Prediction of Schizophrenia from Activity Data using hidden Markov Model Parameters

Matthias Boeker Hugo L. Hammer Michael A. Riegler
Pål Halvorsen Petter Jakobsen

September 27, 2022

Abstract

In this paper, we address the problem of predicting schizophrenia based on a persons measured motor activity over time. A key challenge to achieve this is how to extract features from the activity data that can efficiently separate schizophrenia patients from healthy subjects. To achieve this, we suggest to fit time dependent hidden Markov models with and without integrated covariates and letting the estimated model parameters represent our features. To further evaluate the efficiency of these features, we suggest to use them as features in a classification method (logistic regression) to separate schizophrenia patients from healthy subjects. The results show that the estimated hidden Markov model parameters are well-performing in predicting schizophrenia, and outperform features derived from other methods in the literature in terms of goodness-of-fit and classification performance.

Keywords— schizophrenia circadian rhythm motor activity time series classification hidden Markov model logistic regression

1 Introduction

Mental health is a fundamental component of health and essential to people's collective and individual ability to think, feel, and interact with one another's. The World Health Organization defines mental health as "a state of well-being in which an individual realises his or her abilities, can cope with the normal stresses of life, can work productively and can make a contribution to his or her community" [1].

Schizophrenia is a severe mental illness characterised by temporary, fundamental disturbances of thinking, perception and experience with impairments up to the loss of the reference to reality [2]. Affected persons exhibit patterns of disturbances in the areas of attention, perception, thinking, ego function, drive, affectivity and psychomotor function [3]. The symptoms of schizophrenia can make it difficult to participate in everyday activities [4].

The diagnostic practice of schizophrenia largely depends on subjective tools, like self-reports, clinical assessments and observations [5]. However, sensor data collected from motor activity recordings could be an objective and reliable method with a huge potential to either relieve or support existing subjectively diagnostic methods [6]. Motor activity is commonly recorded with wrist-worn piezoelectric accelerometers, measuring movements in the three-dimensional space [7].

Sleep disturbances and related disruptions of the circadian rhythm is a common symptom of psychiatric disorders [8, 9]. The circadian rhythm synchronizes humans to the diurnal rhythm of day and night, and cues a complex system of recurring interlocked biological rhythms, like the sleep-wake rhythm, shorter rest-activity patterns, regulation of hormone levels, as well as numerous other internal processes [10]. In schizophrenia, sleep disturbance is a common symptom of a disturbed and unsynchronized circadian system [11]. In motor activity, both diurnal fluctuations of the circadian system and social rhythm patterns are recognized [12].

In this paper we address the problem of using motor activity records of schizophrenic individuals and a control group to predict whether a subject has been diagnosed with schizophrenia. A core challenge to obtain this is on how to extract useful features from the activity data that are able to efficiently separate schizophrenia patients from healthy controls, and have been addressed in several studies [13, 14, 15, 16, 17, 18].

An interesting family of methods to model motor activity when the activity state of a subject is unobserved are hidden Markov models (HMMs). HMMs have been applied in several studies [19, 20]. Huang et al. suggested the use of time dependent HMMs with integrated covariates [21]. Carr et al. applied time-dependent HMM with integrated covariates on activity time series to predict the depression score of bipolar patients [22].

Extracting features from activity data that are able to efficiently separate schizophrenia patients from healthy subjects is the core challenge addressed in this paper. We suggest a two-step machine learning procedure. In the first step, the circadian rhythm is modelled by two states, the resting state and the active state, by using two time dependent HMMs with and without incorporated covariates similar to the models in Huang et al. and Carr et al. In the second step, the estimated parameters of the HMMs are used as features in a classification method (logistic regression) to predict schizophrenia. The extracted HMM features' ability to predict schizophrenia is further compared to other sets of features suggested in the literature.

The modelling of the activity data in this paper is closely related to the ones in Huang et al. and Carr et al. However, the novelty in this paper, lies in the second step described above, using the estimated parameters as features in a machine learning classification method. Time dependent HMM observes the transition probabilities of the HMM over time. The fluctuations in the time varying transition probabilities are assumed to give valuable information about the difference between schizophrenic and non-schizophrenic persons. The second HMM with incorporated covariates allows the integration of a covariate into the time depending transition probabilities of an HMM.

This paper is an extension of our previously published conference paper

[23]. The conference paper applied a standard HMM, while in this paper we develop two time dependent HMMs. The model parameters from the time dependent HMM achieved better prediction performance than the standard HMM. The time dependent HMM with integrated covariate structure did not perform as good as the time dependent HMM model, but we found its parameters to be informative in explaining the difference between schizophrenic and non-schizophrenic.

2 Related Work

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. 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 [24]. Actigraphy can be useful in detecting sleep patterns, sleep disorders or neurobehavioral disorders [25, 26]. However, although there is a certain lack of sufficient studies on the relationship between circadian rhythms and schizophrenia in motor activity data [27], it has been hypothesized that circadian disturbance is an essential feature of schizophrenia [11]. Unlike in this work, 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. [28] examined the difference between patients with schizophrenia and a control group based on sleep patterns. They analysed the motor activity, sleep and life quality of 68 subjects. Apiquian et al. [29] investigated on the impact of medical treatment on the pattern of motor activity of 20 schizophrenic patients against 20 healthy controls. Docx et al. [30] study the relationship between qualitative psychomotor performance levels and motor activity in 49 schizophrenic patients. Robillard et al. [31] present a bigger study on motor activity of 342 subjects with various mental disorders, like anxiety disorder bipolar, disorder, unipolar depression and psychotic disorders. Kume et al. [32] analyse the patterns of the circadian rhythm and cognition functions of 35 schizophrenic patients.

Previous studies have successfully been able to discriminate between depressed patients and controls when applying various machine learning techniques [33]. Studies comparing the motor activity of schizophrenic patients to mood disorder patients and healthy controls utilizing nonlinear mathematical models have identified schizophrenia as a distinctive subtype in motor activity. Schizophrenia is characterized by complexity and irregularity in activity patterns [34, 35]. The activity patterns illustrate a distinct profile regarding the distribution of active and inactive periods [36]. In addition, a recent systematic review found schizophrenic patients to be associated with reduced mean motor activity, irregular activity patterns and reduced quality of sleep [37].

However, there is research on the qualitative analysis of the circadian rhythm only. Witting [38] introduced the intraday variance (IV) and the interday stability (IS) as characteristics of the wake/sleep patterns and to assess possible alterations. Witting examined the motor activity of 12 Alzheimer's disease patients and 19 controls. These characteristics have been used to determine the prevalence of schizophrenia by [39]. Witting also introduced the measure of the most active 10 hours (M10) and the least active 5 hours (L5). These measures are applied by [40] in a study analyzing sleep/wake patterns of schizophrenic patients. Similar approaches are applied when studying circadian function in bipolar disorder, a topic substantially more studied than schizophrenia [41]. A more complex approach has been introduced by Martin [42] which fits cosine functions to the recorded activity of 28 older schizophrenic and 28 control participants. The parameters of the cosine function give insight into the mental state of a person. To gain a better understanding of the sleep/wake pattern, HMMs can be used to extract these patterns. HMMs have already been applied for this purpose [43, 23, 44].

Boeker et al. [23] applied HMMs to describe the sleep/wake pattern as two states. The HMM parameters are extracted and used as features for logistic regression. Boeker et al. shows that the transition probabilities of an HMM, as well as the mean and variance of a state, can be seen as competitive characteristics to differentiate between a schizophrenia patients and healthy subjects.

3 Methodology

This work proposes to use estimated parameter values of HMM models as features in a classification method (logistic regression). First, we apply HMMs to model the motor activity patterns of individuals. A general introduction to HMMs is given in Section 3.1. In this study, two time dependent HMMs with and without integrated covariates are fitted to the motor activity data.

Secondly, the estimated model parameters of the HMMs are used as features in a logistic regression classification method to identify schizophrenic patients from healthy controls. The parameters are chosen to reflect a possible disruption of the measured activity from the circadian rhythm. This disruption can be interpreted as an indicator of schizophrenia. Further details are given in Section 3.5. The suggested HMM features are compared with features suggested in the literature in terms of classification performance.

3.1 Standard HMM

An HMM describes a random process which fulfils the Markov property. The Markov property states that the future is independent of the past, given the present [45]. Let Z 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 [46]. If the process is in state i at time n, it is noted as $z_n = i$. 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 $n+1$, z_{n+1} given only the previous state $z_n = i$. The probabilistic dependence on the past states is only connected to the future through the present state, i.e

$$P(z_{n+1} = j | z_n = i) = P(z_{n+1} = j | z_n = i, z_{n-1} = i_{n-1}, ..., z_0 = i_0) \ \forall \, i, j \in S \tag{1}$$

The HMM extends the Markov process by differentiating between the observed stochastic process and a latent stochastic process. The latent or hidden stochastic process fulfils the Markov property [45] and can only be deduced from the observable process. For observable random variable, x_n, is derived from a latent variable z_n. The standard HMM model is given as

$$P(x_1, ..., x_n, z_1, ..., z_N) = P(z_1) \prod_{n=2}^{N} P(z_n | z_{n-1}) \prod_{n=1}^{N} P(x_n | z_n) \tag{2}$$

Especially, we see that the observations are conditionally independent given the the latent Markov process.

The initial state has a special status since it does not have a previous state. The initial state probability is given by π_i. The conditional probability $P(X|Z)$ is called the emission probability. The emission probability is a probability distribution for each time step on X given the state Z, $b_i(n) = P(X | z_n = i)$ where the emission probabilities represent the probability that the observation at time t was generated by state i. In the case of a Gaussian HMM, the emission probability is Gaussian distributed. Therefore, the emission probability can be described by the first two moments of the Gaussian distribution, mean and variance. The factors $P(z_n | z_{n-1})$ represent the Markov chain and are the probabilities 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 $A \in \mathbb{R}_{kxk}$. The equation is given by

$$a_{ij} = P(z_n = i | z_{n-1} = j) \tag{3}$$

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

3.2 Time dependent HMM

Time dependent HMM refers to a family of HMMs where the transition probabilities vary with time. This can for example be useful to model the circadian rhythm. It is reasonable to assume that the probability to switch from a resting state to an active state would increase in the morning hours. The probability to switch into a resting state from an active state would increase during the evening. The probability to stay in a given state, resting or active would stay high during night or day, accordingly. The time dependent model is a Gaussian

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

One approach to model the time dependency is to let a covariate be integrated into the time dependent transition probabilities. For the problem considered in this paper, we suggest using trigonometric functions with a daily period which makes sense from our visualizations of the data in Figure 1. We propose to model the disturbance by considering the deviation between a cosine function and the course of the transition probabilities. The covariate will be integrated by link coefficients. We assume that these link coefficients are suitable to represent the deviation between cosine function and transition probabilities. In Section 5, we present an interpretation of the link coefficients.

The time dependent transition probability matrix is denoted as

$$A(n) = \begin{bmatrix} a_{00}(n) & a_{01}(n) \\ a_{10}(n) & a_{11}(n) \end{bmatrix} \tag{4}$$

Other than in Equation 3 where the matrix is stationary, each element of the matrix states the probability to transit from state i to state j at time n.

3.3 Covariate Model

The time dependent HMM with incorporated covariate includes one further model extension. The idea of incorporating a covariate follows the approach of Banachewicz et al. [47]. The approach has been applied to motor activity time series in [22] and [21]. We follow these approaches of and propose to use the parameters as classification features and present an interpretation of those in this special case. The covariate is incorporated into the transition probabilities through a link function. The multinomial logistic link function is described as

$$a_{ij}(n) = p(Z_n = i | Z_{n-1} = j, C_n) = \frac{\exp(\phi^0_{ij} + \phi^1_{ij} C_n)}{\sum_{j=1}^{m} \exp(\phi^0_{ij} + \phi^1_{ij} C_n)} \tag{5}$$

$a_{ij}(n)$ denotes the time dependent transition probability. C_n is the covariate that is supposed to be linked to the transition probabilities. The coefficients ϕ^0_{ij} and ϕ^1_{ij} are vectors of link coefficients with which the covariate will be linked to the transition probabilities. We apply a binomial logistic link function because we model two hidden states - rest or active. For the case of two hidden states $m = 2$, Equation 5 can be simplified to the binomial form, where

$$a_{i0}(n) = \frac{\exp(\phi^0_{00} + \phi^1_{00} C_n)}{1 + \exp(\phi^0_{00} + \phi^1_{00} C_n)} \tag{6}$$

and

$$a_{i1}(n) = \frac{\exp(\phi^0_{10} + \phi^1_{10} C_n)}{1 + \exp(\phi^0_{10} + \phi^1_{10} C_n)} \tag{7}$$

The covariate is a trigonometric function which approximates the course of a healthy circadian cycle [21, 22, 48]. Based on the properties of the motor activity data, in this paper, we suggest to use the sum of a sine and a cosine function. The period of the function is fitted to one day, which is 1,440 minutes long.

$$C_n = \phi^0 + \phi^1 \left[\sin\left(\frac{M}{60 \cdot 24} 2\pi x_n\right) + \cos\left(\frac{M}{60 \cdot 24} 2\pi x_n\right) \right] \tag{8}$$

The equation contains the linkage coefficients ϕ. ϕ^0 is the offset of the function, while ϕ^1 is the amplitude. The link coefficients are supposed to quantify the disruption of the circadian rhythm. Our hypothesis states that the link coefficients are larger the more the actual course of the transition probabilities deviates from the assumed healthy course. The magnitude of the linking coefficients indicates a systematic or non-systemic rhythm. Patients with schizophrenia have a less systematic circadian rhythm. In conclusion, schizophrenic persons would have in magnitude greater link coefficients. The uncertainty among subjects is quantified by the variance of the transition probabilities. The less systematic circadian rhythm of schizophrenics, the larger the variance in the observational layer is expected to be.

3.4 Baum-Welch Algorithm

The Baum-Welch algorithm approximately derives the maximum likelihood estimate of the model parameters of the HMM, and follows the structure and idea of the Expectation-Maximization algorithm (EM) [49]. The EM is applied, when the maximum likelihood estimate cannot be derived analytically. This is the case, when the data is incomplete. The case of incomplete data is given when the model includes unobserved variables, such as the latent layer of the HMMs. The EM algorithm works in two steps. First, the probability for a set of parameters θ given an observed sequence X is derived. Secondly, this probability is maximized for the parameter set θ. The two steps are called the expectation and maximization steps. The algorithm is not guaranteed to converge, but proven efficient in practice.

The integration of the covariate is possible by extending the Baum-Welch algorithm. Let $p(\mathbf{X}, \mathbf{z}|\theta)$ represent the joint probability of the observed sequence and a specific state sequence, given the model parameters θ. The EM algorithm aims to maximise its log transformed expectation with regards to the model parameters. This joint probability can be written in terms of the model parameters

$$p(\mathbf{X}, \mathbf{z}|\theta) = \pi_{z_0} \prod_{n=1}^{N} a_{z_{n-1} z_n} b_{z_n}(X_n) \tag{9}$$

As the log transformation of equation (9) is a sum of the parameter terms, each term of the function can be maximized independently. The initial state probability π_i and the emission probabilities b_i can be optimized with the conventional algorithm. The transition probabilities however are optimized separately.

Equation (10) is maximised to determine the optimal transition probabilities then given the linking coefficients of the covariate [21]. The vectors of the link coefficients are numerically optimized in an additional step of the Baum-Welch algorithm with the Newton–Raphson method. Since the number of link coefficients increases with the number of covariates integrated, only one covariate is implemented.

$$\sum_{n=1}^{N-1}\sum_{i=0}^{m}\sum_{j=0}^{m}\xi_{ij}(n)\log p(Z_{n+1}=i|Z_n=j,C_n,A_{ij}(n)) \tag{10}$$

3.5 Classification

In this article we suggest to use the estimated features from the models described above as features in machine learning methods to distinguish patients from controls. More specifically, we suggest to use a logistic regression classifier with $L1$ regularization.Logistic regression is a statistical model that predicts the likelihood of a binary dependent variable. It describes the linear relationship between the binary dependent variable and various nominal, ordinal, or real independent variables. Regularization is included in the classifier, since this in general improves classification by avoiding overfitting. The suggested $L1$ regularization results in a sparse and interpretable model. Especially interpretability is important in this work, as we derive an interpretation of the link coefficients from the regression coefficients. The $L1$ regularization is scaled by the regularization rate λ. λ is the rate to which the regularization penalises the complexity of the model. $L1$ regularization shrinks the regression coefficient of less important features to zero and provides the best combination of features. This proved useful insight of which parameters of the HMMs that are most useful for the classification task. To achieve comparable results among different models, we applied feature selection down to three features for each model. Since the sample size of 54 patients and the control subjects is rather small, the application of regularization is recommended [50, 51].

The performance of the HMM parameters is evaluated by comparing their performance to a baseline classification. The baseline classification includes well-known features from the literature, summarized in Table 1.

For the time dependent HMM, the values of the estimated time dependent transition probabilities and their variability over time are used as classification features. This is summarized into the features - mean transition probabilities over time and the respective variance over time. The following features are considered.

(i) The mean value of the resting and active state

(ii) The variance of the resting and active state

(iii) The variance of the transition probabilities over time

(iv) The mean transition probabilities over time

For the time dependent HMM with integrated covariate, the link coefficients of the covariate are used as classification features. Moreover, it includes

Features for Baseline Classification	**Literature Source**
Mean activity & Standard Deviation	[22], [21], [52], [53], [21], [39], [54], [55], [38], [48], [33]
Interdaily Stability(IS) & Intradaily Variability (IV)	[38], [21], [56], [57]
Least active 5 hours (L5) & Most active 10 hours (M10)	[21], [56]
Root mean square successive differences (RMSSD)	[34], [55]
Autocorrelation coefficient	[55]

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

the mean and variance of the resting and active states as well as the transition probabilities. As in the set of features above, the mean of the transition probabilities over time approximates the transition probabilities. Overall, the following features of the time dependent HMM are considered for the logistic regression.

(i) The mean value of the resting and active state

(ii) The variance of the resting and active state

(iii) The mean transition probabilities over time

(iv) The link coefficients

Feature selection is applied to identify the three best performing features within each feature group. We evaluate the model performance of different regularization rates λ. First, we reduce the number of features by $L1$-regularization. The λ and the corresponding feature set that maximizes the Area under the Receiver Operating Characteristic Curve (AUC) are selected. It is then ensured that all feature sets contain only three features. Each feature in each set is analyzed individually for its predictive performance and model fit. The feature with the best performance is selected as the basis for step-wise regression. In step-wise forward selection, one additional feature is included at a time until each feature set contains three features. Classification prediction is evaluated using cross-validation.

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 [58] and is the quotient of the maximum likelihood estimator and the LL-Null model. The pseudo R^2 denotes how much of the total variation in the data can be explained by the model and standardization is applied before each model is implemented.

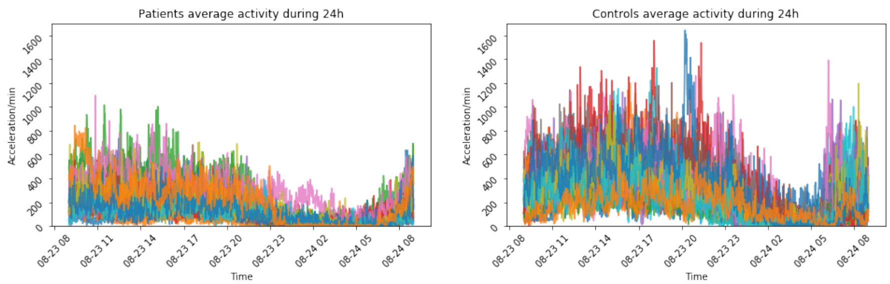

Figure 1: The figure illustrates the intensity of acceleration per minute over time. The acceleration is referred to as activity.
The average activity over 24h, starting from 9 am to 8.59 am. The left side represents the patients' average activity over 24h and the right side for the control group. The full data set is displayed.

The overall performance is assessed based on the pseudo R^2, the significance of single coefficients based on the p-values of the t-test and the Matthews Correlation Coefficient (MCC).

4 Data Description & Analysis

The data set contains actigraphy data from 22 subjects diagnosed with schizophrenia and 32 healthy controls. The overall sample size of 54 subjects is rather small whereas the sample size for each subjects is large with in average 18,320 observation points. $L1$ regularization prevents overfitting of the logistic regression model for small sample sizes [51, 50].

The activity was 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 per minute. The recording duration varied from subject to subject ranging from 9 to 20 days [54]. The data set is publicly available via [59]. We refer to the recorded acceleration as activity.

Figure 1 illustrates the average activity over 24h, from 9 am to 8.59 am the next day for patients and the controls. The average activity of the control group indicates a higher activity during the day and a lower activity during the night. Moreover, the transition from the resting to the active period appears to be more abrupt for the controls compared to the patient group. The patient group reveals a more flat course of activity throughout the day. The distinction between day and night is smaller. In general, activity is lower and the transition from resting to the active period emerges smoother than in the control group.

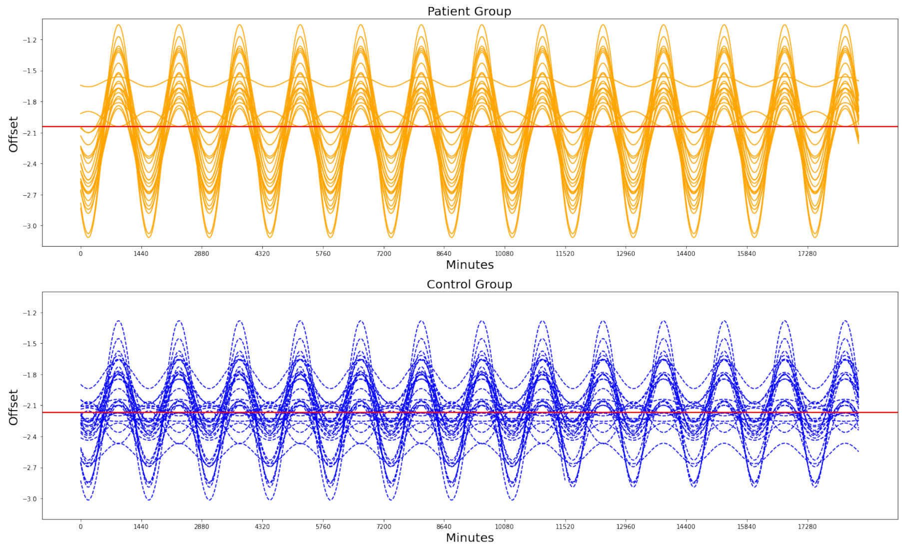

Figure 2: Visualisation of the covariate in Equation 8 for the individual optimized link coefficients. The dashed blue lines show the control group, while the orange lines represent patients. The average offset link coefficient is presented as a red line in both graphs. The optimal link coefficients are derived by maximising Equation 10.

5 Results

In Section 3, we introduced two time dependent HMMs. Figure 2 displays the integrated trigonometric functions for each subject with their respective optimal link coefficients. The blue dashed lines represent subjects from the control group, while the orange lines show schizophrenic subjects. The red line illustrates the average offset for each group. Figure 2 underlines our hypothesis from Section 3.3 that the link coefficients are larger the more the actual course of the transition probabilities deviates from the assumed healthy course. Indeed, the orange patient's covariates tend to have a larger amplitude and a higher offset than the control group.

We use the HMM parameters as classification features and evaluate their performance. The results of the classification are presented in Table 2. To get a comprehensive assessment of the regression model performance, four different metrics are applied. While the MCC and the AUC reflects the classification performance, the pseudo R^2 presents the goodness-of-fit and the p-values show the significance of single features. MCC was introduced by Matthews and is a balanced metric for classification performance. The given dataset is slightly unbalanced, thus the accuracy metrics would give us a biased picture of the performance. The MCC is calculated from the confusion matrix and ranges from $[-1, 1]$ [60]. A MCC equal to one means a perfect prediction in each class hence no false positive nor false negative. A MCC equal to zero indicates random guessing. If the MCC is equal to -1, the classifier would misclassify in any case.

Model	Features	Pseudo R^2	SE	p-value of t-test	MCC	AUC
Literatue Feature Model	Mean	0.584	0.799	0.001	0.82	0.93
	IS		0.637	0.034		
	IV		0.639	0.052		
Standard HM	Var1	0.674	1.329	0.004	0.82	0.96
	Trans01		0.829	0.008		
	Trans10		0.541	0.098		
Time dependent Model	Var1	0.740	5.861	0.006	0.91	0.97
	Trans01		3.553	0.029		
	Var01		1.408	0.020		
Model with integrated covariate	Mean1	0.716	2.019	0.005	0.83	0.98
	Trans10		1.216	0.017		
	Link inter.		1.043	0.262		

Table 2: The overall comparison of the performance statistics between the best performing feature sets of each model.

For comparison reasons, we calculated the random baseline for the classification. Random guessing achieves an average precision of 0.45 and a MCC of 0.11. Beating the random baseline can be seen as an important contribution to mental health-related applications due to the fact that diagnosing a patient with a mental health disease can be very difficult also for medical professionals. Thus, anything better than random guessing can be useful. Additionally, we added the results of the standard HMM parameters derived in the earlier study by Boeker et al. [23] to Table 2.

Feature selection identified three feature sets which contain the three best features derived from the two HMMs and the literature. Feature selection of the literature feature identified the intradaily variability, the overall mean and root mean successive differences as best features. Feature selection of the time dependent HMM parameters identified the variance of the active state, the transition probability to change from the resting state into the active state and the variance of the time dependent transition probability to change from the resting state into the active state as best features. Finally, a feature selection of the time dependent HMM with integrated covariate parameters identifies the mean value of the active state, the transition probability for the change from the active state to the resting state, and the intercept link coefficient as the three best features.

According to the MCC, the feature set of the time dependent HMM performs best in classification. The classification includes the variance of the active state, the transition probability from resting to being active, and the variability of the same transition probability over time achieving a MCC of 0.91. Features from the HMM with integrated covariate achieved a MCC of 0.83 and performed

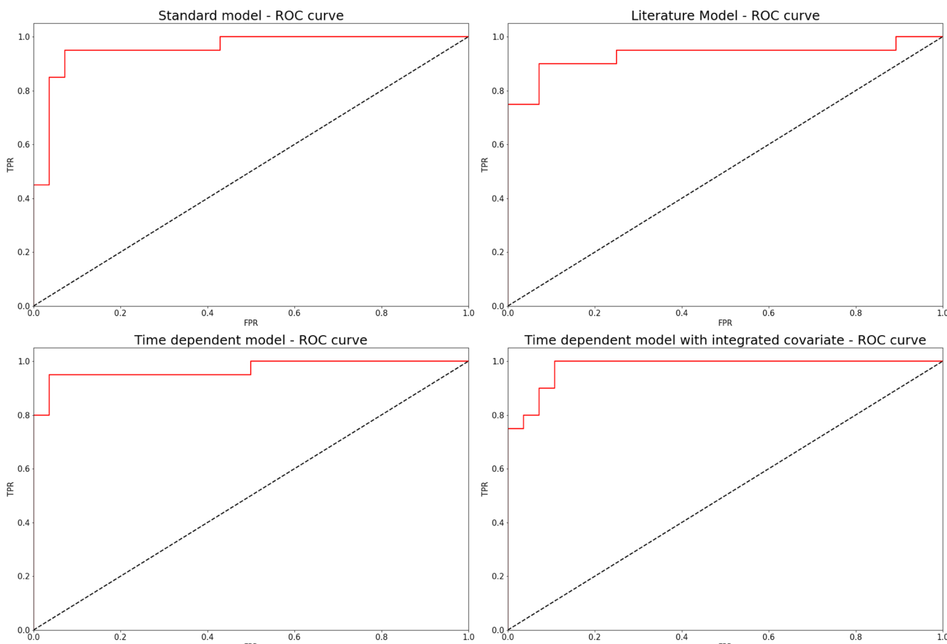

Figure 3: The figure shows the Receiver Operating Characteristic (ROC) Curve of the regression models for the four identified feature sets. The ROC curve plots the True Positive Rate against the False Positive Rate at different probability thresholds.

second-best within the comparison. The models including HMM parameters score fairly similar AUC but better then the literature feature model. Figure 3 illustrates the ROC curves of the four different regression models. The predictive performance is slightly better than the features found in the literature and the standard HMM parameters. The feature set of the time dependent HMM provides the best fit to the data with a R^2 of 0.74, followed by the features of the time dependent HMM with integrated covariate and the literature features. When analysing the p-values we differentiate between a strong significance of $p < 0.05$ and a weak significance of $p < 0.1$. All features of the time dependent model are found to be strongly statistically significant in explaining the difference between schizophrenics and the control group. For the regression model which includes the features of the HMM with integrated covariate two of the variables are significant. The mean value of the active state and the transition probability to change from active to resting are significant, whereas the first is strong and the latter weak significant is. The regression model based on the literature feature found Mean and IS to be strongly significant and IV to be weakly significant. Overall, the features of the time dependent HMM performs better than the other feature sets regarding to all the comparison metrics. Moreover, the time dependent extension of the HMM outperforms the standard HMM parameters previously presented in [23] These results only account for the given data set and the applied classification method of logistic regression.

We introduced various HMM parameters as classification features for predicting the presence of schizophrenia. The results of the logistic regression models based on different HMM parameters have been evaluated.

We further investigate the classification performance of only the link coefficients. A new classification is performed without feature selection. The results of the new logistic regression model is presented in Table 3. To better understand if the link coefficients have an impact on the classification, the regression coefficients are interpreted.

According to Equation (5) 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 the difference between high and low of the trigonometric function.

Since the logistic regression is applied, the regression coefficients are interpreted differently than in ordinary regression. If the value of the coefficient increases by one, the expected change in the log odd ratio changes by the coefficient value

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{11}$$

The coefficients in Table 3 are given as 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 means that if the vertical shift of the probability to stay active increases by one the likelihood of being schizophrenic increases by 2.84 times. The more vertical shift is needed to be integrated, the more likely it is that one is schizophrenic. 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. The bigger the magnitude of the integration coefficient to increase the amplitude, the likelier it is that the person is schizophrenic.

Hence, the link coefficients could indeed present the measurements of disruption of the circadian rhythm of schizophrenic persons. However, these results need further investigation, as the link coefficients did not significantly contribute to the classification results in Table 2.

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

6 Discussion & Conclusion

The main aim of this work has been to improve the classification of schizophrenia based on a subject's motor activity. We introduce the approach of using HMM parameters as classification features for logistic regression. A time dependent HMM and a time-dependent HMM with integrated covariate are applied to motor activity to identify a person's rest and activity phases. We compare three sets of features, features from time dependent HMM with and without integrated covariate which both are compared with features from the literature. Each classification model includes three features which have been identified by feature selection. The HMM parameters performed better than using features from the literature.

The results in Table 2 show that features of the time dependent HMM achieved best in the classification comparison. The feature set which includes the variability of the transition probability from resting achieves the best classification results. Moreover, the transition probability is significant. Hence, it is concluded that the variability of the transition probability can be used as a valid classification feature.

The features of the HMM with integrated covariate performed overall better than the baseline literature features. However, the results of the evaluation procedure did not find the link coefficients to be significant in explaining the difference between schizophrenic and non-schizophrenic people. Nevertheless, the link coefficients are further investigated. We state a potential interpretation of these features in the context of a quantified circadian rhythm.

The introduced features have only been analyzed by logistic regression with $L1$ regularization. $L1$ regularization is known for the fact that correlated variables are mutually exclusive by regulation. 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 HMM parameters, each variable has to be analyzed independently.

The results are limited to the given data set and its small sample size, the applied feature selection and the classification method used. The introduced approach needs to be further evaluated on other data sets and with other classification methods. In future work, our presented approach of using model parameters of time dependent HMM as classification variables to diagnose schizophrenia or similar mental illnesses based on the motor activity should be challenged and further explored using other data sets and classification methods.

Conflict of interest

The authors declare that they have no conflict of interest.

Data Availability

The dataset analysed during the current study is available in the Simula Datasets repository [59].

Abbreviations

S State space.

X Observable time-discrete stochastic process.

Z Time-discrete stochastic process of hidden states.

λ The regularization rate of the $L1$ regularization.

ϕ^0 Intercept link coefficient.

ϕ^1 Amplitude link coefficient.

π_i Initial probability of state i.

θ Parameter set of a hidden Markov model.

a_{ij} Transition probability to switch from state i to state j.

b_i Emission probability of state i.

AUC Area under the Receiver Operating Characteristic Curve.

HMM hidden Markov Model.

IS Interdaily Stability.

IV Intradaily Variability.

MCC Matthews Correlation Coefficient.

RMSSD Root Mean Square of Successive Differences.

References

[1] World Health Organization, *Promoting mental health: Concepts, emerging evidence, practice: Summary report* (World Health Organization, 2004)

[2] A.H. Ropper, Stephen r. marder, md, and tyrone d. cannon, ph. d., N Engl J Med **381**, 1753 (2019)

[3] N.I. of Mental Health, Schizophrenia, NIH Publication No. 21-MH-8082 (2021)

[4] W. Gaebel, W. Wölwer, Gesundheitsberichterstattung des bundes heft 50, Berlin: Rober Koch Institut (2010)

[5] A.P. Association, et al., *Publication Manual of the American Psychological Association, (2020)* (American Psychological Association, 2019)

[6] E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K.J. Oedegaard, J. Tørresen, Mental health monitoring with multimodal sensing and machine learning: A survey, Pervasive and Mobile Computing **51**, 1 (2018)

[7] P. Jakobsen, A. Stautland, M.A. Riegler, U. Cote-Allard, Z. Sepasdar, T. Nordgreen, J. Torresen, O.B. Fasmer, K.J. Oedegaard, Complexity and variability analyses of motor activity distinguish mood states in bipolar disorder, medRxiv (2021)

[8] A.G. Harvey, G. Murray, R.A. Chandler, A. Soehner, Sleep disturbance as transdiagnostic: consideration of neurobiological mechanisms, Clinical psychology review **31**(2), 225 (2011)

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

[10] C. Dibner, U. Schibler, U. Albrecht, The mammalian circadian timing system: organization and coordination of central and peripheral clocks, Annual review of physiology **72**, 517 (2010)

[11] T.C. Delorme, L.K. Srivastava, N. Cermakian, Are circadian disturbances a core pathophysiological component of schizophrenia?, Journal of Biological Rhythms **35**(4), 325 (2020)

[12] P. Henson, I. Barnett, M. Keshavan, J. Torous, Towards clinically actionable digital phenotyping targets in schizophrenia, npj Schizophrenia **6**(1), 1 (2020)

[13] Y. Motohashi, A. Maeda, H. Wakamatsu, S. Higuchi, T. Yuasa, Circadian rhythm abnormalities of wrist activity of institutionalized dependent elderly persons with dementia, The Journals of Gerontology Series A: Biological Sciences and Medical Sciences **55**(12), M740 (2000)

[14] H. Komijani, M.R. Parsaei, E. Khajeh, M.J. Golkar, H. Zarrabi, Eeg classification using recurrent adaptive neuro-fuzzy network based on time-series prediction, Neural Computing and Applications **31**(7), 2551 (2019)

[15] I. Aydin, M. Karakose, E. Akin, A new method for time series classification using multi-dimensional phase space and a statistical control chart, Neural Computing and Applications **32**(11), 7439 (2020)

[16] O. Abu Arqub, Adaptation of reproducing kernel algorithm for solving fuzzy fredholm–volterra integrodifferential equations, Neural Computing and Applications **28**(7), 1591 (2017)

[17] S. Momani, Z.S. Abo-Hammour, O.M. Alsmadi, Solution of inverse kinematics problem using genetic algorithms, Applied Mathematics & Information Sciences **10**(1), 225 (2016)

[18] Z. Abo-Hammour, O. Abu Arqub, S. Momani, N. Shawagfeh, Optimization solution of troesch's and bratu's problems of ordinary type using novel continuous genetic algorithm, Discrete Dynamics in Nature and Society **2014** (2014)

[19] Z. Xu, E.B. Laber, A.M. Staicu, Hierarchical continuous time hidden markov model, with application in zero-inflated accelerometer data, Statistical Modeling in Biomedical Research pp. 125–142 (2020)

[20] V. Witowski, Hmmpa-package: Analysing accelerometer data using hidden markov models, CRAN R Packages (2018)

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

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

[23] M. Boeker, M.A. Riegler, H.L. Hammer, P. Halvorsen, O.B. Fasmer, P. Jakobsen, in *Proceedings of IEEE International Symposium on Computer-Based Medical Systems (CBMS)* (IEEE, 2021), pp. 432–437

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

[25] A. Sadeh, The role and validity of actigraphy in sleep medicine: an update, Sleep medicine reviews **15**(4), 259 (2011)

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

[27] M. Tahmasian, H. Khazaie, S. Golshani, K.T. Avis, Clinical application of actigraphy in psychotic disorders: a systematic review, Current psychiatry reports **15**(6), 359 (2013)

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

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

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

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

[32] Y. Kume, T. Sugita, K. Oga, K. Kagami, H. Igarashi, 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 (2015)

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

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

[35] K. Krane-Gartiser, T.E. Henriksen, G. Morken, A.E. Vaaler, O.B. Fasmer, Motor activity patterns in acute schizophrenia and other psychotic disorders can be differentiated from bipolar mania and unipolar depression, Psychiatry research **270**, 418 (2018)

[36] O.B. Fasmer, E. Hauge, J.Ø. Berle, S. Dilsaver, K.J. Oedegaard, Distribution of active and resting periods in the motor activity of patients with depression and schizophrenia, Psychiatry investigation **13**(1), 112 (2016)

[37] Z.Y. Wee, S.W.L. Yong, Q.H. Chew, C. Guan, T.S. Lee, K. Sim, Actigraphy studies and clinical and biobehavioural correlates in schizophrenia: a systematic review, Journal of Neural Transmission **126**(5), 531 (2019)

[38] W. Witting, I. Kwa, P. Eikelenboom, M. Mirmiran, D.F. Swaab, Alterations in the circadian rest-activity rhythm in aging and alzheimer's disease, Biological psychiatry **27**(6), 563 (1990)

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

[40] K. Wulff, S. Gatti, J.G. Wettstein, R.G. Foster, Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease, Nature Reviews Neuroscience **11**(8), 589 (2010)

[41] G. Murray, J. Gottlieb, M.P. Hidalgo, B. Etain, P. Ritter, D.J. Skene, C. Garbazza, B. Bullock, K. Merikangas, V. Zipunnikov, et al., Measuring circadian function in bipolar disorders: Empirical and conceptual review of physiological, actigraphic, and self-report approaches, Bipolar Disorders **22**(7), 693 (2020)

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

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

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

[45] Y. Bengio, et al., Markovian models for sequential data, Neural computing surveys **2**(199), 129 (1999)

[46] P. Brémaud, *Probability Theory and Stochastic Processes* (Springer, 2020)

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

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

[49] J.A. Bilmes, et al., 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 (1998)

[50] E. Vittinghoff, C.E. McCulloch, Relaxing the rule of ten events per variable in logistic and cox regression, American journal of epidemiology **165**(6), 710 (2007)

[51] M. Pavlou, G. Ambler, S. Seaman, M. De Iorio, R.Z. Omar, Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events, Statistics in medicine **35**(7), 1159 (2016)

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

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

[54] P. Jakobsen, E. Garcia-Ceja, L.A. Stabell, K.J. Oedegaard, J.O. Berle, V. Thambawita, S.A. Hicks, P. Halvorsen, O.B. Fasmer, M.A. Riegler,

in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)* (IEEE, 2020), pp. 303–308

[55] J. Scott, A.E. Vaaler, O.B. Fasmer, G. Morken, K. Krane-Gartiser, 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 (2017)

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

[57] L.A. Zuurbier, A.I. Luik, A. Hofman, O.H. Franco, E.J. Van Someren, H. Tiemeier, Fragmentation and stability of circadian activity rhythms predict mortality: the rotterdam study, American journal of epidemiology **181**(1), 54 (2015)

[58] D. McFadden, et al., Conditional logit analysis of qualitative choice behavior (1973)

[59] SimulaMet. Simula datasets. URL `https://datasets.simula.no`

[60] B.W. Matthews, Comparison of the predicted and observed secondary structure of t4 phage lysozyme, Biochimica et Biophysica Acta (BBA)-Protein Structure **405**(2), 442 (1975)