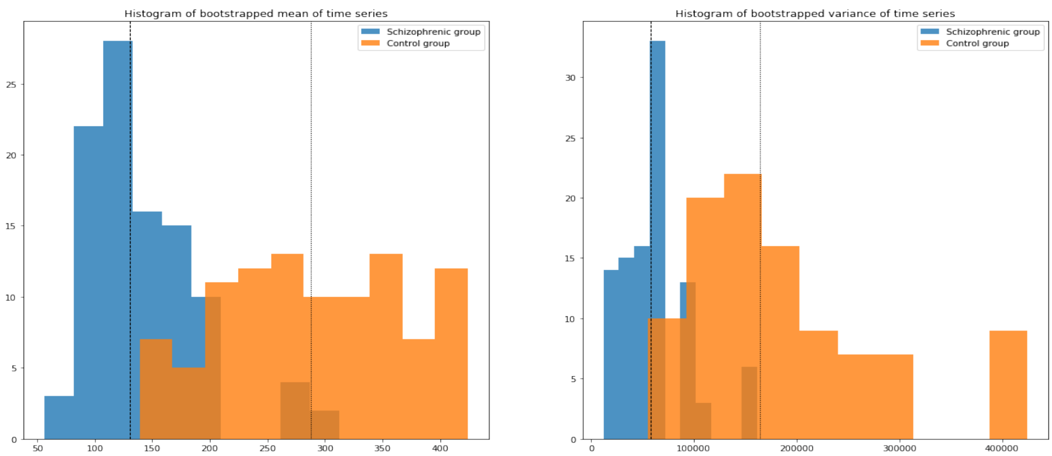


Figure 4: Histograms of the sample mean to the left and the sample variance to the right, for Schizophrenic in blue and control group in orange. The dashed lines show the median of both groups. The histograms indicate that the two subgroups probably belong to different distributions and thus a classification is possible. Source: own illustration

Figure 4 shows two histograms of the overall sample mean and the sample variance of each time series. To improve the parameter estimate, bootstrapping was applied to increase the sample size from 22 and 32 from the schizophrenic and control group to 50, respectively.

The histograms show that the control group has indeed a higher overall mean and variance of activity than the schizophrenic group. A t-test is conducted to test whether there is a difference in scale between patients and the control group. A t-test examines if the means of two groups are equal. With a p-Value: $p = 2.7e^{-08}$ close to zero, the null hypothesis, that the two groups come from the same distribution, can be rejected. This proofs, that there is a difference between schizophrenic patients and the control group. This gives rise to a classification.

3.2 Stationarity and Autocorrelation

Given the nature of the data, seasonality can be assumed. The analyzed circadian rhythm is assumed to be the reason for the seasonality. Figure 33 presents four example autocorrelation charts with lag 2000. There is a certain periodicity within the autocorrelation lags. It can be seen in figure 33 that the negative correlation to $x(n)$ is highest around $x(n - 750)$. 720 minutes are 12 hours. The positive autocorrelation finds a peak close to $x(t - 1440)$, which are 24 hours.
The autocorrelation coefficients for different lags might give us insight into the difference between schizophrenic patients and the control group. The autocorrelation coefficient will be used as features for the latter baseline classification.

The data does not appear to have any trend. The Augmented-Dickey-Fuller (ADF) as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for trend-stationarity are conducted. Both tests show that there is no unit root in any of the time series. Thus, each time series is not trend stationery.

3.3 Analysis on the Time Series Sample Distribution

In this chapter, the time series distributions of the patient and control group are analyzed. To do so, the histograms of the time series and its transformations are analyzed.

Figure 6 shows 12 different histograms of the full-time series of example patients and control subjects. The patients' histograms are plotted on the left side, while the control groups' histograms are plotted on the right. The histograms are provided for 12 example time series.

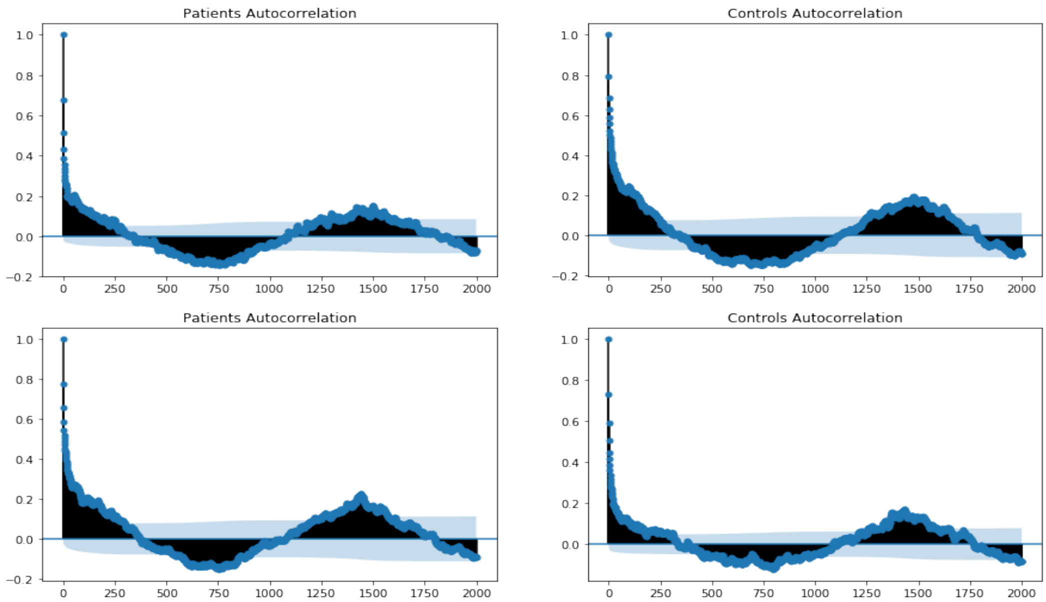


Figure 5: Autocorrelation chart with lag 2000 for the patient group on left the control group on the right. Source: own illustration

All histograms center at zero and dropping off in the next bin. From there on, a tailing-off can be observed. The accumulation of zeros might be due to the data collected through the sensors. By nature, the data is non-negative, since we can not observe the negative activity. Due to the large number of zero counts within the data, it is difficult to extract a given sample distribution out of these histograms.

To overcome the restriction of non-negativity, one can differentiate the time series by one. Differencing the time series by one will give us a slightly different interpretation of it. The differenced time series is no longer the measured activity per minute, but the change in activity compared to the last minute. The differencing is done by, $y'_t = y_t - y_{t-1}$. Positive values can be interpreted as positive change inactivity. The person has been less active before compared to now. On the other hand, negative values indicate that the persons are less active now than before.
An example histogram of the differenced time series can be found in the appendix.

The distribution of the differenced time series still contains a large number of zeros, but the distribution tails-off in a positive and negative direction.

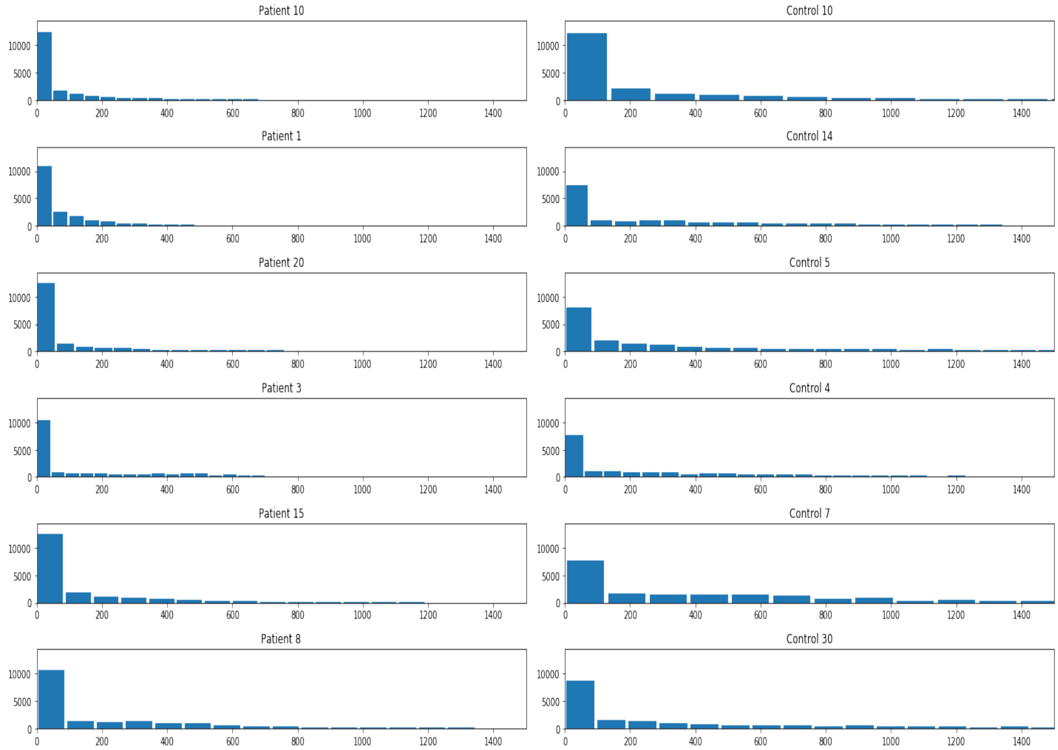


Figure 6: Histograms of example time series from the patients and control group to investigate on the time series sample distribution. Patients time series are presented on the left, the control group to the right. Source: own illustration

After conducting the Kolmogorov-Smirnoff test for normality as well as the Shapiro-Wilk test for normality, it can be concluded that the data is not normally distributed.

Another common transformation to suggest is the log-transformation of the time series. Due to the zeros within the data, the data is shifted by one to be able to calculate the natural logarithm.
The histograms of the log-transformed time series are presented in Figure 7. The log-transformed histograms reveal a high number of zeros, too. Besides a large number of zeros, the distribution of the log-transformation reveals a second smaller accumulation of observations located on the right-hand side of the histogram.

As already indicated at an earlier stage, it is suspected, that observations

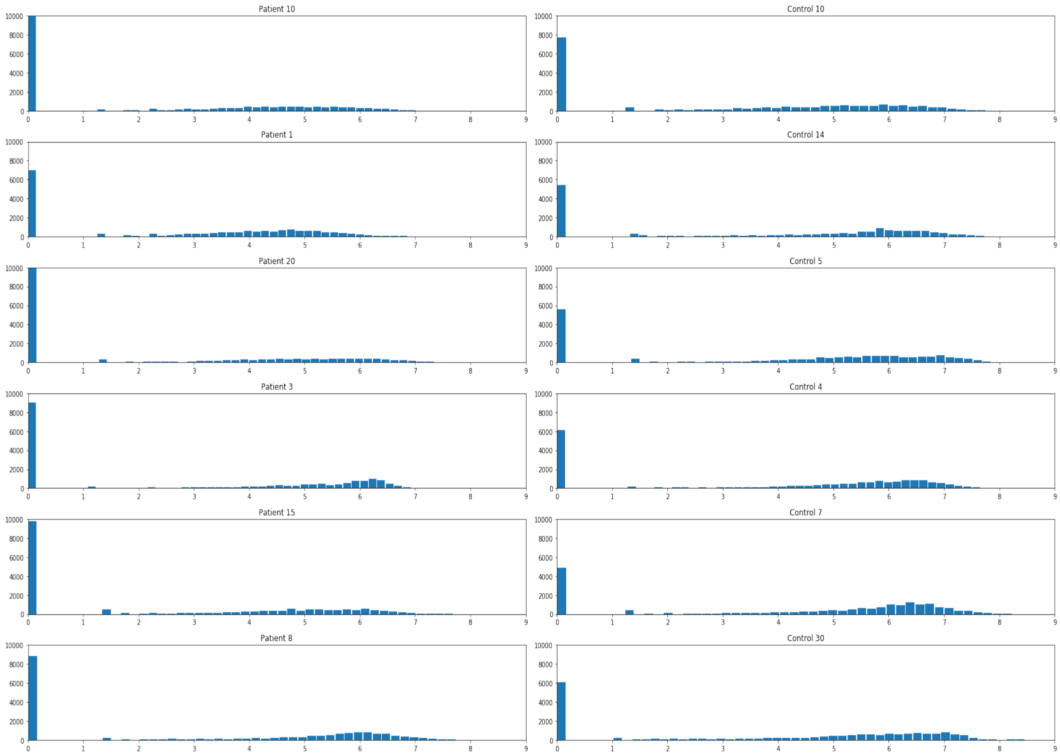


Figure 7: Histograms of the log-transformed time series of patients and the control group. Source: own illustration

of active and resting periods come from different distributions. This presumption is supported by the histogram in figure 7. The resting and active periods are not directly observed. To get a better insight, the daily structure, as well as the day and night pattern of the time series and its distribution, are analyzed in the following.

3.4 Daily 24h structure of the Time Series

This chapter will investigate a person's average activity for 24 hours.
The circadian rhythm, the inner-clock regulates the biological process within a 24-hour cycle and is related to light perception (**kolmos2007circadian**). It is, therefore, reasonable to examine a person's daily activity.

Figure 8 illustrates the average activity over 24h, from 9 am to 8.59 am the next day, of patients and the control group. The plot shows the average

activity during the day.
Especially the control group reveals a course of average daily activity. There is an increase at the beginning of the day, around 5 pm. After, there is a decrease until the course reaches a low around 3 am.

Since figure 8 presents the average activity, there is no abrupt change between day and night observable. But figure 8 shows a different pattern between day and night.

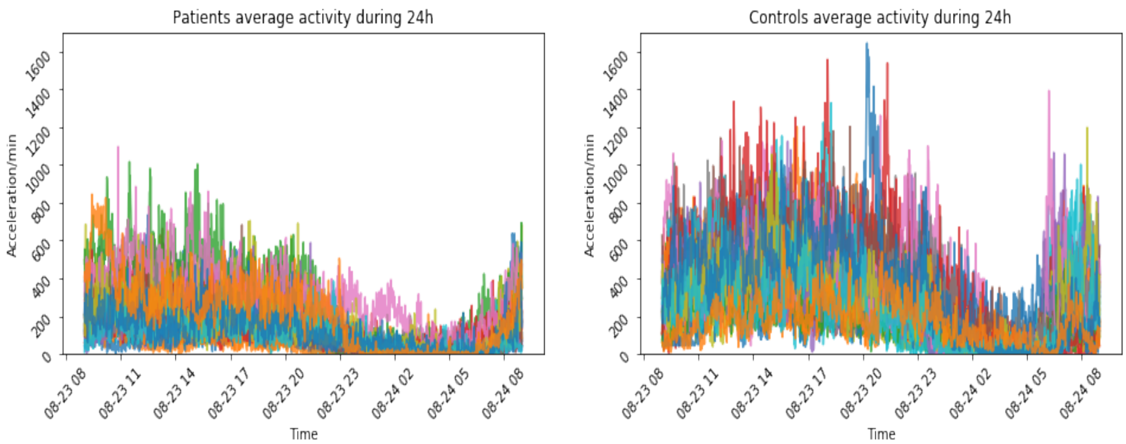


Figure 8: The average activity over 24h, starting from 9 am to 8.59 am. The left side represent the patients average activity over 24h, the right side for the control group. The full data set is displayed. Source: own illustration

Within the group of patients, we can see a similar course of activity but the difference between day and night is not as remarkable as for the control group. The average activity of patients during the day is lower than that of the control group. Thus, the distinction between day and night is smaller, too. The literature refers to the distinction between night and day as intraday variability and plays a significant role in the classification of sleep disturbances (Meyer et al., 2020).

To get a better overview of the average 24 hours activity of the two groups, box plots for each individual are presented in figure 9. Figure 9 shows, that the median of the daily average activity of the patient group is smaller. Moreover is the range of values smaller than for the control group. According to the whiskers of the box plots, the control group appears to have higher maximums than the control group.

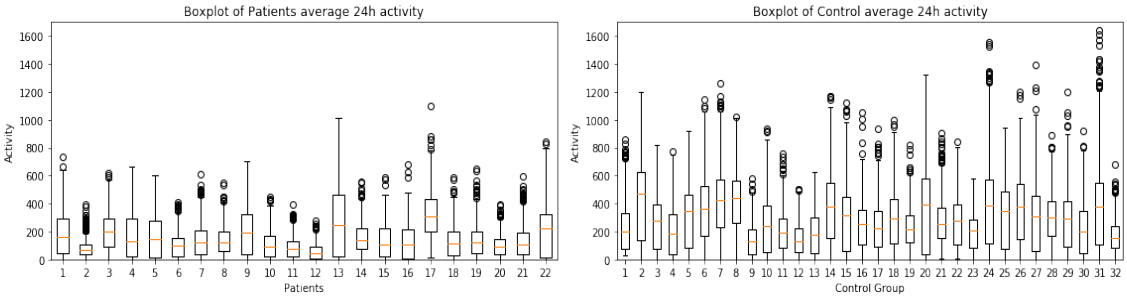


Figure 9: Box plot of the average daily activity of each person. The control group is shown on the right side, while the patient group is shown on the left side. Source: own illustration

In conclusion, there is already a visual distinction between the two groups on which the classification will be based on. This work will investigate if there are more distinct features within the data than the overall mean and variance of the time series and the quantiles of the average 24h activity.
It is still a matter of investigating whether day and night are subject to different distributions. To do so, the next chapter will focus on the 12h activity during day and night.

3.5 Day and Night Structure of the Time Series

It is assumed, that the active and resting period of an individual comes from two different distributions. To substantiate this, the activity of the two groups during the day and the night is examined. While the day is defined from 9 am until 9 pm, the night period is defined from 9 pm until 9 am.
Figure 10 visualizes the activity of the patient and the control group for two defined periods, day and night. As already shown in other illustrations, the average of the patients' data seems to be lower. However, fig 10 shows the difference in the course of the activity.
The controls averaged 12h daily activity follows are more curved course than the patients' average 12h daily activity. The difference between the groups is even more emphasized during the night period, illustrated in the two bottom graphs of figure 10. The patients' activity is lower and less variable. The edges of the graphs show the times of going to sleep and getting up. These transitions of the resting and active state are more visible for the control

group.

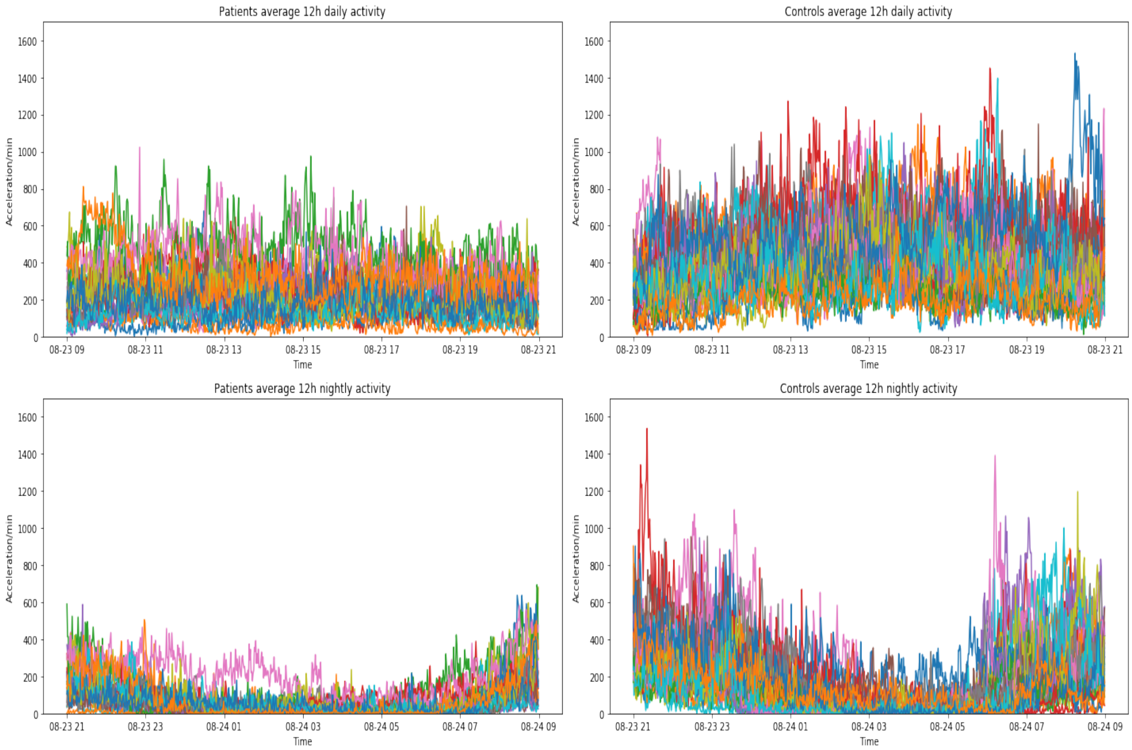


Figure 10: Visualization of the average nightly and daily 12h activity. The visualization are presented for the patient group on the left side and control group on the right side. The day was defined as a 12h interval from 9 am until 9 pm. The night was defined as a 12h from 9pm until 9am. Source: own illustration

Figure 11 shows histograms of the day and night log-transformed average activity of example individuals of the patient and control group. The histograms of day and night are compared within a plot. We can observe, that the daily histogram appears to be slightly left-skewed, while the night histograms reveal strong left skewness. According to figure 11, the day and night structure have different moments of their distribution. This observation underlines the hypothesis, that the daily and nightly activity belong to different distributions.

Thus, figure 12 visualizes how the first two moments of the daily and nightly average activity of the control and patient group are spatially distributed on

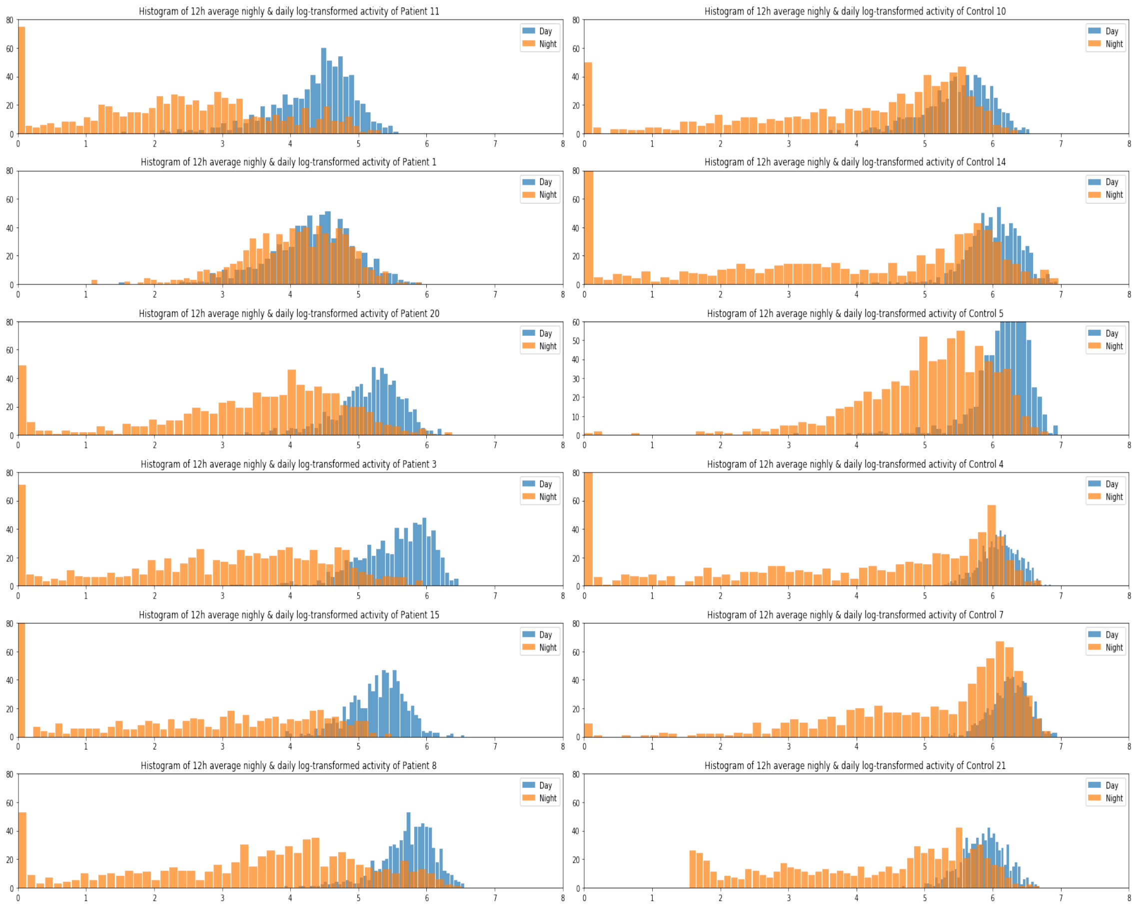


Figure 11: Comparing day and night activity distributions of example individuals from the control and patient group. Source: own illustration

a plane. Especially, more of the control group's average daily activity has a higher mean value of 300 accelerations/minute, while most of the control groups nightly mean range between 100 and 300 accelerations/minute.

Notably, the standard deviation of the nightly period tends to be higher than the one of the daily period. The reason for the higher standard deviation might be the included transition between the two states, rest and active. The period lasts from 9 pm until 9 am, but most people go to bed later and get up earlier. The resting and being active periods are not separated in 12h periods. The two bottom visualizations of figure 10 show the included transitions between the resting and being active.

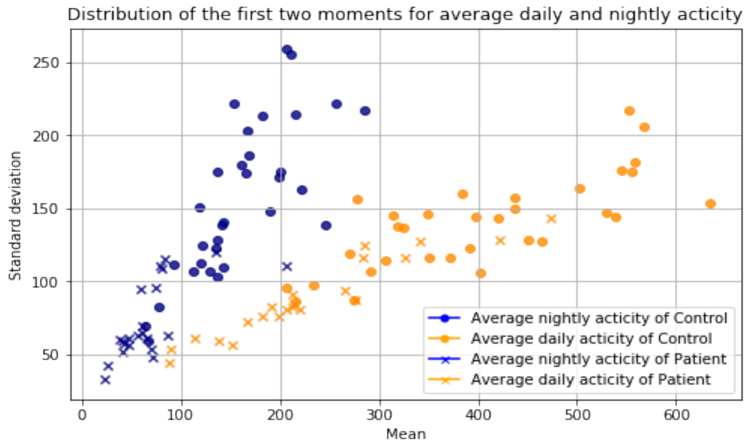


Figure 12: Scatter diagram of the first two moments of the average 12h daily and nightly activity for the patient and control group. The blue-colored marks represent the average nightly activity. The orange-colored markers show the daily activity. Control and patient groups are visually separated by dots and crosses, respectively. The different observations between daily and nightly activity seem to be linearly separable. Source: own illustration

In the two euclidean dimensional hyperplanes of standard deviation on the one axis and the mean value on the other axis, the different observations between day and night are linearly separable. Therefore, it will be assumed that the nightly and daily average activity of the control and patient group is generated by different distributions.

There is the hypothesis, that the active and rest phases of a person are generated by different distributions. These two phases cannot be observed but according to a normal life cycle, a person is more likely resting during the night than during the day. Out of this reason, the daily and nightly activity was analyzed, to see differences between the two periods.

According to figure 12, there is a difference in the first two moments of the day and night structures for both the control and patient group. This strengthens the assumption that the resting and being active periods are generated by different distributions. As a matter of course, this has to be further investigated.
But with his strong hypothesis, the Hidden Markov model will be applied to the data to model the circadian cycle as the two different activity periods.

4 Method

The classification of schizophrenic and non-schizophrenic persons is assumed to be improved by using the parameters of an appropriate model. In chapter 3, the idea is derived that the circadian rhythm can be described by two states - the rest and the active state. Even though the states cannot be observed, the activity can be observed.
As stated in chapter 1.3, Hidden Markov models have already been applied to model unobserved rest/active states through the observed activity.
According to the hypotheses in chapter 1.2, three different variations of the Hidden Markov model are applied. Two extensions of the Hidden Markov model are implemented in this thesis. All three variations differ in their transition probability matrix.

The first Hidden Markov Model corresponds to the general model and has a constant or time-independent transition probability. The second proposed Hidden Markov model has a time-dependent transition probability matrix. According to hypothesis 3, the time-dependent structure of the transition probability is ought to give valuable information for classification. The last proposed Hidden Markov model incorporates a trigonometric function into the time-dependent structure of the transition probability. It is assumed that the transition probabilities are related to the circadian rhythm. The probability of resting increases at night, while it decreases during the day. A trigonometric function that follows a 24-hour course seems to be a suitable way to model the circadian rhythm of transition probabilities. According to hypothesis 4, the disruption of the circadian rhythm can be captured in the link coefficients with which the trigonometric function is integrated with the transition probabilities.

The two applied extensions of the Hidden Markov model are not offered by any Python library. Therefore, the models and the according algorithms were implemented within the scope of this thesis.

The method chapter firstly introduces the time-independent Hidden Markov model. The approach is outlined. Secondly, the approach of the time-dependent Hidden Markov models is presented. The extension of the Baum-Welch algorithm is described mathematically. After the model of the co-variate is presented. This is followed by a presentation of the actual implementation of the extended Hidden Markov model in Python. Finally, the limitations of the implementation used in this thesis are outlined.

4.1 The time-independent Hidden Markov Model

The approach and model adaptation of the time-independent Hidden Markov model is illustrated. In chapter 3, the idea is developed that the rest/active states are described by two different distributions. Usually, the choice of the number of hidden states can be difficult. However, the number of hidden states is already determined by the assumptions of only two states of activity. Since the rest/active states are ought to be modeled, the Hidden Markov model includes two hidden variables.
The observed activity data is a continuous-time series. As introduced in the theory chapter 2.5, a Gaussian Mixture Hidden Markov with diagonal covariance matrix is applied.

In chapter 3, the distribution of the activity time series is discussed. As observed in figure 11, the day and night activity reveals skewed normal distributions. Therefore, it is assumed that the actual unobserved rest and active states can be represented by a Gaussian process and described by their mean and variance.

The diagonal covariance matrix represents only the variance of a single state, respectively. No covariance between different states is given. Computational complexity is an issue in this work as it will be stated in the chapter 4.2.4. Thus, to reduce the number of parameters, only the variance of each state will be estimated.

The python library *hmmlearn* is used for the implementation. The transition probabilities and the two moments describing the state distribution are extracted for the classification.

4.2 The time dependent Hidden Markov Model

To extend the potential classification features, a time-dependent Hidden Markov model is presented. The time dependence refers to the transition probability matrix. Each element of the matrix is observed over each time step n. According to hypothesis 3, it is assumed that the time dependence can bring valuable additional information to improve the classification.

Compared to the earlier model the only proposed modification is indeed the time dependence of the transition matrix. Hence, a Gaussian Mixture Model with a diagonal covariance matrix is applied. In the following, two different models are proposed.

First, a Gaussian Mixture Model with a diagonal covariance matrix is implemented, whose transition probabilities are observed over time.

Secondly, the same Gaussian Mixture model is extended by modeling the time-dependent transition probabilities as a function of time. This function of time incorporates a covariate. The two models are referred to as

- Time-dependent Hidden Markov model
- Time-dependent Hidden Markov model with incorporated covariate

The Sergey Lebedev, n.d. package does not provide the two modifications. For this purpose, the Hidden Markov model, including the forward-backward algorithm, Viterbi and Baum-Welch algorithm with the corresponding modifications are implemented within the scale of this thesis. The extension of the Baum-Welch algorithm and the implementation of the whole model is described in the following chapters.

4.2.1 Extension of the Baum-Welch Algorithm

In the theory chapter 2.5 the Baum-Welch algorithm is introduced. Equation (19) presents the likelihood function which is ought to be maximized. The equation is a sum of the parameter terms. Each term of the function can be maximized independently. This allows one to focus solely on the transition probabilities.

The time dependent transition probability matrix is denoted as

$$A(n) = \begin{bmatrix} a_{00}(n) & a_{01}(n) \\ a_{10}(n) & a_{11}(n) \end{bmatrix} \tag{35}$$

Each element of the matrix states the probability to transit from state i to state j at time n.

The time-dependent Hidden Markov model analyses these transition probabilities. To capture the variability of the transition probabilities over time, the variance of each element overtime is calculated. The variance is used as an additional feature for the classification.

The time-dependent Hidden Markov model with incorporated covariate includes one further model extension. The idea of incorporating a covariate follows the approach of, Banachewicz et al., 2008, see. The approach has been applied to activity time series in Carr et al., 2020 and Huang et al., 2018.

The covariate is incorporated into the transition probabilities through a link function. There is various link function to choose from Banachewicz et al., 2008. In this work, a binomial logistic link function is applied. The link function is binomial as the outcome of the hidden variable is binary - rest or active. The binomial logistic link function is described as

$$a_{ij}(n) = p(Z_n = i|Z_{n-1} = j, C_n) = \frac{\exp(\phi_{ij}^0 + \phi_{ij}^1 C_n)}{\sum_{j=1}^{m} \exp(\phi_{ij}^0 + \phi_{ij}^1 C_n)} \tag{36}$$

$a_{ij}(n)$ denotes the time dependent transition probability. C_n is a vector of covariates that is supposed to be linked to the transition probabilities. The coefficients ϕ_{ij}^0 and ϕ_{ij}^1 are vectors of link coefficients with which the covariate will be linked to the transition probabilities. How the covariate is modelled and how the link coefficients can be interpreted in the case of trigonometric function will be discussed in detail in chapter 4.2.2.

The vectors of the link coefficients are numerically optimized in an additional step of the Baum-Welch algorithm (Huang et al., 2018). Since the number of link coefficients increases with the number of covariates integrated, only one covariate is implemented.

The term of the log likelihood function in (19), which is to be optimized is the following Carr et al., 2020,

$$\sum_{n=1}^{N-1} \sum_{i=0}^{m} \sum_{j=0}^{m} \xi_{ij}(n) \log p(Z_{n+1} = i|Z_n = j, C_n, A_{ij}(n)) \tag{37}$$

Equation (37) is optimized numerically for the link coefficients within each step of the Baum-Welch algorithm.

4.2.2 Modelling the Covariate

Trigonometric functions have already been used to model the circadian rhythm, see Phillips et al., 2010, Phillips et al., 2011 and Skeldon et al., 2014. It is assumed that the activity of a healthy subject increases during the day and decreases to its minimum during the night, before rising again. Translated to the case of transition probabilities, the likelihood to be active is higher during the day than at night and vice versa. This can be described by a sine and cosine function.
As already shown in chapter 3, the 24-hour structure of the average activity indeed could be approximated by a trigonometric function.

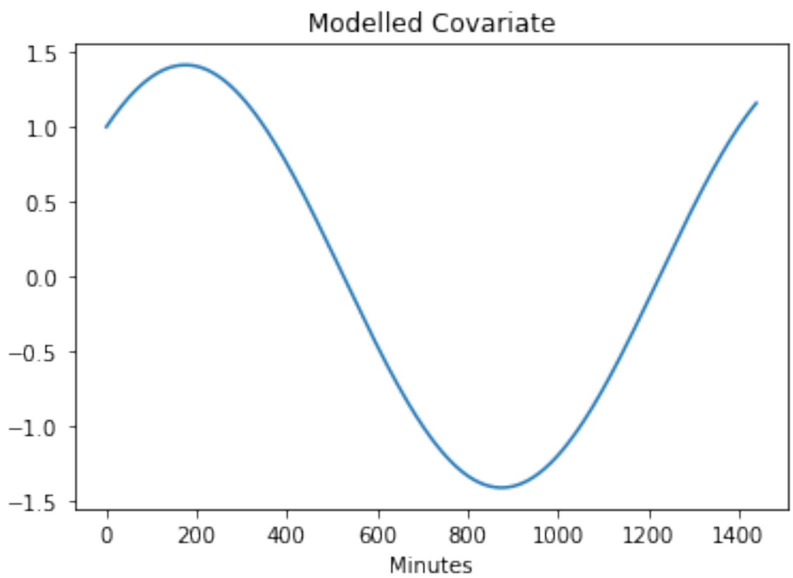


Figure 13: The integrated covariate, modelled as a trigonometric function. The function is shown for 24h, starting from 9am to 8:59am. Source: own illustration

The trigonometric function is assumed to approximate the circadian rhythm of a healthy person. Consequently, the disruption of the circadian rhythm of a schizophrenic person should be embedded in the link coefficients. Thus, the link coefficients are supposed to quantify this disruption. The hypothesis is that the link coefficients are larger the more the actual course of the transition probabilities deviates from the assumed healthy course. In conclusion, schizophrenic subjects would have greater link coefficients.

Figure 13 shows the covariate for a 24 hours interval. The covariate is ought to approximate the daily average activity of a healthy subject.
Equation 38 shows the trigonometric function which is integrated as the covariate. It is the sum of a sine and a cosine function. The period of the function is fitted to one day, which is 1440 minutes long.

$$C_n = \phi^0 + \phi^1 \left[\sin\left(\frac{M}{60 \cdot 24} 2\pi x_n\right) + \cos\left(\frac{M}{60 \cdot 24} 2\pi x_n\right) \right] \tag{38}$$

The equation contains the linkage coefficients ϕ to show their possible interpretation.
ϕ^0 is the offset of the function, while ϕ^1 is the amplitude. This representation is useful for the later interpretation of the results of the optimized link coefficients.

4.2.3 Implementation of the time dependent Hidden Markov Model

In this chapter, some technicalities of the implementation of the Hidden Markov model are described.
The implementation can be accessed through the following Github repository: Classification of Schizophrenia - Project

The Baum-Welch algorithm is implemented according to Bilmes et al., 1998. As already stated in chapter 2, in the scale of the Baum-Welch algorithm, the Forward-Backward algorithm is implemented. Moreover, to solve the problem of decoding, the Viterbi algorithm is implemented, too. However, the core of this work and its implementation is the Baum-Welch algorithm. The Baum-Welch algorithm is implemented as a function of the states or hidden variables, a model object, the specific time series, the number of maximal iterations, and a threshold for the likelihood convergence. Additionally, if the covariate should be incorporated, the model of covariate itself and some boundaries for the external numerical optimization of the loglikelihood function (19) must be given as parameters.

The model object contains all the parameters which are then updated throughout the algorithm. The updated parameters can be accessed after the conduction of the algorithm as attributes of the object. The implemented algorithm can be applied for more than only two states.
Some of the structures of the *hmmlearn* Sergey Lebedev, n.d. package are used. In order to compute the emission probabilities, a subpackage of the *hmmlearn* package called *hmmc* is applied Sergey Lebedev, n.d.
The object function of equation (37) is numerically optimized by a Sequential Least Squares Programming optimizer provided by the Virtanen et al., 2020.

For the time-dependent Hidden Markov model, the maximum number of iterations is 200. For the time-dependent Hidden Markov model with an integrated covariate, the maximum number of iterations is ten. The reason for the small number of iterations in the time-dependent Hidden Markov model with integrated covariate is the computing time of the Baum-Welch with the additional optimization step. The additional optimization step refers to the maximization of the objective function in (37), which is performed by the Sequential Least-Squares Programming optimizer by Virtanen et al., 2020. The limitation will be discussed in more detail in chapter 4.2.4.

The output of the implemented Baum-Welch is the calculated transition probabilities over time, the log-likelihood of the model, and the model parameters. Moreover, the optimized link coefficients are saved.
The results are automatically saved in CSV files. Through a loading function, the results can be load into a result object. The implemented Viterbi algorithm can be applied to a result object to decode the hidden states for the corresponding time series.

4.2.4 Limitations

The proposed implementation of the Baum-Welch algorithm faces some limitations. The computation time is slow. Especially the integration of the covariate costs much time. Due to this limitation, only ten optimization steps for the time-dependent Hidden Markov model with integrated covariate were conducted. This affects the model fit of the proposed model. The time-dependent Hidden Markov model with integrated covariate is not considered as appropriately fitted.
However, for the overall comparison based on the classification performance,

the evaluation and the poor model fit is not considered as problematic. But the implementation can be improved in the future to achieve better model fits of the time-dependent Hidden Markov model with an integrated covariate.
The computation was performed on a Mac Book Pro (mid-2012) with a 2.5 GHz Intel Core i5 processor and a 4 GB 1600 MHz DDR3 RAM.

5 Model Evaluation

In the subsequent chapter, the time-dependent and time-independent Hidden Markov models are evaluated based on their fit to the given data.
The model fit of the Hidden Markov model provided by the *hmmlearn* and the model fit of the Hidden Markov model implemented in the scale of this work will be compared. Due to the small number of iterations used, the time-dependent Hidden Markov model with integrated covariate is not considered in the comparison.
For the comparison, Akaike's information criterion (AIC), introduced in Akaike, 1974, is calculated according to

$$AIC = -2\log(likelihood) + 2K \tag{39}$$

K is the number of model parameters, Akaike, 1974, see. The main purpose of this chapter is to show the validity of the implemented Hidden Markov model.

5.1 Convergence Rates of the time-dependent Hidden Markov model

The Baum-Welch algorithm of the time-dependent Hidden Markov model is performed on 200 iterations. The convergence of the Baum-Welch regarding the maximized likelihood is analyzed. The fact, that the implemented Baum-Welch algorithm converges is key for its success- and meaningful application.

According to Bilmes et al., 1998, the Expectation-Maximization algorithm is guaranteed to find a local maximum as soon as the likelihood converges. Convergence is given, when a parameter set Θ^i is found for which the following equations counts.

$$|Q(\Theta^{i+1}, \Theta^i) - Q(\Theta^i, \Theta^{i-1})| < \tau \tag{40}$$

According to equation (40), convergence is obtained as soon as the likelihood of the updated parameter set is smaller than some threshold τ.

Figure 14 presents the course of the log-likelihood for the control subjects and the schizophrenic subjects. The given log-likelihood is interpreted in a way, that the higher the log-likelihood, the better does the model fit.

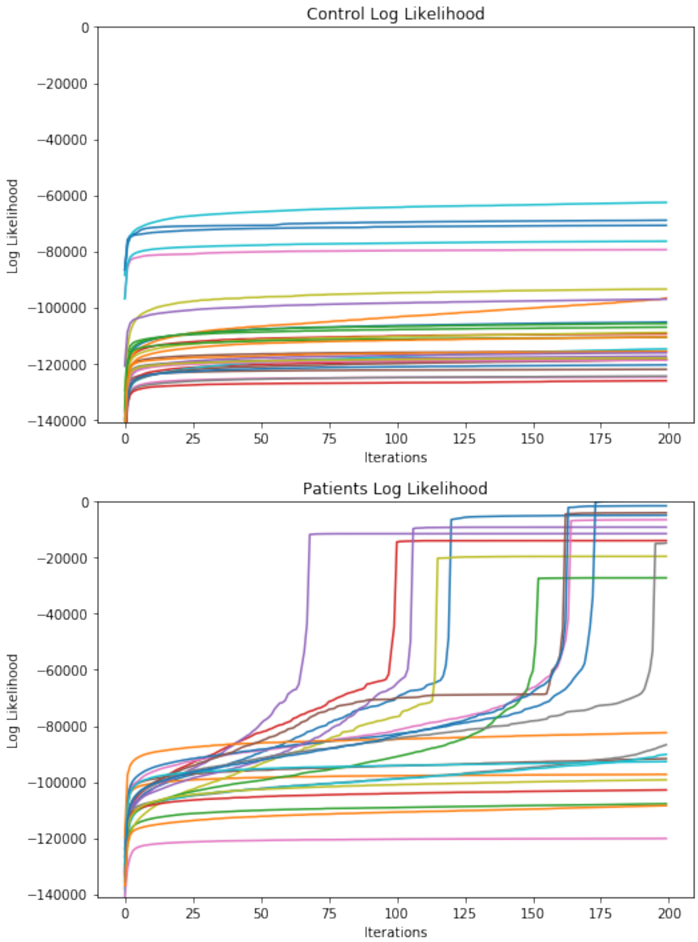


Figure 14: The convergence rates of the log likelihood of the implemented Baum-Welch algorithm. The rates are presented for both the control subjects and the Schizophrenic subjects. Source: own illustration

Two different shapes of convergence can be observed. The convergence rates of the control group and some of the convergence rates of schizophrenic patients assimilating within the first few iterations to a constant level. After reaching the constant level, it is only marginally improved.

However, some of the convergence rates within the group of schizophrenic patients show a different course. The likelihood gradually increases to a certain point. After the gradual increase, the likelihood raises nearly vertical to a new level.
The concrete reasons for this behavior would need further investigation. A

potential reason is a fact, that the EM-Algorithm only guarantees to converge to a local maximum. The observed straight raise could be the finding of another local maximum in the parameter search space.

However, the implemented algorithm shows convergence in the log-likelihood for all analyzed time series. Hence, the implemented Baum-Welch algorithm can be seen as valid for modeling the given data.

5.2 Comparison of the model fit

The model fit of the time-independent and the time-dependent Hidden Markov model is evaluated based on Akaike's Information Criterion.

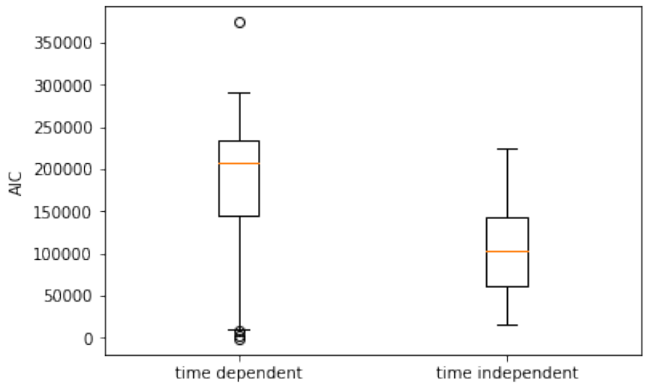


Figure 15: Comparison of the model fit between the time dependent and time independent Hidden Markov model. Boxplots of the AIC of the time independent and time dependent Hidden Markov model. Source: own illustration

Figure 15 shows a boxplot graphic of the AIC for each fitted activity time series. Figure 15 compares the time-independent and the time-dependent Hidden Markov model based on their model fit.
The figure shows that the time-dependent Hidden Markov model obtains a higher AIC in the median. This means, that the model fit of the time-dependent Hidden Markov model is poorer. Moreover, the variations of the AICs obtained by the time-dependent Hidden Markov model is bigger than for the time-dependent model.

The time-independent model provided by the (Sergey Lebedev, n.d.) package achieves a better model adaptation. However, the conclusion must not be, that the implemented time-dependent Hidden Markov Model is not valid. The boxplots in figure 15 and the according AICs show that the model fit of the implemented time-dependent Hidden Markov model is not too far off the ones provided by the package *hmmlearn*.

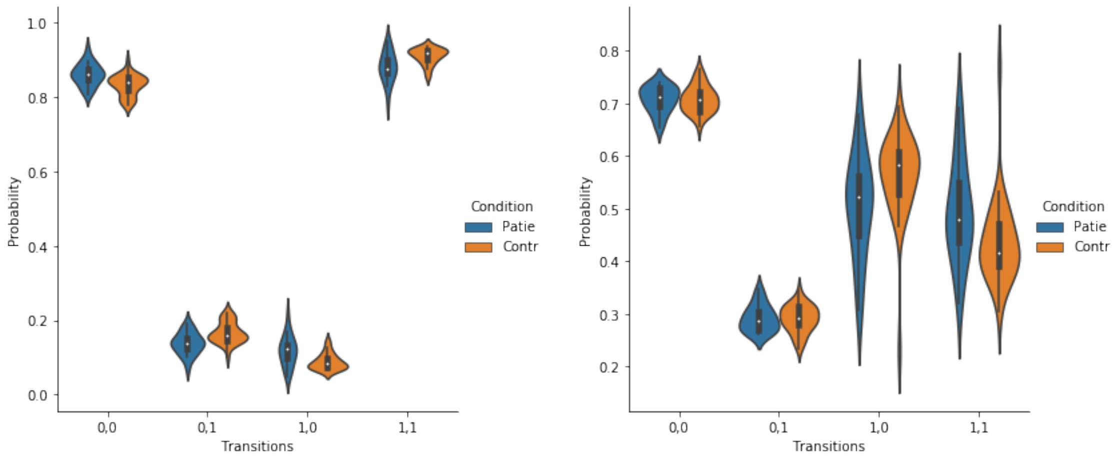


Figure 16: Visualization of the fitted transition probability matrix for patient and control group. Violin plots are shown for each transition probability matrix element. The left side is the fitted transition probability matrix of the time-independent Hidden Markov model. The right side shows the violin plot of the time-dependent Hidden Markov model. Source: own illustration

Figure 16 shows a so-called violin plot. A violin plot is similar to a box plot but shows the full density instead of the summary statistics.
Figure 16 visualizes the estimated transition probabilities for each transition matrix element. Additionally, it compares the control group and the schizophrenic patients for each transition probability element estimate.

Again it can be shown, that the model fit of the time-independent Hidden Markov model appears to be better. The matrix element estimates are overall consistently denser than the ones estimated by the time dependent Hidden Markov model. While the transition elements a_{00} and a_{01} are similar for both models, the estimates of a_{10} and a_{11} are majorly different. The estimates of the time dependent Hidden Markov model for a_{10} and a_{11} exhibit a large

variability. Furthermore, the median of the a_{11} estimate is lower than the a_{10} estimate.
For the time independent Hidden Markov model, the estimates of a_{00} and a_{11} clearly settled around $80\% - 90\%$, whereas the parameter estimate of a_{11} in the time dependent model is located around $40\% - 50\%$.

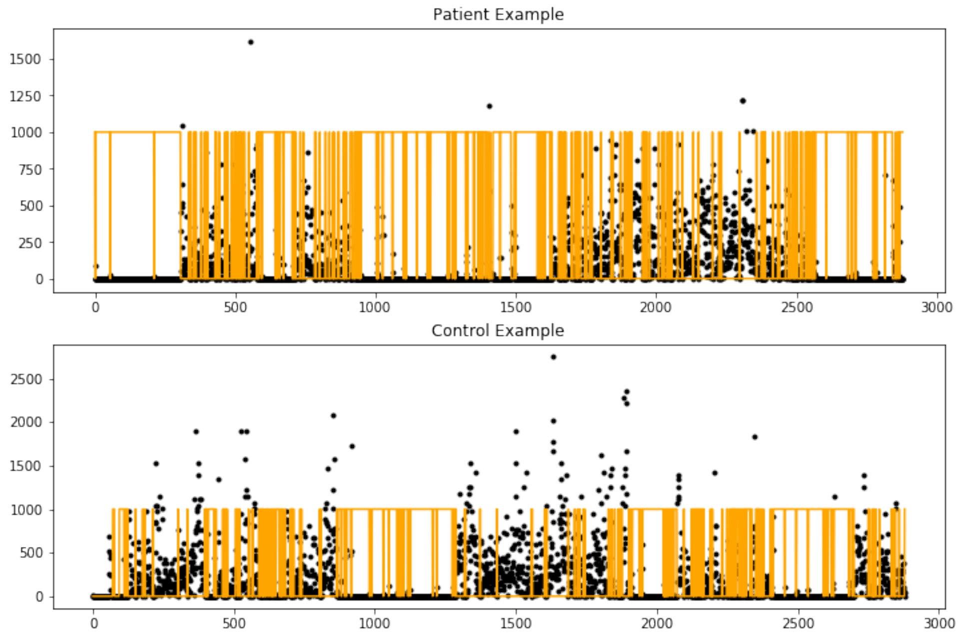
In conclusion, the time-independent model estimates a person's transition probability to rather stay in a certain state of activity. The time-dependent model in turn estimates a higher transition away from the active state.
In both proposed models, a difference in the median of the estimated parameters between the control group and the group of schizophrenic patients can be observed.

Figure 17 presents exemplary the estimated states for the activity time series. To obtain the state estimate, the so-called decoding problem was solved with the Viterbi algorithm. The binary signal in orange represents the estimated states. The activity measured per minute is displayed with a black dot per measurement.
It can be shown that the implemented time-independent Hidden Markov model estimates the states of resting and being active in a reasonable manner. This in turn speaks for the validity of the implemented algorithm.

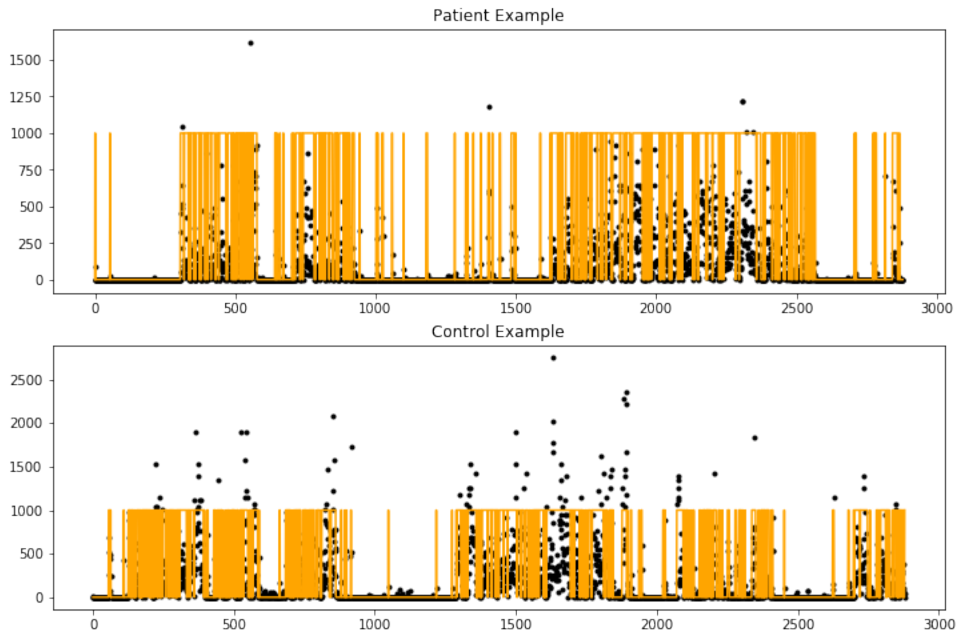
5.3 Analysing indicators of the model fit for the time-independent Hidden Markov model

Due to computation time, only ten iterations of the extended Baum-Welch algorithm which integrates the covariate are conducted. Consequently, the model is not appropriately fitted. Thus, a comparison with the other models would make little sense. This chapter focuses on presenting the results of the time-dependent Hidden Markov model with integrated covariate and to argue for their validity.

Similar to the previous chapter, the estimated transition probabilities for each transition matrix element is visually analyzed. Figure 18 shows the estimated transition probabilities. Unlike in the time-independent model, the probability to stay in the active state, element a_{11}, is lower than the probability to change from being active to resting. The likelihood of resting and



The state estimate of the time-independent Hidden Markov model.



The state estimate of the time-dependent Hidden Markov model

Figure 17: Example visualization of a state estimation in orange together with the actual measured activity in black over two days. The estimated hidden states are exemplary shown for the time series of patient subject Nr.10 and control subject Nr.10. Source: own illustration

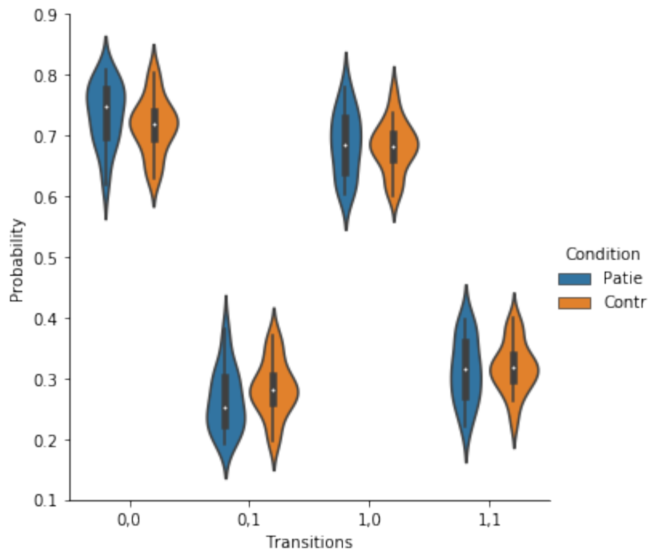


Figure 18: Violin plots are shown for each transition probability matrix element, grouped in the control group and Schizophrenic group of subjects. Source: own illustration

the likelihood of switching back to the resting state are higher than their counterparts. The model estimates, that a person rather stays resting and to transfer back to the resting state. The probability of being in the active state and the probability to transfer back to the active state are low.

A valid model would adjust its estimated transition probabilities according to a daily course. The likelihood of being in the resting state is assumed to be lower during the day and higher during the night. To verify this assumption, the averaged estimated transition probabilities for a 24h daily time interval were analyzed.

Figure 19 shows the average daily transition probability of switching from the resting to the active state for 8 exemplary subjects. The transition probability is similar to the average 24h activity in figure 8. The figure shows a day starting from 9 am until 8:59 am. The likelihood of switching from resting to being active indeed raises after 9 am and stays on a high level during the day. Around 3 am, the probability is the lowest.

It can be concluded that the estimated transition probabilities contain a daily

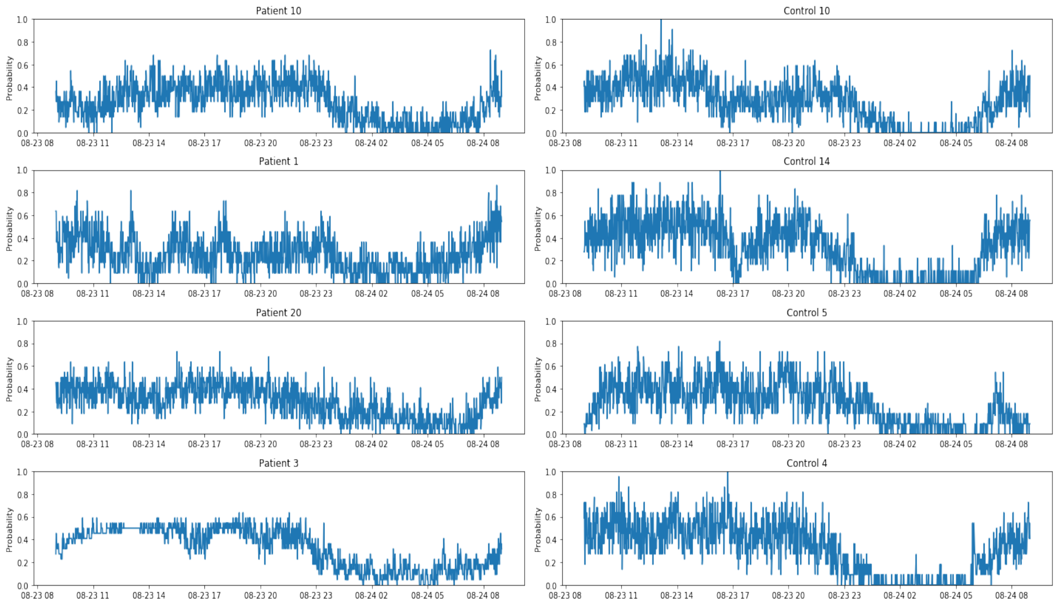


Figure 19: The average estimated transition probability from state resting to state active during 24h. The transition probability is visualized for 8 example subjects, where the patient group is located on the left and the control group on the right. Source: own illustration

course, even if only 10 iterations of Baum-Welch were carried out.
The other transition probabilities are also analyzed and underline this finding.

Even though, the Baum-Welch algorithm for the time-dependent Hidden Markov model with integrated covariate was only performed on ten iterations. The estimated transition probabilities seem to be. To obtain features for the classification, the optimized link coefficients are extracted. Based on the previously listed findings, it is assumed that the link coefficients could give reasonable results while comparing them to the control and the schizophrenic group.

6 Classification

The classification of schizophrenic and non-schizophrenic persons is based on extracted features of their activity time series. As stated in chapter 1.3, the classification of mental health disorders based on activity is already applied broadly in the literature.
The model parameters of the Hidden Markov model and its extensions are used as classification features. The hypotheses put forward in chapter 1.2 are analyzed within this chapter.

With the purpose of evaluating the suggested features, their performance is assessed by comparison of their classification performance with a baseline classification. The classification will be performed by logistic regression. Logistic regression was chosen as easy and interpretable method. Moreover, the $l1$ penalization of the parameters is seen as a valuable addition for the feature evaluation.

Feature selection through a LASSO regression is used to evaluate which set of features derived from the literature performs best. The features from the literature are introduced in chapter 6.2. Second, the parameters of the time-independent Hidden Markov model are evaluated as classification features. Third, the time-dependent transition matrix and its variability over time are used as classification parameters. The transition probabilities and their variance over time are considered as classification features. Finally, the time-dependent transition Hidden Markov model with integrated covariate is used to extract classification features. The additional classification features are the link coefficients between the time-dependent transition probability matrix and the integrated trigonometric function. After, the classification performance of the different Hidden Markov model parameters is compared to the baseline classification.

6.1 Baseline Classification

The features listed in the top part of table 1 in chapter 6.2 are included in the baseline classification. A LASSO regression is performed to classify the schizophrenic and control group. Logistic regression is considered as a method, which provides interpretability of the model fit. Moreover, $l1$ regularization can be applied to select features. The scaling parameter λ of the

$l1$ regularization has to be determined to find an optimal feature selection.

To find the optimal λ, the model performance was evaluated for different scaling parameters λ. The λ which achieved the best Area under the ROC curve (AUC) and average precision is chosen. Standardization is applied before each model is implemented.

First, logistic regression with $l1$ regularization will be conducted to perform feature selection. The left-over features will be analyzed individually, according to their prediction performance and model fit. The feature with the best performance is selected as the basis for stepwise regression. A step-wise forward selection will incorporate one additional feature at a time and evaluate the performance of the model. The classification prediction is evaluated based on leave-one-out cross-validation. The average precision of the leave-one-out cross-validation, the area under the AUC are used as performance metrics. The significance of the single coefficients is evaluated through their t-test p-values. The model fit is evaluated through the pseudo R^2. The pseudo R^2 is developed by McFadden, 1974 and is the quotient of the maximum likelihood estimator and the LL-Null model (McFadden, 1974). The LL-Null model is the model without any independent variable included in the model. In the end, the baseline classification with the best performing features is selected.

Figure 20 shows the correlation matrix of the considered features for the baseline classification. High correlations up to more than 90% can be observed between mean, variance, autocorrelation and RMSSD. All these statistics are very closely related. To avoid multicollinearity, the selection will take the correlation between the features into consideration.

Figure 21 shows the different performance of the LASSO regression for different penalty scaling parameters C. The highest area under the ROC curve (AUC) is achieved at $C = 0.1$, while the average precision reaches maximum at $C = 0.3$. Only features, selected by the LASSO regression with penalty values $C < 0.3$, are considered in the further feature selection.

The course of the coefficient values with increasing penalty parameter C is visualized in figure 22. Figure 22a shows the full course of the coefficient values. The coefficient values of the features mean, variance and autocorrelation

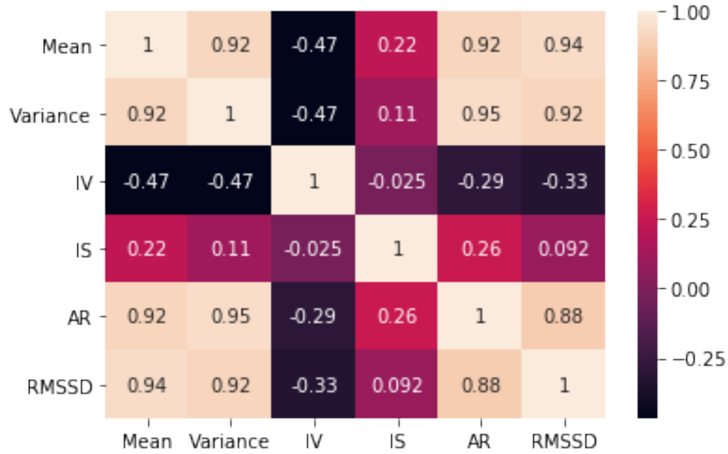


Figure 20: Correlation matrix of baseline classification features. Especially, the mean value and the variance correlating highly with other features like the autocorrelation and the RMSSD. High multicolineraty is implied.

shrinking towards zero before $C > 1$. The interdaily stability (IS), the intradaily variability and the RMSSD are converging towards zero after $C < 10$. To get a better insight, figure 22b provides a cut-out of figure 22b in the interval $C \in [0.5, 0.001]$. It is notable, that the coefficient value of the mean, which already converged to zero, alters from zero between $C \in [0.3, 0.05]$.

Since the optimal AUC and average precision of the LASSO regression are achieved for values of $C \in 0.3, 0.1$, all features which are non-zero in this interval are considered for further feature selection. The coefficients for the intraday variability (IV), the mean and the RMSSD are non-zero during the interval $[0.3, 0.05]$.

The three left-over feature are evaluated according to their classification performance and model fit, independently. A new classification performance statistic is introduced, the Matthew correlation coefficient (MCC).

Table 2 illustrates three different performance and model fit statistics for only one feature each. According to the p-values of t-test for regression coefficients in table 2 each feature was significant. However, the intraday variability (IV) performed weaker for the classification as stated by the MCC. Furthermore, the pseudo R^2 of the IV is the smallest. With 0.742 the mean value scores the same value of MCC, as RMSSD. The mean value provides a better model fit with a pseudo R^2 of 0.436. In conclusion, the mean value will be the base

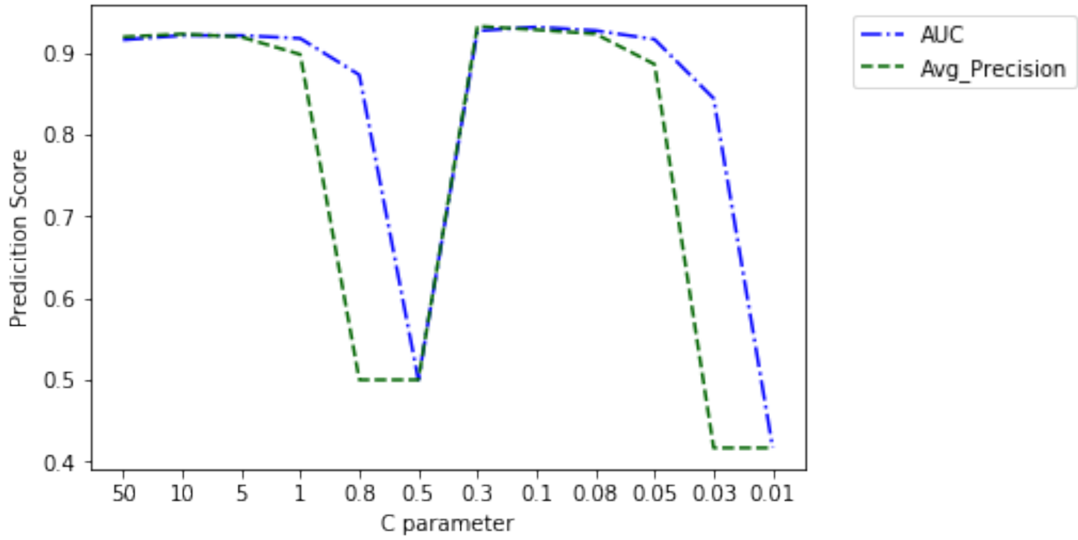
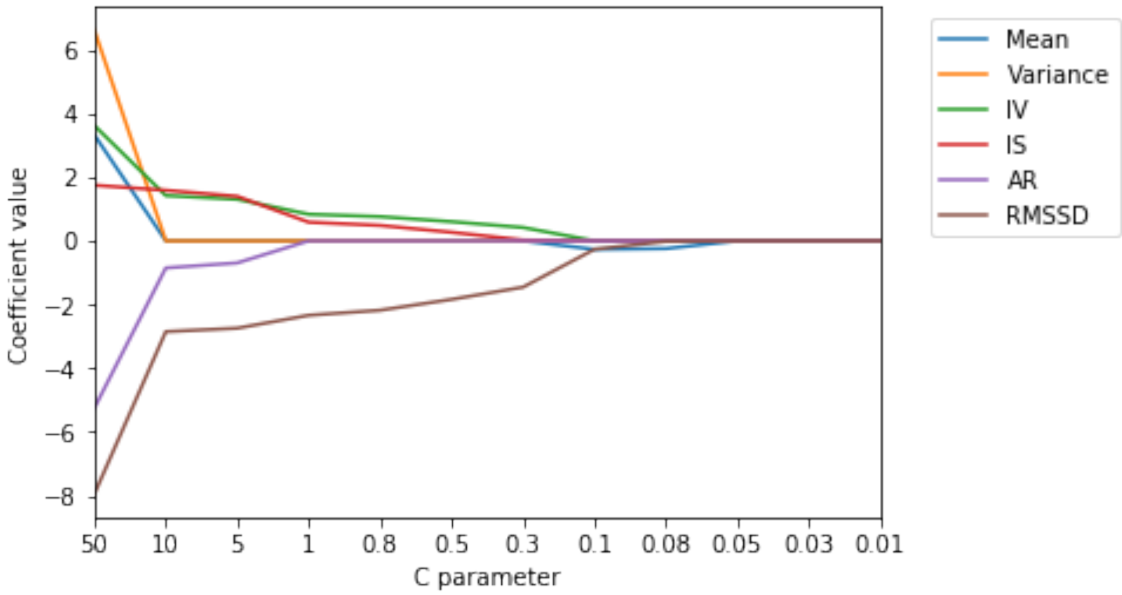


Figure 21: The course of the prediction score AUC for different penalty parameters of C. C denotes the inverse regularization strength, thus the inverse λ (Pedregosa et al., 2011). Source: own illustration

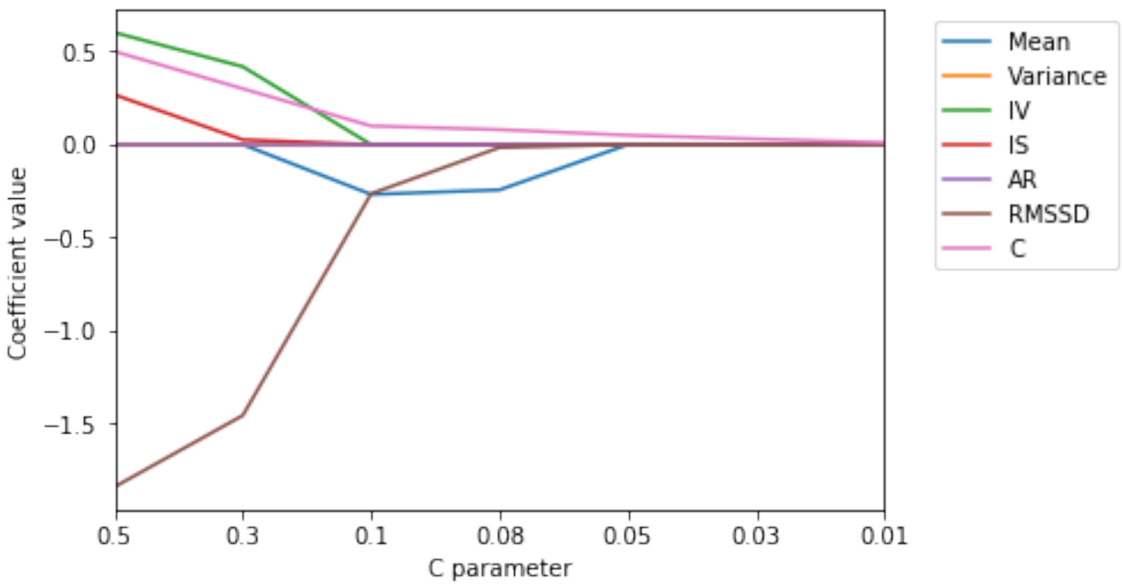
feature for a forward stepwise regression.

The overview of results for the forward stepwise regression is listed in table 3. Since the number of all features is considerably small, each combination with the feature mean as basis was considered. The subset including the {mean, RMSSD} as features performs the weakest. The MCC and the Pseudo R^2 are smaller than for the model including {Mean, IV}. The best MCC and Pseudo R^2 are achieved by the model with all three features included.

However, when the p-value of the t-test of the regressions is analyzed, an interesting observation can be made. Considering the combination of {mean, RMSSD}, the parameter RMSSD turns insignificant. This could be an indication of the high correlation between the two variables. However, the significance level of the parameter IV in the { mean, IV } subset also changes but would be still significant with an $\alpha = 10\%$. The combination of all three features affects the p-value of the mean and turns it insignificant. Due to the high correlation of these two features, RMSSD will not be included in the model.



(a) The coefficient values of the LASSO regression for different penalty values C.



(b) Cut-out of the top graphic, with a focus around the values $C < 0.3$

Figure 22: The coefficient values assigned to different penalty coefficients C values. Features with non-zero coefficient value for $C < 0.3$ are considered for further selection. Source: own illustration

The logistic regression model including the variables mean activity and interday variability are chosen to be the baseline classification model. In the following chapter's this baseline model is ought to be the comparison to the newly introduced features.
In the next chapter, the time-independent transition matrix is introduced as a potential feature for the classification.

6.2 Classification with parameters derived from the time-independent Hidden Markov Model.

A Hidden Markov model with time-independent transition probability is fitted to the activity time series in order to model the resting and active states of a person.

As stated in hypothesis 1, it is presumed that the distribution of the resting and active period might be a valuable feature for classification. Chapter 3 already investigated on the difference between daily and nightly activity since the active and resting periods were unknown. With the Hidden Markov Model the active and resting periods were determined. Thus, the first two moments, which describe the distribution of the two states in the Hidden Markov model are used as features for the LASSO regression model.

Referring to hypothesis 2, it is studied if there might be valuable information within the probabilities of switching between the activity states. Due to the disturbance of the circadian rhythm of a schizophrenic person, the probabilities of the transition matrix are considered as informative classification features.

The derived parameters from the Hidden Markov Model are listed below.

- The mean value of the two states
- The variance of the two states
- The transition probabilities

The same process as in chapter is used to achieve comparable results in feature performance. A LASSO logistic regression is applied.

Before the logistic regression is applied, the correlation matrix of the used features is analyzed. The correlation matrix is shown in figure **??**. Figure **??** shows, that there is a perfect linear relationship between the transition probabilities of staying in a state and switching. The transition probabilities of staying in a state, active or resting, are removed, to avoid multicollinearity and allow a feasible parameter estimation. This high correlation occurs due to the normalization of the rows of the transition probability matrix. The probability to stay in the resting state and the probability of transit from the resting state to the active state sum up to one.
The correlation between the mean value of the active state and the variance of the active state is 0.88. But will not be taken out of the feature set, since the LASSO regression should cope with correlated feature sets by penalizing the coefficients.

The logistic regression is conducted several times for different parameters of C to find the optimal penalty parameter. It is considered, that the optimal penalty parameters are the one maximizing the AUC and average precision. Figure 24 presents the course of the AUC and average parameters for different C parameters.

Similar to the baseline model, the model performance drops in the interval $C \in [1, 0.5]$. The performance rises after to a second but marginally lower peak for $C \in [0.1, 0.08]$. The difference in performance between the two peaks lays by 0.001 for the average precision and 0.01 for the AUC. Thus, there will be a closer look on the feature selected by $C \in [1, 0.5]$ and $C \in [0.1, 0.08]$. The model with little features but nearly equal performance will be selected. The overall performance evaluation will be done after the single features are selected.

Figure 25 shows a comparable plot to chapter 6.1 - the coefficient values for each feature for different $l1$ penalty parameters C.
It can be observed, that all features but the mean value for the active state shrinking toward zero for $C < 0.3$. In figure 24 it is observed, that the performance of the classification is nearly as good for $C < 0.3$ than for the maximum score obtain for $C \in [1, 0.5]$.
In conclusion, a model including only the feature of the mean activity during the active state has as much predictive power as all features together.

In the following, the fit of the proposed logistic regression model to the data

Features	Pseudo R^2	p-value of t-test	MCC
Mean value	0.436	0.00	0.742
RMSSD	0.306	0.00	0.742
IV	0.2436	0.01	0.524

Table 2: Performance measures of the literature feature. The features derived by literature are chosen by LASSO feature selection.

Combination	Features	Pseudo R^2	p-value of t-test	MCC
{Mean, IV}	Mean IV	0.490	0.002 0.069	0.698
{Mean, RMSSD}	Mean RMSSD	0.438	0.098 0.701	0.698
{Mean, RMSSD, IV}	Mean RMSSD IV	0.5462	0.526 0.021 0.069	0.702

Table 3: The performance statistics of the literature features as sets of three variables.

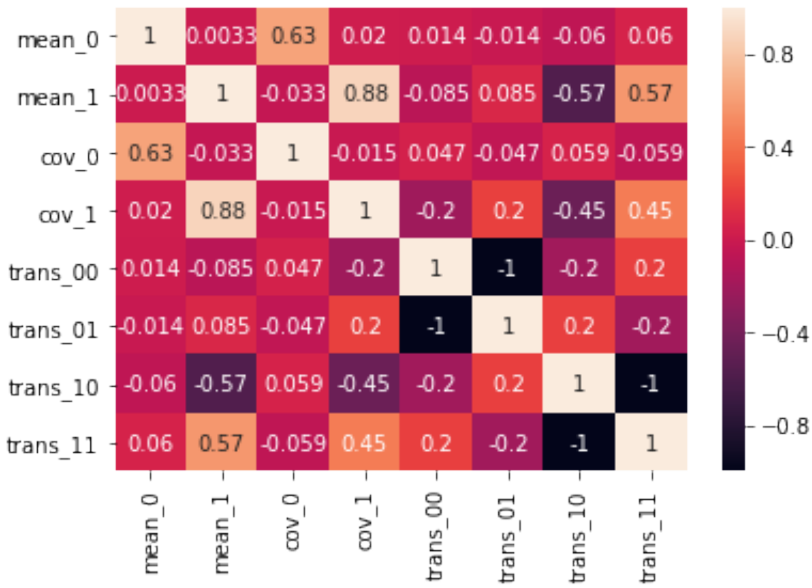


Figure 23: Correlation matrix of features derived from the Hidden Markov model. The resting state is annotated with 0, the active state with 1. Source: own illustration

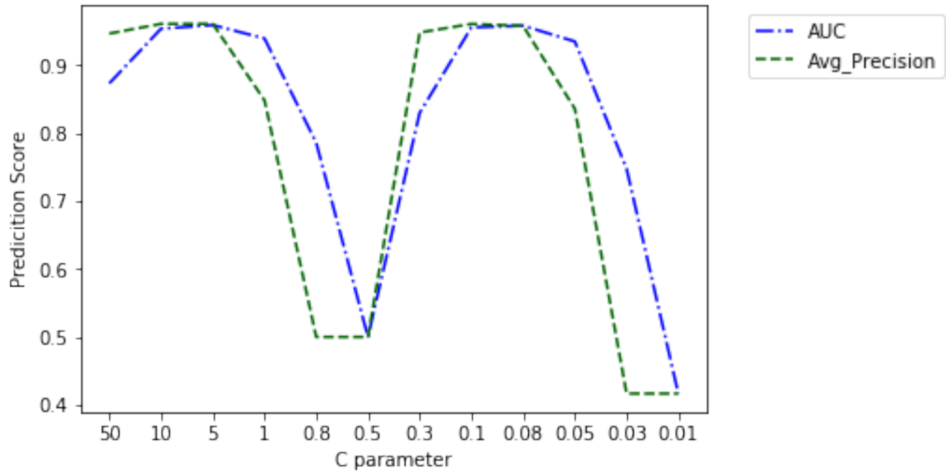


Figure 24: The AUC and the average precision obtained by the LASSO regression for different penalty parameters C. Source: own illustration

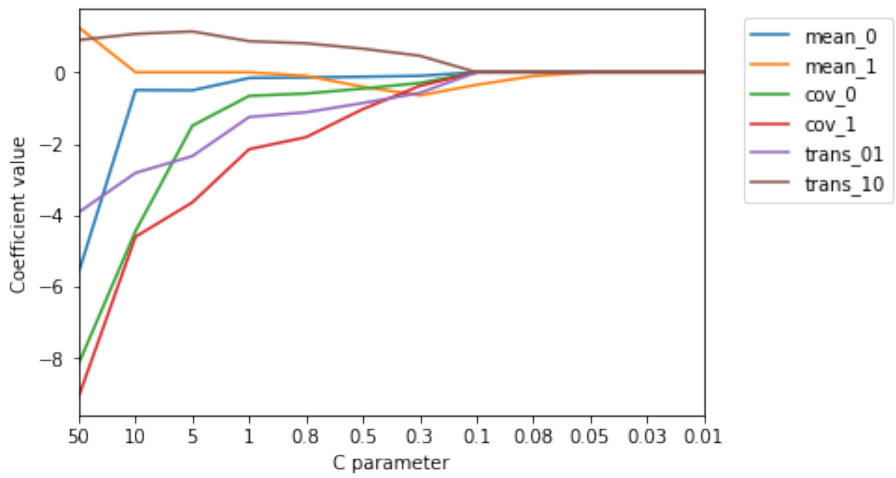


Figure 25: Evaluation of different feature coefficient for different penalty coefficients C values. Source: own illustration

is analyzed. As in chapter 6.1, the significance of the single coefficients is evaluated through their t-test p-values. The model fit is evaluated through the pseudo R^2. A stepwise forward regression is applied.

The selection criterion is the same as in chapter 6.1, namely the Matthews' correlation (MCC) coefficient, t-test p-values for significance and the pseudo R^2.

The results of the last iteration of the stepwise logistic regression are presented in table 4. The model, including the variance of the active state, the transition probability to change from the active to the resting state and vice versa, has a higher MCC and Pseudo R^2.
The p-values reveal, that the variance of the active state and the transition probability from resting to active are significant in the first model. The transition probability from active to resting meets the significance level of $\alpha = 10\%$.

The second model in table 4 has a lower MCC but confirms the significance of the features variance of the active state and the transition probability from resting to active. Both, meeting a significance level of $\alpha = 5\%$. The transition probability from resting to active is not considered as significant. These results confirm the hypothesis 1 and the hypothesis 2.

The significance of the variance and mean value of the active state underline the hypothesis 1. Hypothesis 1 stated, that the distribution of the two different activity states of a subject can be a valuable classification feature.

Hypothesis 2 claims, that the transition probabilities between the activity states withhold valuable information for the classification. In fact, especially the transition probability from the resting to the active state turned out to be a significant variable for the classification.

In the next chapter, the validity of hypothesis 3 and 4 is studied. Features derived from the time-dependent Hidden Markov model and from the time-dependent Hidden Markov Model with integrated covariate are analyzed by their classification performance and model fit.

6.3 Classification with parameters derived from the extended Hidden Markov model.

After investigating on the validity of the first two hypotheses, the validity of hypothesis 3 and hypothesis 4 are studied.
The evaluation procedure follows that of the previous two chapters. The penalty parameter C is determined by the models' classification performance by according penalty parameter.
The left-over features and their significance are further analyzed by applying a stepwise regression model. This will determine the features with the best performance, according to selection criteria.

First, the model parameters of time-dependent Hidden Markov model are analyzed. In the second part, the model parameters time-dependent Hidden Markov model with integrated covariate are extracted and evaluated as features for the classification task.

6.3.1 Classification with model parameters derived from the time dependent Hidden Markov Model.

Hypothesis 3 claims that the fluctuation of the time-dependent transition probability might be a relevant feature to classify between Schizophrenic and non-Schizophrenic subjects.

The derived parameters from the time dependent Hidden Markov Model are listed below.

- The mean value of the two states
- The variance of the two states
- The transition probabilities over time:
  - The variance of the transition probabilities over time
  - The mean transition probability

While the mean and variance of the two estimated activity states is similar to the features in the chapter above, the time dependent transition probabilities might introduce new information. It is assumed that the fluctuation

Combination	Features	Pseudo R^2	p-value of t-test	MCC
{Var1, Trans01, Trans10}	Var1 Trans01 Trans10	0.674	0.004 0.008 0.098	0.828
{Var1, Mean1, Trans01}	Var1 Trans01 Mean1	0.643	0.004 0.041 0.351	0.784

Table 4: The performance statistics for the two best performing subsets. A step wise forward logistic regression has been applied.

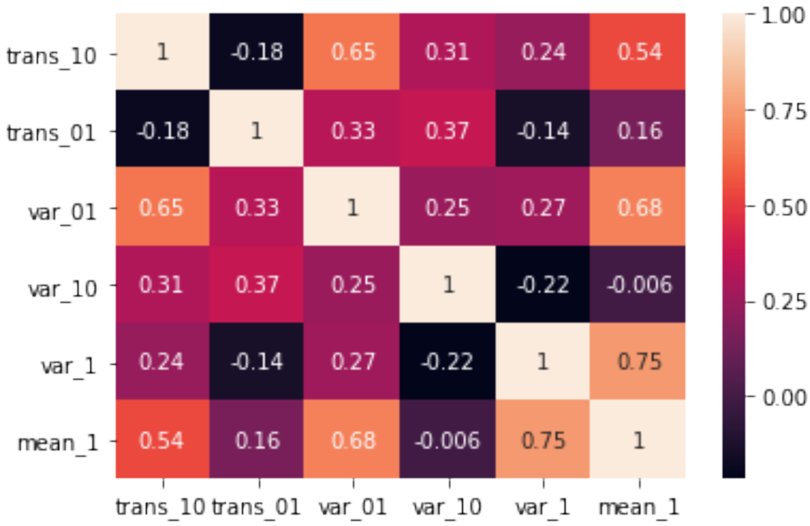


Figure 26: The correlation matrix of features extracted from the Hidden Markov model with time dependent transition probabilities. Source: own illustration

of the transition probabilities can be capture through the overall variance of the time dependant transition probabilities. The mean values of the time dependant transition probabilities are basically the estimated time-independent transition probabilities.

Figure 26 shows the correlation matrix of the considered features. Due to multicollinearity, the time-dependent mean transition probabilities for staying in each state were excluded. Also, the variance of these transition probabilities as well as the mean and variance of the resting period were excluded. The mean and variance of the resting period are causing high correlation with the ones of the active period. From the earlier analysis, it is known, that the parameter estimates of the active state are significant. Thus, the ones from the resting state were excluded.

The classification performance is highest in the interval of $C \in [50, 10]$, according to the AUC and average precision, respectively. Figure 27 visualizes the classification performance for different parameters of C. The feature which have not been shrunk towards zero at $C \in [50, 10]$ are further considered for the modelling.

The course of the coefficient values for decreasing values of the penalty parameter C are presented in figure 28. The parameter estimate of the mean in the active state, the variance of the time-dependent transition probability from resting to active and active to resting, and the transition probability from resting to active, are selected.

Stepwise forward regression was conducted on the subset of the four stated features. The results of the last iteration is presented in table 5.

The model presented in the first row of table 5 shows, that all included variables are significant at a level of $\alpha = 5\%$. This does not count for the second model, in which no variable is significant. The pseudo R^2 is with a 0.08 difference slightly better in the second model.
A greater difference is between the Matthews' correlation coefficients. The first model achieves the highest MCC yet. The first model is clearly in favor, due to the better classification performance and the significance of the included variables.

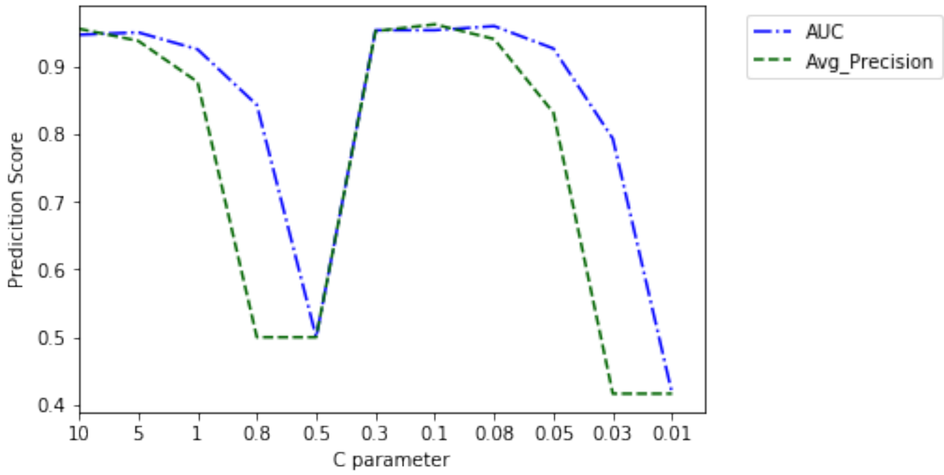


Figure 27: The penalty parameter C is evaluated by the classification performance, measured by AUC and the average precision. Source: own illustration

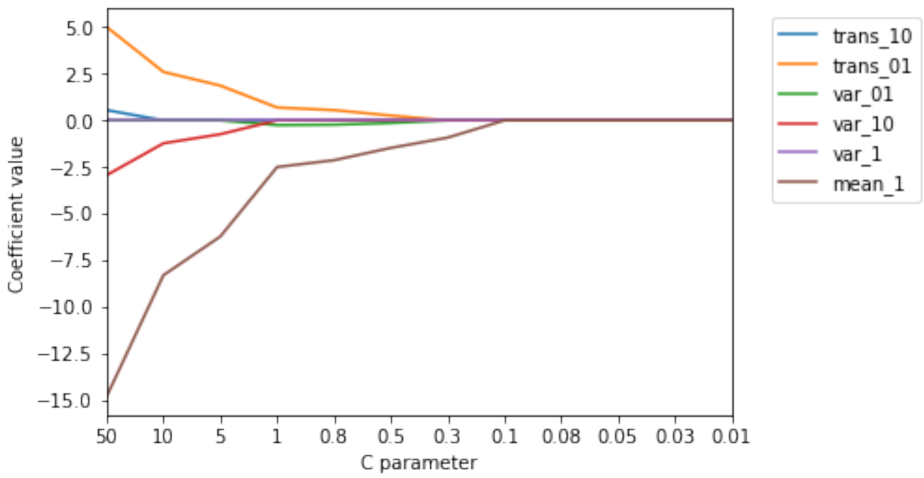


Figure 28: The coefficient values of each included features for different penalty parameter values C. According to figure 27, the cut off for the feature selection is at $C = 0.1$. Source: own illustration

The results underline hypothesis 2 since the transition probability is found to contribute to a good classification. Hypothesis 3 is confirmed by this method. The variability in the time-dependent transition probabilities is selected as the valuable feature variable for a good performing classification model. So far, three out of four hypotheses are confirmed within this work. The next chapter will examine the validity of the last hypothesis 4.

6.3.2 Classification with model parameters derived from the time dependent Hidden Markov Model with integrated covariate.

Before the classification results of the last proposed model are presented, it must be referred to the model fit. A limitation of the proposed model is the small number of expectation-maximization iterations executed. Due to the high computation time of the extended maximization, to incorporate the covariate, only ten iteration steps of the proposed extended Baum-Welch algorithm are executed. This can effect the goodness of fit of the model.

This chapter studies the validity of hypothesis 4 introduced in chapter 1.2. Hypothesis 4 states, that the link coefficients used to integrated a trigonometric function might capture the disruption of the circadian rhythm of a schizophrenic person, compared to a non-schizophrenic person. It is presumed, that the time-dependent transition probabilities of a non-Schizophrenic subject possibly follow a trigonometric course already. Hence, the linking coefficients might differ between the two groups.

Newly introduced in this chapter are the link coefficients with which the trigonometric function is integrated into the time-dependent transition probabilities during the Baum-Welch algorithm. The evaluation procedure follows the chapters above. A LASSO regression for different penalty parameter values C is applied first.

The correlation matrix of the considered features of the time-dependent Hidden Markov model with integrated covariate is presented in figure 29. The same problem as introduced in chapter is faced. The moment estimates for the resting or active state cause a high correlation among each other. Thus, the estimates of the resting state are excluded. Unlike the time-dependent Hidden Markov model from the above chapter, the transition probabilities revealing high negative correlations. This might be implied by the weak

Combination	Features	Pseudo R^2	p-value of t-test	MCC
{Mean1 ,Trans01, ,Var01}	Mean1 Trans01 Var01	0.882	0.03 0.036 0.035	0.914
{Mean1, Trans01, Var10}	Mean1 Trans01 Var10	0.890	0.191 0.185 0.211	0.749

Table 5: The performance statistics of the two best performing subsets of the time dependent Hidden Markov model parameters.

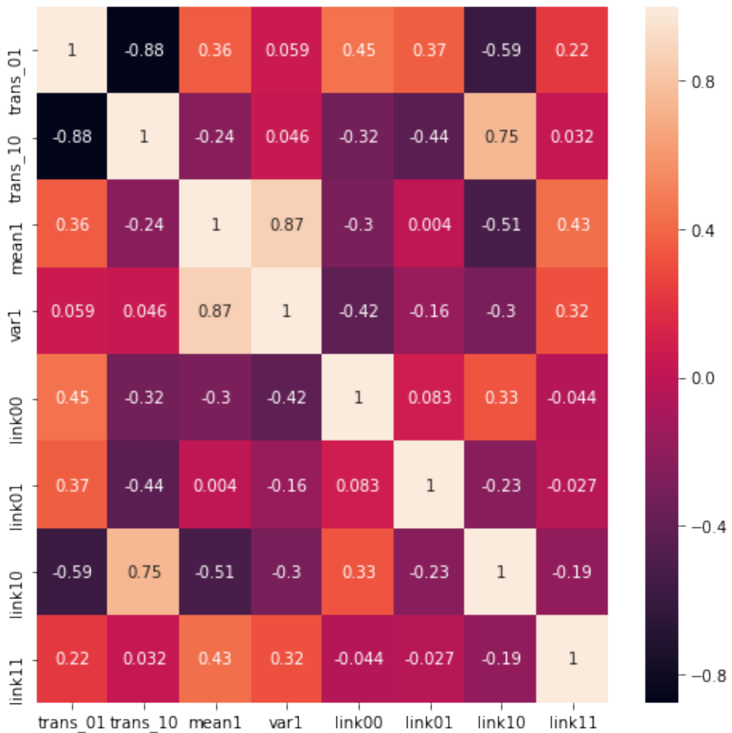


Figure 29: Correlation matrix of features derived from the Hidden Markov model with integrated covariate. The resting state is annotated with 0, the active state with 1. The moment estimates of the resting state are excluded. Source: own illustration

model fit. The transition probabilities seem to not have converged enough.

The penalty coefficient C, which maximizes the classification performance measured by AUC and the average precision is chosen as the cut-off criterion for the feature selection. Since the procedure follows strictly the ones from the chapters above, only the results will be presented.
A maximum AUC and average precision is measured for the values $C = \{10, 1\}$, respectively.

The selected features are

- The mean value of the active state
- The link coefficients of the intercept and the covariate coefficient
- The transition probability from active to resting

The features are selected because their coefficients are non-zero at the penalty parameter of value $C = 1$.
The results of the last iteration of the stepwise forward regression are listed in table 6. According to the pseudo R^2 the model including the link-coefficient for the intercept of the covariate fits better to the given data. The same model performs better in the classification, too, according to the MCC.
In both models are the mean value of the active state as well as the transition probability, switching from active to resting, found to be significant. The link coefficients of the integrated covariate however are not found to be significant.

6.4 Comparison of Feature Performance

After hypotheses 1 to 4 are examined for validity, the best performing feature sets of each proposed model will be compared. The focus will lay on the comparison of the new proposed feature and the features used in the literature.

Table 7 summarises the best performing feature sets of each proposed model. According to the MCC, the feature set of the time-dependent Hidden Markov model performs best in classification. The set of the mean value of the active state, the transition probability of changing from resting to being active, and

Combination	Features	Pseudo R^2	p-value of t-test	MCC
{Mean1 ,Trans10, ,Link inter.}	Mean1 Trans10 Link inter.	0.722	0.003 0.084 0.204	0.784
{Mean1, Trans10, Link coef.}	Mean1 Trans10 Link coef.	0.698	0.003 0.038 0.738	0.749

Table 6: The performance statistics for the two best performing subsets of the time dependent Hidden Markov model parameters with integrated covariate. The performance of the link parameters are additionally evaluated.

Combination	Features	Pseudo R^2	p-value of t-test	MCC
{Mean, RMSSD, IV}	Mean RMSSD IV	0.5462	0.526 0.021 0.069	0.702
{Var1, Trans01, Trans10}	Var1 Trans01 Trans10	0.674	0.004 0.008 0.098	0.828
{Mean1 ,Trans01, ,Var01}	Mean1 Trans01 Var01	0.882	0.03 0.036 0.035	0.914
{Mean1 ,Trans10, ,Link inter.}	Mean1 Trans10 Link inter.	0.722	0.003 0.084 0.204	0.784

Table 7: The overall comparison of the performance statistics between the best performing feature sets of each model.

the variability of the same transition probability over time achieve a MCC of 0.914. Moreover, the same feature set achieves the highest pseudo R^2 of 0.882. The included features in the logistic regression model are also significant. Thus, this feature set outperforms the other feature sets by every performance statistics.

The feature set provided by the time-independent model achieves the second highest MCC with 0.828. However, the model fit is according to the pseudo R^2 not as good as the one of the feature set provided by the time-dependent Hidden Markov model with an integrated covariate.

More importantly, all three proposed models achieve better performance statistics than the feature set of the literature. Therefore, the newly introduced features obtained by the three different Hidden Markov models improved the classification performance of conventional features of literature. These results only account for the given data set and the applied classification method of logistic regression.

7 Discussion

In the beginning of the work in chapter 1.2, four different hypotheses are proposed. The presented results in the antecedent chapter and their implications to the hypotheses are discussed. In addition, the results are discussed with regard to the work of Jakobsen, Garcia-Ceja, et al., 2020. Jakobsen, Garcia-Ceja, et al., 2020 provides a baseline classification for the data set on which this thesis is based.

The validity of the hypothesis 1 is shown throughout all models applied. At least one of the selected and significant features is one of the two moments of the estimated activity states. The variance of the active state of the time-independent Hidden Markov Model is included in the best performing regression model.
The mean value of the active state of the two time-dependent Hidden Markov models is included in the classification model.
Both, variance and mean of the active state showed to have a high predictive value. Due to high correlation among themselves, either mean or variance is implemented.

The validity of hypothesis 2 is also confirmed throughout all the models applied. Both transition probabilities present the switch between states are found to be significant features in the logistic regression models. The hypothesis 2 focuses on the transition probabilities between states, rather than on the one describing to stay in a certain state. To avoid multicollinearity, these transition probabilities were not considered for modeling the logistic regression.

The validity of the hypothesis 3 is studied in chapter 6.3.2. The feature of the variability of the transition probability from resting to active was found to be significant and not highly correlated to the according estimated time-independent transition probability. Hence, it is concluded that the variability of the transition probability can be used as a valid classification feature.

According to chapter 6.3.2, the validity of hypothesis 4 cannot be confirmed through the evaluation procedure applied in this work. The results of the evaluation procedure did not find the features of hypothesis 4 to be significant in explaining the difference schizophrenic and non-schizophrenic person. Thus, hypothesis 4 can not be validated within this work.

Feature	Pseudo R^2	coefficient	p-value of t-test	MCC	AUC
Link coef.	0.288	-1.31	0.003	0.306	0.77
Link inter.		1.05	0.008		

Table 8: The performance statistics and coefficients of the logistic regression model with the link coefficient of the intercept of the covariate and the coefficient of the covariate.

In order to further investigate the significance of the variables, the link coefficients are used as the only variables in a logistic regression model.
Indeed, both link coefficients turn out to be significant variables at a confidence level of $\alpha < 5\%$. The pseudo R^2 of the model is $R^2 = 0.288$. The Matthew's correlation coefficient is 0.31, while the AUC is 0.77. The results are summarised in table 8.

To better understand if the link coefficients have an impact on the classification, the regression coefficients are interpreted.

According to equation (37) the link coefficients integrate a trigonometric function into the time-dependent transition probabilities. The coefficients for integration are an intercept and a covariate coefficient. These two link coefficients can be interpreted as the vertical shift and the amplitude of the trigonometric function. The vertical shift moves the trigonometric function in direction of the y-axis. The amplitude measures the change of the trigonometric function during a single period. The higher the amplitude, the higher is the difference between high and low of the trigonometric function.

Since the logistic regression is applied, the regression coefficients have to be interpreted differently. If the value of the coefficient increases by one, the expected change in the log odd ratio changes by the coefficient value.
This is formulated by the following equation:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{41}$$

The coefficients in table 8 are log odd ratio. Thus the exponent has to be calculated. If the vertical shift increases by one, so increase the likelihood of being schizophrenic by 2.84 times. If the amplitude increases by one, so increase the likelihood of being schizophrenic by 0.27 times. This interpretation has to be adapted to the transition probabilities, into which the trigonometric function is integrated. During the feature selection, two link coefficients are selected. The intercept link coefficient of the transition probability of staying active is selected. Additionally, the coefficient of the covariate for the transition probability of staying in the state rest is selected. Now, the interpretation can be translated in terms of the transition probabilities.

If the vertical shift of the probability to stay active increases by one the likelihood of being Schizophrenic increases by 2.84 times. The increase of the amplitude of the probability to stay in state rest by one increases the likelihood of being schizophrenic by 0.27 times. To clarify, the link coefficients represent the magnitude needed to integrate a trigonometric function.

Another formulation could be the following:
If the probability to stay active is shifted vertically by one to integrate the covariate, then the likelihood of being schizophrenic is 2.84 times higher. Or, the more vertical shift is needed to be integrated, the more likely it is that one is schizophrenic.
Thus, the link coefficients could indeed present the measurements of disruption of the circadian cycle of schizophrenic persons. However, it does not contribute significantly to the classification task, compared to other features.

In conclusion, the parameter estimates from the time-dependent Hidden Markov model with integrated covariate are found to be informative in explaining the difference between schizophrenic and non-schizophrenic. Even though, the predictive value of the link parameters is not comparable with the other model parameters.

The potential of newly introduced features has only been analyzed by one classification method, namely the LASSO logistic regression. Lasso regression is known for the fact that correlated variables are mutually exclusive by regulation. As shown in chapter 6.2, some included variables are correlated with each other. The presented approach allowed the regularization of potentially excluded valuable feature variables, before evaluating them.
Thus, for an overall conclusion on the significance and the predictive value of the Hidden Markov model parameters, each variable has to be analyzed independently. Moreover, more than only one classification method has to be applied.

At the time of the delivery of this thesis, Jakobsen, Garcia-Ceja, et al., 2020 and Jakobsen, Ceja, et al., 2020, are the only comparative analyses of the given data. Jakobsen, Garcia-Ceja, et al., 2020 provides a comparison of different machine learning classifiers. The intention is to provide a baseline performance for future work on the data set (Jakobsen, Ceja, et al., 2020). The classification is done on the ground of three extracted features: mean, standard deviation and proportion of zeros. The features are calculated per full day per patient. The data set used thus comprises 687 data points, whereas 285 are schizophrenic and 402 are non-schizophrenic. Leave-one-out cross-validation is used for evaluation. Precision, False-Positive-Rate, Recall, F1-Score and the Matthews Correlation Coefficient are the chosen evaluation metrics.

It is important to mention that the Matthews Correlation Coefficient provided by Jakobsen, Garcia-Ceja, et al., 2020 and the Matthews Correlation Coefficient obtained within this work are not directly comparable. Jakobsen, Garcia-Ceja, et al., 2020 break up the time series into days for the feature extraction and in the end to obtain more data. The presented approach within this master thesis extracts the features from the model parameters of Hidden Markov models, which are fitted to the time series. Thus, the amount of data is less.

Nevertheless, both approaches pursue the same goal of predicting the mental illness schizophrenia. In the work of Jakobsen, Garcia-Ceja, et al., 2020, the best classifier is Random Forest, which achieves a Matthews Correlation Coefficient of 66.2% on leave-one-out cross-validation Jakobsen, Ceja, et al., 2020.
The best classifier presented in this work is obtained with logistic regression on leave-one-out cross-validation. The included features are the mean, the transition probability to switch from the resting state into the active state and the variance of exactly this time-dependent transition probability. The classifier achieves a Matthews Correlation Coefficient of 91%. In general, the applied logistic regression on the presented features sets, from the literature and the model parameters of the Hidden Markov model, performed better than the obtained Matthews Correlation Coefficient of 66.2% in Jakobsen, Garcia-Ceja, et al., 2020.

Needless to say, comparing the two results comes close to comparing pears and apples. However, the contrast between the two different approaches should underline the need for a broader evaluation of the suggested model parameters as classification features. Additionally, it should show the potential of the here presented features on the course of finding a good classification method to detect Schizophrenia on the base of activity time series.

8 Conclusion

Symptoms of schizophrenia include disorganized behavior, behavioral abnormalities and especially the disruption of the circadian rhythm. The circadian rhythm and a potential disruption are embedded in a persons' activity record. The identification of schizophrenia through activity is approached by modeling the circadian rhythm and quantifying its potential disruption.

This work presents features derived from parameters of a Hidden Markov model to improve the classification of the disease of schizophrenia by a persons' activity. The Hidden Markov Model is extended by modeling the transition probabilities of being in a certain state as time-dependent and describing them by a trigonometric function. The extension of the Hidden Markov model and the according extension of the Baum-Welch algorithm is implemented in the scale of this work.
The three different versions of the Hidden Markov model are applied to the activity time series of each participant. Four hypotheses were put forward, each introducing different model parameter of the Hidden Markov model as potential classification feature.

For the validation of the hypotheses, already applied features for schizophrenia classification derived from the literature are used as baseline comparison. The results are limited to the given data set and the feature selection and classification method used. Hypotheses 1 to 3 are found to be valid in the given conditions. The mean and variance of the active state distribution, the transition probability from the resting to the active state and vice versa, the variance over the time-dependent transition probability from the resting to the active state and the intercept link coefficient are found to be significant. Moreover, each model including the model parameters obtained better classification results according to the Matthews Correlation coefficient, then the features derived from the literature.
Clearly, these results need to be further evaluated on other data sets and with other classification methods. It can be concluded that the chosen approach

introduces new potential classification features in the quantitative research on schizophrenia.

Finally, a personal note is added. Next to the goal of improving the identification of schizophrenia with quantitative methods, I would like to raise awareness on mental health diseases. My hope for the future is that there will be more research on mental health and, above all, a general acceptance of mental illness in society.

9 Outlook

During the process of writing, different approaches are evaluated. One approach is implemented but not included in the thesis to not expand the scale of the work. The outlook will briefly present this approach.

A Gaussian mixture model on a time-embedded time series representation was considered to model the different activity states. The time series $\mathbf{x} = x_n$ is represented by a trajectory matrix, where each row is defined by Povinelli et al., 2006 Eirola and Lendasse, 2013.

$$x_n = [x_{n-(d-1)\tau}, ...x_{n-\tau}, x_n] \tag{42}$$

According to chapter (3), the two states, active and resting, follow two different normal distributions. The Gaussian Mixture model classifies the two assumed activity states - resting and active, for each time step. The model provides labeling over time. The obtained binary time series is then modeled by a Markov Chain.
The proposed model is a simplified Hidden Markov model. Similar model parameters, from the fitted Gaussian Mixture model and the Markov Chain as the Hidden Markov model are expected to be extracted as features for the classification.

One of the main points within this work is the integration of a covariate in the time-dependent transition probabilities. The covariate was modeled as a trigonometric function 38. The link parameters are the offset and the amplitude of the trigonometric function. Potentially different link coefficients can be included like the phase shift.

Moreover, a trigonometric function is not the only possible covariate that can be included. The covariate gives space to include additional information into the Hidden Markov model. Potentially, there are better models to describe the circadian rhythm than a trigonometric function. Hammer et al., 2017 used an autocorrelation process as modeled covariate to forecast CPU consumption of virtual machines. This approach can be extended to an integrated autoregression moving average model for example. However, the likelihood function of the transition probabilities has to be maximized in an external step. The more parameters are included the more computational complex is the extended EM algorithm.

The evaluation of the suggested features is done on one classification algorithm. The discussion in chapter 7 concluded that the proposed features should be evaluated using a more comprehensive number of classification algorithms than just one. The classification algorithms applied for a baseline performance in Jakobsen, Garcia-Ceja, et al., 2020 can be used as a reference.

Acknowledgment

I would like to pay my special regards to the Simula Metropolitan Center for Digital Engineering and my extraordinary supervision by:
Dr. Michael Riegler
Dr. Paal Halvorsen
Dr. Hugo Lewi Hammer

Moreover, I would like to thank my supervisor at the Institute of Operations Research - Analytics and Statistics at the Karlsruhe Institute of Technology:
Prof. Dr. Oliver Grothe

References

Afonso, P., Figueira, M. L., & Paiva, T. (2014). Sleep–wake patterns in schizophrenia patients compared to healthy controls. *The World Journal of Biological Psychiatry*, *15*(7), 517–524.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Albert, M. V., Sugianto, A., Nickele, K., Zavos, P., Sindu, P., Ali, M., & Kwon, S. (2020). Hidden markov model-based activity recognition for toddlers. *Physiological Measurement*, *41*(2), 025003.

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, *26*(3), 342–392.

Apiquian, R., Fresán, A., Muñoz-Delgado, J., Kiang, M., Ulloa, R. E., & Kapur, S. (2008). Variations of rest–activity rhythm and sleep–wake in schizophrenic patients versus healthy subjects: An actigraphic comparative study. *Biological Rhythm Research*, *39*(1), 69–78.

Banachewicz, K., Lucas, A., & Van Der Vaart, A. (2008). Modelling portfolio defaults using hidden markov models with covariates. *The Econometrics Journal*, *11*(1), 155–171.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, *41*(1), 164–171.

Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., & Zoran, M. J. (2005). Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nature Reviews Genetics*, *6*(7), 544–556.

Bengio, Y. (1999). Markovian models for sequential data. *Neural computing surveys*, *2*(199), 129–162.

Berle, J. O., Hauge, E. R., Oedegaard, K. J., Holsten, F., & Fasmer, O. B. (2010). Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. *BMC research notes*, *3*(1), 1–7.

Billingsley, P. (1961). Statistical methods in markov chains. *The Annals of Mathematical Statistics*, 12–40.

Bilmes, J. A. Et al. (1998). A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, *4*(510), 126.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Brémaud, P. (2020). *Probability theory and stochastic processes*. Springer.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry research*, *28*(2), 193–213.

Carr, O., Andreotti, F., Saunders, K. E., Palmius, N., Goodwin, G. M., & De Vos, M. (2020). Monitoring depression in bipolar disorder using circadian measures from smartphone accelerometers. *arXiv preprint arXiv:2007.02064*.

De los Santos, H., Collins, E. J., Hurley, J. M., & Bennett, K. P. (2017). Circadian rhythms in neurospora exhibit biologically relevant driven and damped harmonic oscillations, In *Proceedings of the 8th acm international conference on bioinformatics, computational biology, and health informatics*.

Docx, L., Sabbe, B., Provinciael, P., Merckx, N., & Morrens, M. (2013). Quantitative psychomotor dysfunction in schizophrenia: A loss of drive, impaired movement execution or both? *Neuropsychobiology*, *68*(4), 221–227.

Domingues, A., Paiva, T., & Sanches, J. M. (2013). Sleep and wakefulness state detection in nocturnal actigraphy based on movement information. *IEEE Transactions on Biomedical Engineering*, *61*(2), 426–434.

Eirola, E., & Lendasse, A. (2013). Gaussian mixture models for time series modelling, forecasting, and interpolation, In *International symposium on intelligent data analysis*. Springer.

Destatis. (2019). Todesursachen: Suizide [https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Todesursachen/Tabellen/suizide.html].

Forney, G. D. (1973). The viterbi algorithm. *Proceedings of the IEEE*, *61*(3), 268–278.

Gaebel, W., Möller, H.-J., & Rössler, W. (2005). Stigma-diskriminierung-bewältigung. *psychoneuro*, *31*(03), 159–159.

Gaebel, W., & Wölwer, W. (2010). Gesundheitsberichterstattung des bundes heft 50. *Berlin: Rober Koch Institut*.

Gonçalves, B. S., Cavalcanti, P. R., Tavares, G. R., Campos, T. F., & Araujo, J. F. (2014). Nonparametric methods in actigraphy: An update. *Sleep Science*, *7*(3), 158–164.

Hammer, H. L., Yazidi, A., & Begnum, K. (2017). An inhomogeneous hidden markov model for efficient virtual machine placement in cloud computing environments. *Journal of Forecasting*, *36*(4), 407–420.

Hauge, E. R., Berle, J. Ø., Oedegaard, K. J., Holsten, F., & Fasmer, O. B. (2011). Nonlinear analysis of motor activity shows differences between schizophrenia and depression: A study using fourier analysis and sample entropy. *PloS one*, *6*(1), e16291.

Huang, Q., Cohen, D., Komarzynski, S., Li, X.-M., Innominato, P., Lévi, F., & Finkenstädt, B. (2018). Hidden markov models for monitoring circadian rhythmicity in telemetric activity data. *Journal of The Royal Society Interface*, *15*(139), 20170885.

Jakobsen, P., Ceja, E. G., Stabell, L. A., Oedegaard, K. J., Berle, J. O., Thambawita, V., Hicks, S., Halvorsen, P., Riegler, M., & Fasmer, O. B. (2020). Psykose: A motor activity database of patients with schizophrenia.

Jakobsen, P., Garcia-Ceja, E., Riegler, M., Stabell, L. A., Nordgreen, T., Torresen, J., Fasmer, O. B., & Oedegaard, K. J. (2020). Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. *Plos one*, *15*(8), e0231995.

Keselj, V. (2009). Speech and language processing daniel jurafsky and james h. martin (stanford university and university of colorado at boulder) pearson prentice hall, 2009, xxxi+ 988 pp; hardbound, isbn 978-0-13-187321-6, 115.00. MIT Press.

Kume, Y., Sugita, T., Oga, K., Kagami, K., & Igarashi, H. (2015). A pilot study: Comparative research of social functioning, circadian rhythm parameters, and cognitive function among institutional inpatients, and outpatients with chronic schizophrenia and healthy elderly people. *International psychogeriatrics*, *27*(1), 135.

Last, G., & Penrose, M. (2017). *Lectures on the poisson process* (Vol. 7). Cambridge University Press.

Lauerma, H., Niskanen, L., Lehtinen, I., & Holmstroem, R. (1994). Abnormal lateralization of motor activity during sleep in schizophrenia. *Schizophrenia research*, *14*(1), 65–71.

Li, X., Zhang, Y., Jiang, F., & Zhao, H. (2020). A novel machine learning unsupervised algorithm for sleep/wake identification using actigraphy. *Chronobiology International*, 1–14.

Lindamer, L. A., McKibbin, C., Norman, G. J., Jordan, L., Harrison, K., Abeyesinhe, S., & Patrick, K. (2008). Assessment of physical activity in middle-aged and older adults with schizophrenia. *Schizophrenia research*, *104*(1-3), 294–301.

Liu, J., Zhao, Y., Lai, B., & Tsui, K. L. (2019). A two-stage hmm model for sleep/wake identification via commercial wearable device, In *2019 ieee international conference on systems, man and cybernetics (smc)*. IEEE.

Martin, J. L., Jeste, D. V., & Ancoli-Israel, S. (2005). Older schizophrenia patients have more disrupted sleep and circadian rhythms than age-matched comparison subjects. *Journal of psychiatric research*, *39*(3), 251–259.

Martin, J., Marler, M., Shochat, T., & Ancoli-Israel, S. (2000). Circadian rhythms of agitation in institutionalized patients with alzheimer's disease. *Chronobiology international*, *17*(3), 405–418.

McFadden, D. (1974). Frontiers in econometrics, chapter conditional logit analysis of qualitative choice behavior. Academic Press.

Meyer, N., Faulkner, S. M., McCutcheon, R. A., Pillinger, T., Dijk, D.-J., & MacCabe, J. H. (2020). Sleep and circadian rhythm disturbance in remitted schizophrenia and bipolar disorder: A systematic review and meta-analysis. *Schizophrenia Bulletin*.

Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., Brown, T., Chesson Jr, A., Coleman, J., Lee-Chiong, T., Et al. (2007). Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: An update for 2007. *Sleep*, *30*(4), 519–529.

Naitoh, P., Englund, C. E., & Ryman, D. H. (1985). Circadian rhythms determined by cosine curve fitting: Analysis of continuous work and sleep-loss data. *Behavior Research Methods, Instruments, & Computers*, *17*(6), 630–641.

of Health, U. D., & Service, H. (n.d.). *What makes us sleep?* https://www.nichd.nih.gov/health/topics/sleep/conditioninfo/causes (accessed: 14.12.2020)

Palmer, B. A., Pankratz, V. S., & Bostwick, J. M. (2005). The lifetime risk of suicide in schizophrenia: A reexamination. *Archives of general psychiatry*, *62*(3), 247–253.

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

Phillips, A. J., Chen, P., & Robinson, P. (2010). Probing the mechanisms of chronotype using quantitative modeling. *Journal of Biological Rhythms*, *25*(3), 217–227.

Phillips, A. J., Czeisler, C. A., & Klerman, E. B. (2011). Revisiting spontaneous internal desynchrony using a quantitative model of sleep physiology. *Journal of biological rhythms*, *26*(5), 441–453.

Povinelli, R. J., Johnson, M. T., Lindgren, A. C., Roberts, F. M., & Ye, J. (2006). Statistical models of reconstructed phase spaces for signal classification. *IEEE Transactions on Signal processing*, *54*(6), 2178–2186.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Robillard, R., Hermens, D. F., Naismith, S. L., White, D., Rogers, N. L., Ip, T. K., Mullin, S. J., Alvares, G. A., Guastella, A. J., Smith, K. L., Et al. (2015). Ambulatory sleep-wake patterns and variability in young people with emerging mental disorders. *Journal of psychiatry & neuroscience: JPN*, *40*(1), 28.

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. *Sleep medicine reviews*, *15*(4), 259–267.

Sano, W., Nakamura, T., Yoshiuchi, K., Kitajima, T., Tsuchiya, A., Esaki, Y., Yamamoto, Y., & Iwata, N. (2012). Enhanced persistency of resting and active periods of locomotor activity in schizophrenia. *PloS one*, *7*(8), e43539.

Saß, H. (2007). *Diagnostisches und statistisches manual psychischer störungen*.

Scott, J., Vaaler, A. E., Fasmer, O. B., Morken, G., & Krane-Gartiser, K. (2017). A pilot study to determine whether combinations of objectively measured activity parameters can be used to differentiate between mixed states, mania, and bipolar depression. *International journal of bipolar disorders*, *5*(1), 5.

Sergey Lebedev, C. F., Daniela Huppenkothen. (n.d.). *Hmmlearn 0.2.2*. https://github.com/hmmlearn/hmmlearn (accessed: 21.10.2020)

Shumway, R. H., & Stoffer, D. S. (2000). Time series analysis and its applications. *Studies In Informatics And Control*, *9*(4), 375–376.
Skeldon, A. C., Dijk, D.-J., & Derks, G. (2014). Mathematical models for sleep-wake dynamics: Comparison of the two-process model and a mutual inhibition neuronal model. *PloS one*, *9*(8), e103877.
Tahmasian, M., Khazaie, H., Golshani, S., & Avis, K. T. (2013). Clinical application of actigraphy in psychotic disorders: A systematic review. *Current psychiatry reports*, *15*(6), 359.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2
Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, *13*(2), 260–269.
Von Neumann, J., Kent, R., Bellinson, H., & Hart, B. t. (1941). The mean square successive difference. *The Annals of Mathematical Statistics*, *12*(2), 153–162.
Witowski, V. (2018). Hmmpa-package: Analysing accelerometer data using hidden markov models.
Witowski, V., Foraita, R., Pitsiladis, Y., Pigeot, I., & Wirsik, N. (2014). Using hidden markov models to improve quantifying physical activity in accelerometer data–a simulation study. *PloS one*, *9*(12), e114089.
Witting, W., Kwa, I., Eikelenboom, P., Mirmiran, M., & Swaab, D. F. (1990). Alterations in the circadian rest-activity rhythm in aging and alzheimer's disease. *Biological psychiatry*, *27*(6), 563–572.
Wulff, K., Gatti, S., Wettstein, J. G., & Foster, R. G. (2010). Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nature Reviews Neuroscience*, *11*(8), 589–599.
Xu, Z., Laber, E. B., & Staicu, A.-M. (2020). Hierarchical continuous time hidden markov model, with application in zero-inflated accelerometer data, In *Statistical modeling in biomedical research*. Springer.
Zuurbier, L. A., Luik, A. I., Hofman, A., Franco, O. H., Van Someren, E. J., & Tiemeier, H. (2015). Fragmentation and stability of circadian activity

rhythms predict mortality: The rotterdam study. *American journal of epidemiology*, *181*(1), 54–63.

Appendix

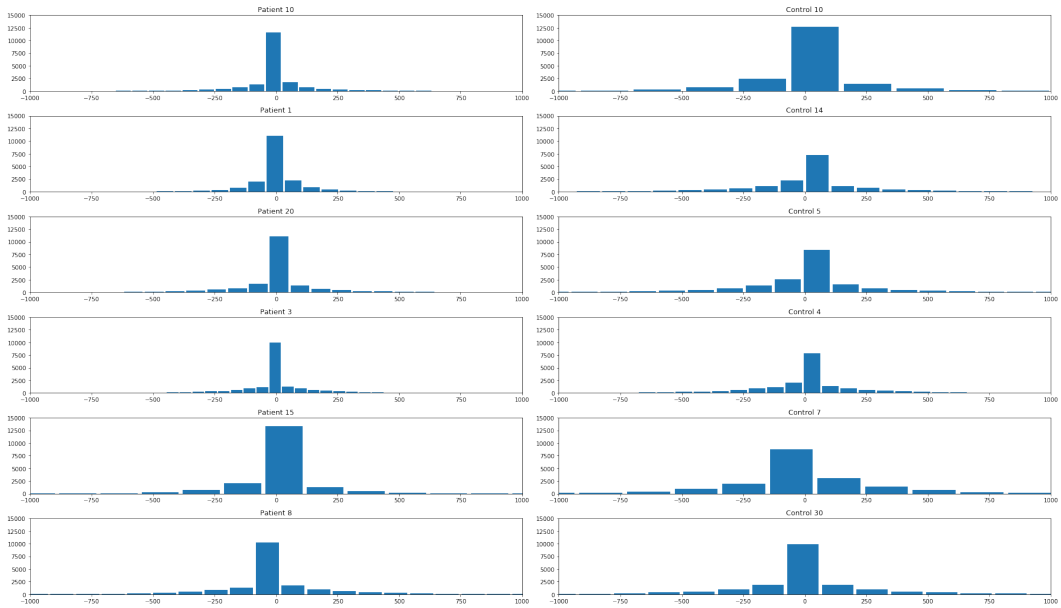


Figure 30: Example histograms of the differenced time series. It can be observed that the non-negativity is overcome by differenciating the activity time series.

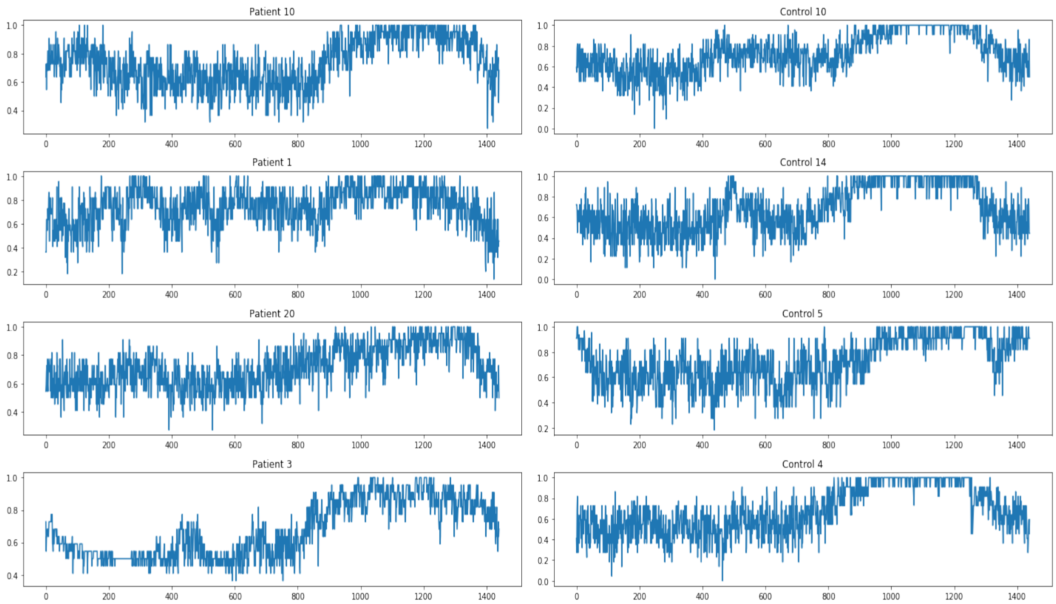


Figure 31: The average estimated transition probability to stay in state resting during 24h. The transition probability is visualized for 8 example subjects, where the patient group is located on the left and the control group on the right. Source: own illustration

Index	KPSS	ADF	Index	KPSS	ADF
1	0.100000	2.903650e-19	33	0.100000	4.388916e-28
2	0.010000	3.870852e-24	34	0.010000	3.486300e-21
3	0.100000	3.873316e-13	35	0.068342	4.865477e-19
4	0.100000	2.043466e-14	36	0.100000	5.044018e-24
5	0.100000	5.667646e-15	37	0.100000	5.077496e-21
6	0.100000	2.564178e-19	38	0.100000	8.350532e-22
7	0.010000	4.201741e-25	39	0.100000	2.937968e-20
8	0.100000	6.147830e-22	40	0.100000	6.020969e-18
9	0.100000	1.270108e-19	41	0.100000	2.645864e-23
10	0.010000	6.238952e-25	42	0.100000	1.558507e-19
11	0.100000	1.033441e-18	43	0.100000	9.014214e-16
12	0.100000	1.793844e-20	44	0.100000	7.098700e-23
13	0.100000	3.182136e-12	45	0.100000	3.063391e-26
14	0.100000	8.492540e-21	46	0.015913	1.275977e-21
15	0.100000	3.260825e-14	47	0.077687	9.274895e-25
16	0.100000	4.683741e-17	48	0.010000	3.059349e-19
17	0.100000	1.842905e-19	49	0.085235	1.314988e-24
18	0.049940	1.319204e-19	50	0.100000	3.861816e-30
19	0.010000	7.101954e-18	51	0.100000	7.270375e-24
20	0.100000	1.025360e-23	52	0.084323	1.108467e-24
21	0.100000	5.847979e-22	53	0.100000	1.486257e-27
22	0.100000	2.898210e-17	54	0.100000	1.214645e-28
23	0.100000	4.038317e-19			
24	0.040539	4.709262e-21			
25	0.100000	5.595913e-20			
26	0.100000	2.086683e-27			
27	0.100000	5.725446e-20			
28	0.100000	9.651874e-24			
29	0.100000	1.882582e-29			
30	0.077913	4.086290e-26			
31	0.031744	8.122117e-23			
32	0.080492	2.401377e-14			

Table 9: Table of p-Values of KPSS and ADF test for each time series calculated.

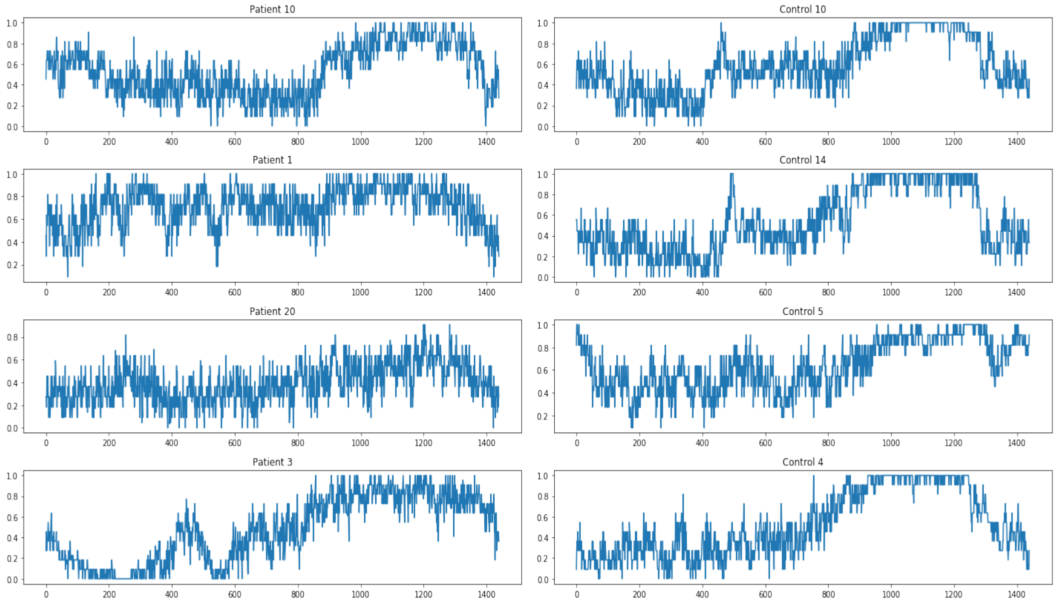


Figure 32: The average estimated transition probability from state active to state resting during 24h. The transition probability is visualized for 8 example subjects, where the patient group is located on the left and the control group on the right. Source: own illustration

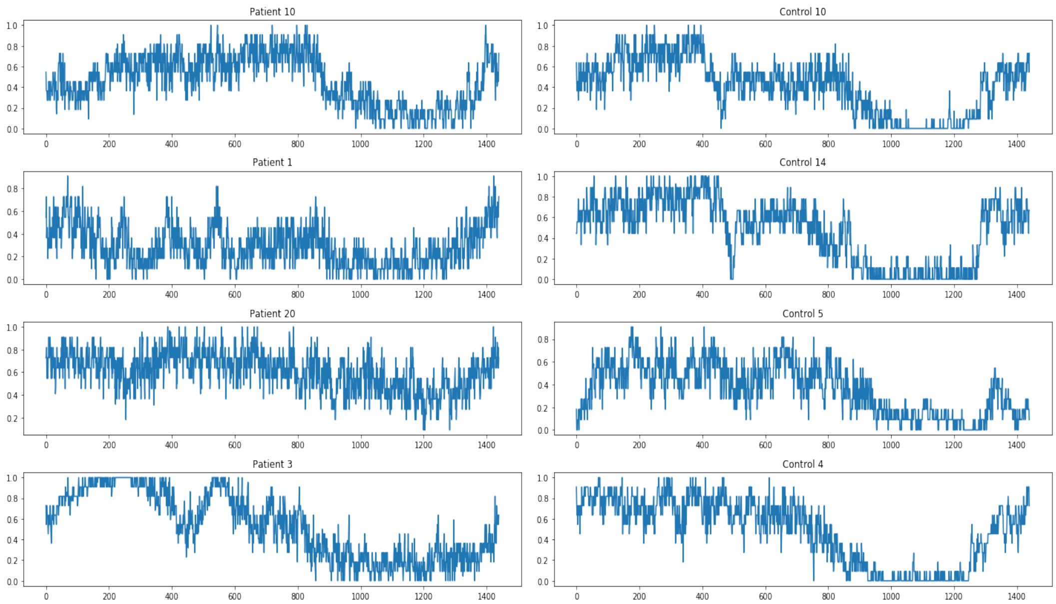


Figure 33: The average estimated transition probability to stay in state active during 24h. The transition probability is visualized for 8 example subjects, where the patient group is located on the left and the control group on the right. Source: own illustration