Applying machine learning in motor activity time series of depressed bipolar and unipolar patients.

Petter Jakobsen[1,2*], Enrique Garcia-Ceja[3], Michael Riegler[4,5], Lena Antonsen Stabell[1,2], Tine Nordgreen[6,7], Jim Torresen[5], Ole Bernt Fasmer[1,2] & Ketil Joachim Oedegaard[1,2]

[1]NORMENT, Division of Psychiatry, Haukeland University Hospital, Bergen, Norway

[2]Department of Clinical Medicine, University of Bergen, Bergen, Norway

[3]SINTEF Digital, Oslo, Norway

[4]Simula Metropolitan Center for Digitalisation, Oslo, Norway

[5]Department of Informatics, University of Oslo, Norway

[6]Division of Psychiatry, Haukeland University Hospital, Bergen, Norway

[7]Department of Clinical Psychology, Faculty of Psychology, University of Bergen, Norway

* Corresponding author

E-mail: petter.jakobsen@helse-bergen.no (PJ)

ABSTRACT

Current practice of assessing mood episodes in affective disorders largely depends on subjective observations combined with semi-structured clinical rating scales. Motor activity is an objective observation of the inner physiological state expressed in behavior patterns. Alterations of motor activity are essential features of bipolar and unipolar depression. The aim was to investigate if objective measures of motor activity can aid existing diagnostic practice, by applying machine-learning techniques to analyze activity patterns in depressed patients and healthy controls. Random Forrest, Deep Neural Network and Convolutional Neural Network algorithms were used to analyze 14 days of actigraph recorded motor activity from 23 depressed patients and 32 healthy controls. Statistical features analyzed in the dataset were mean activity, standard deviation of mean activity and proportion of zero activity. Various techniques to handle data imbalance were applied, and to ensure generalizability and avoid overfitting a Leave-One-User-Out validation strategy was utilized. All outcomes reports as measures of accuracy for binary tests. A Deep Neural Network combined with random oversampling class balancing technique performed a cut above the rest with a true positive rate of 0.82 (sensitivity) and a true negative rate of 0.84 (specificity). Accuracy was 0.84 and the Matthews Correlation Coefficient 0.65. Misclassifications appear related to data overlapping among the classes, so an appropriate future approach will be to compare mood states intra-individualistic. In summary, machine-learning techniques present promising abilities in discriminating between depressed patients and healthy controls in motor activity time series.

Introduction

The current practice of assessing mood episodes in affective disorders are subjective observations combined with semi-structured clinical rating scales. Objective methods for assessing affective symptoms are desired (1). Motor activity is an objective observation of the inner physiological state expressed in behavior patterns, and alterations in activation are essential symptoms of bipolar and unipolar depression (2, 3). The depressive state is typically associated with reduced daytime motor-activity, increased variability in activity levels and less complexity in activity patterns compared to healthy controls (2). However, in some bipolar and unipolar depressed patients contradictory motor activity patterns have been observed, characterized by increased mean activity levels, reduced variability and an augmented complexity in activity patterns more similar to that observed in manic patients (4). Such depressions are commonly associated with irritability, restlessness, and aroused inner tension, in contrast to the general loss of initiative and interest characterizing psychomotor retarded depressions (5).

It has been suggested by Sabelli et al. (6) that mood disorders are diseases of energy fluctuations, and a thermodynamic model of bipolar disorder has been proposed. Simplified the model represents two energies emanating out of a mutual zero point of down-regulated motor retarded depression. The first euphoric energy represents arousal of manic symptoms like inflated self-esteem and increased goal-directed actions. The second agitated energy is associated with aroused inner tension, anxiety and restlessness. The euthymic condition oscillates within a healthy range on both energies. There is evidential support for the thermodynamic hypothesis as amplified levels of euphoric and agitated energy seems present within the manic state (7), and agitated energy seems present in approximately one out of five depressions, regardless of polarity (8).

Motor activity is indisputable an articulation of repeated daily social rhythms in interaction with cyclical biological rhythms, driven by the 24-hour circadian clock interlocked with numerous ultradian rhythmic cycles of 2 to 6 hours (9). Out of sync biological rhythmic patterns are suggested as essential symptoms of mood episodes (10). Time series of recurring biological rhythms and day-to-day life patterns are to be considered complex dynamical systems (11). Complex dynamical systems rarely categorizes by simple linear models. Therefore, mathematical tools obtained from the field of non-linear complex and chaotic systems have been the traditional method for analyzing and evaluating motor activity recordings (12-14). Machine learning (ML) techniques have displayed promising results in analyzing data of complex dynamical systems (15, 16), and MLs ability to reveal non-obvious patterns has fairly accurately classified mood state in long-term heart rate variability analysis of bipolar patients (17). Nonlinear heart rate variability analyses have similarly identified altered cardiovascular autonomic functions in manic patients (18). Accelerometer recordings are considerably more noisy than heart rate data (19). Still, motor activity time series hold prodigious potential for various ML approaches. Techniques like Random Forest (20) and neural networks (21, 22) have revealed promising abilities to handle time series of activation data.

A neural network might be understood as a mathematical model, where millions of parameters automatically fine-tunes to optimize the models' performance (23, 24). Consequently, insight into the lines of argument are virtually impossible (25). Within medical science, there is skepticism of such a black-box method generating calculations without an explanation (26). However, outcomes from analyses of essential variables of high quality ought to be considered trustworthy, at least when measures to counteract overfitting have been applied (27). The ensemble learning method of the Random Forest algorithm is more flexible and less data-sensitive than neural networks. The approach might be understood as a

woodland of decision trees, where multiple trees look at stochastic parts of the data (28), and the algorithm has been found to predict with approximate similar quality to neural networks (29). Decision trees' decisions are transparent, and lines of argument interpretable (30).

The aim was to investigate if objective biological measures can aid existing diagnostic practice, by applying machine-learning techniques to analyze motor activity patterns from depressed patients and healthy controls.

Materials and methods

Sample characteristics

This is a reanalysis of motor activity recordings originating from a study presented in previous papers (12, 13, 31). The study group consisted of 23 bipolar and unipolar outpatients and inpatients at Haukeland University Hospital, Bergen, Norway. All fulfilled the criteria for a major depression, according to a semi-structured interview based on DSM-IV criteria for mood disorders (32). The severity of the depressive symptoms was evaluated on the Montgomery and Aasberg Depression Rating Scale (MADRS) at the beginning and conclusion of the motor-activity recordings (33). Further description of the study group is presented in previous papers.

The control group consisted of 32 heathy individuals, all without a history of either psychotic or affective disorders. Both datasets are available for other researchers (34). The Norwegian Regional Medical Research Ethics Committee West approved the study protocol, a written informed consent was obtained from all participants involved in the study, and all processes were in accordance with the Helsinki Declaration of 1975.

Recording of motor activity

Motor activity was recorded with a wrist-worn actigraph entailing a piezoelectric accelerometer programmed to record the integration of intensity, amount and duration of movement in all directions. The sampling frequency was 32 Hz and movements over 0.05 g recorded. The output was gravitational acceleration units per minute (31).

Machine Learning

The basic framework of our ML approach has earlier been presented in a technological conference paper (35), but the method presented here represents a substantial extension of the previous work. Given that the main objective was to classify a user as depressed or not depressed, we proposed the following approach to accomplish this: Each user collected data for d_i consecutive days where d_i represents the number of days collected by participant i. Then, statistical features capturing overall activity levels and variations from each day were extracted (36), resulting in d_i feature vectors per participant, and then normalized to values between zero and one. The features were extracted in the statistical software R version 3.6.0.

To avoid overfitting, we adopted a Leave-One-User-Out validation strategy, i.e., for each user i use all the data from all other *users* $\neq i$ to train the classifier and test them using the data from user i. In order to obtain the final classification for a particular user, depressed or not depressed, a vector of predictions $\mathbf{p}$ is first obtained from the trained classifier. Each entry of $\mathbf{p}$ corresponds to the prediction of a particular day. The final label was obtained by majority voting, i.e., output the most frequent prediction from $\mathbf{p}$ (27).

Our dataset was imbalanced with 291 depressed and 402 not depressed states, yet it is regarded as a realistic representation of real-world clinical data (37). As ML algorithms generally have a tendency to favor the most represented class (38), we tested two different class balancing oversampling techniques for augmenting the minority class (38). Firstly, we used random oversampling, which duplicates data points selected at random. Secondly, we

used SMOTE (39), which creates new synthetic samples that are generated at random from similar neighboring points. Furthermore, we tested three different machine learning classifiers, Random Forest (40), unweighted and weighted Deep Neural Network (DNN) and a weighted Convolutional Neural Network (CNN) (41). The weighted DNN and CNN use class weights at training time to weight the loss function. The weight for the depressed class was set as $wdepressed = \alpha/\beta$ where α is the number of instances that belong to the majority class (depressed) and β is the number of points that belong to the minority class (not depressed). The weight for the not depressed class was set as $wnondepressed = \alpha/\alpha = 1$. This weighting informs the algorithm to pay more attention to the underrepresented class. For CNN, neither random oversampling nor SMOTE were utilized as the network was trained with image-like representations.

Random forest is an ensemble method that uses multiple learning models to gain better predictive results. It consists of several decision trees. Each decision tree considers a subset of features to solve the problem at hand. Each subset has only access to a subset of the training data points, consequently leading to a more robust overall performance by increasing the diversity in the forests. The subsets are chosen randomly, and the final prediction is an average from all sub decision trees within the forest (40). The code was implemented in the statistical software R with the use of the *randomForest* library (28).

The DNN architecture consisted of two fully connected hidden layers with 128 and 8 units respectively with a rectified linear unit (ReLU) as activation function. After each layer, we applied dropout (p = 0.5) and the last layer has 2 units with a softmax activation function. The CNN architecture entailed two convolutional layers where max pooling and dropout (p = 0.25) were used. Then, two more convolutional layers also applying max pooling and dropout (p = 0.25) followed. Lastly, the data was flattened, and there was a fully connected layer of 512 units with dropout (p = 0.50). Each convolutional layer had a kernel of size 3 with a stride

size of 1. The number of kernels for the first two convolutional layers was 16, and 32 for the last 2 layers. The max pooling size was 2. The activation functions of the convolutional layers and the fully connected layer were ReLUs. Finally, a fully connected layer with 2 units and softmax activation function was used to produce the prediction (41). For the CNN, instead of extracting features, we represented each day as an image with 24 rows and 60 columns. The rows represent the hour of the day, and the columns represent the minute for each particular hour. Each entry is the activity level registered by the device. Missing values were filled with -1. Both networks trained for 30 epochs with a batch size of 32. The code was written in R (version 3.6.0) using the *Keras* library with Tensorflow 1.13 as the backend. For baseline classifier, we used a classifier that outputs a random class only based on their prior probabilities regardless of the input data.

Statistics

The intention of statistical feature extraction from the raw data file is to distillate the dataset into a few variables adaptable for the machine learning algorithms, ideally capturing the essential content of the original dataset (27). As no established practice exists, a common way to find out what features to select for a given dataset is empirically evaluating different features (42). The statistical features extracted for this experiment were mean activity level, the corresponding standard deviation (SD) and the proportion of minutes with an activity level of zero. The estimates were chosen due to previous experiences in analyzing accelerometer data with ML (43). Mean values were calculated from the pre-normalized features per day for each participant, and significance tested with SPSS version 24. Independent Samples T-Test with Levene's Test of Equality of Variances were applied when comparing two groups. One-way ANOVA when comparing more than two groups, followed by Bonferroni corrections to evaluate pairwise differences between groups. A p-value less than 0.05 was considered statistically significant.

Outcome Metrics

Since our ML objective was to classify cases as either conditions or controls, the outcome of machine learning algorithms were given in measures of accuracy for binary tests (44). *Sensitivity* is the fraction of correctly classified conditions related to all conditions and *specificity* the fraction of controls correctly classified as controls. *Weighted recall* is an estimate combining sensitivity and specificity equalized according to sample sizes. The *positive* (PPV) and *negative* (NPV) *predictive values* represent the amount of correct classifications related to the amount of wrong classifications of either conditions (positive) or controls (negative). *Weighted precision* is an estimate combining the predictive values according to sample sizes. Although the estimate Accuracy is a common indicator when reporting outcomes, it does not consider imbalance in the dataset, and therefore potentially presents misrepresentative outcomes. For evaluating the overall performance of the ML classifiers, we used the *Matthews Correlation Coefficient* (MCC) that is recommended when datasets are imbalanced (45). MCC gives a coefficient value between minus one and one, and zero indicates a random estimation.

For the interpretability analysis we used the model-agnostic method Partial Dependence Plots (PDPs) to illustrate separately each of the extracted statistical features' impact on the Random Forest outcome (46). To generate the plots, the *pdp* R library was used (47). Classes were converted to numeric: depressed = 1 and control = 0, and the partial dependence of a set of features of interest *zs* was estimated by averaging the predictions for each unique value of *zs* while keeping the other variables fixed (30) .

Results

The condition group analyzed in the first ML runs consisted of 10 females and 13 males, aged 42.8 ± 11 years (mean ± standard derivation), and with average actigraph recordings of 12.7 ± 2.8 days. Mean MADRS score at the start of registrations was 22.7 ± 4.8, and at the end 20.0 ± 4.7. Fifteen persons were diagnosed with unipolar depression and eight with bipolar disorder. The control group consisted of 20 females and 12 males, average age was 38.2 ± 13, and the group wore the actigraph for an average of 12.6 ± 3.3 days (tab.1).

Table 1. Characteristics of the depressed patients and healthy controls analyzed in the first Machine Learning run.

	Depressed patients	Healthy Controls	t-test*
Label	Condition	Control	
Days[1]	291	402	
N	23	32	
Gender (male/female)	13 / 10	12 / 20	
Age	42.8 ± 11.0	38.2 ± 13.0	p = 0.170
Days used Actigraph	12.7 ± 2.8	12.6 ± 2.3	p = 0.897
Diagnosis (unipolar/bipolar)	15 / 8		
MADRS at start	22.7 ± 4.8		
MADRS at end	20.0 ± 4.7		
Extracted Statistical Features:			
Mean Activity	190.05 ± 81.44	286.59 ± 81.10	***p < 0.001***
SD[2]	300.54 ± 95.86	405.10 ± 99.7	***p < 0.001***
Proportion of Zeros[3]	0.385 ± 0.154	0.299 ± 0.086	***p = 0.010***

All data are given as mean ± standard derivation, if not otherwise specified.

*Independent Samples T-Test with Levene's Test of Equality of Variances, significance level $p < 0.05$

[1]Total number of days collected motor activity

[2]Standard derivation of mean activity

[3]Ratio of minutes with an activity level of zero

The best performing ML algorithm in the first run was Random Forest with SMOTE oversampling technique, correctly classifying 70 % of the depressed patients as conditions (sensitivity: 0.70), with a true negative rate of 0.75 (specificity). The overall capabilities of the algorithm to classify both depressions and controls were 0.73 (weighted recall), and overall performance was 0.44 (MCC). Class balancing techniques improved the performance of both of the unweighted ML approaches. In the DNN experiments, random oversampling performed best, with a weighted recall of 0.71 and MCC of 0.39. Random oversampling did not improve the Random Forest algorithm. Weighted DNN performed best without class balancing techniques (no oversampling), presenting a weighted recall of 0.69 and MCC of 0.37. The weighted CNN approach achieved a weighted recall of 0.69 and a MCC of 0.38 (tab. 2).

Table 2. Machine Learning classification results (1st run) in motor activity time series from depressed patients (n = 23) and healthy controls (n = 32).

Machine Learning Approach	Class Balancing Technique	Classification results by label											
		Sensitivity	Specificity	Weighted Recall	PPV	NPV	Weighted Precision	Accuracy	MCC	TP	TN	FP	FN
	Baseline	0.13	0.69	0.45	0.23	0.52	0.40	0.45	0.21	3	22	10	20
Random Forest	No oversampling	0.52	0.78	0.67	0.63	0.69	0.67	0.67	0.31	12	25	7	11
	Random oversampling	0.52	0.75	0.65	0.60	0.69	0.65	0.65	0.28	12	24	8	11
	SMOTE	**0.70**	**0.75**	**0.73**	0.67	0.77	0.73	0.73	**0.44**	16	24	8	7
	Baseline	0.04	0.69	0.42	0.09	0.5	0.33	0.42	- 0.33	1	22	10	22
Deep	No oversampling	0.43	0.84	0.67	0.67	0.68	0.67	0.67	0.31	10	27	5	13
Neural Network	Random oversampling	**0.57**	**0.81**	**0.71**	0.68	0.72	0.71	0.71	**0.39**	13	26	6	10
	SMOTE	0.57	0.78	0.69	0.65	0.71	0.69	0.69	0.36	13	25	7	10
	Baseline	0.04	0.69	0.42	0.09	0.5	0.33	0.42	- 0.33	1	22	10	22

Weighted Deep	No oversampling	**0.65**	**0.72**	**0.69**	0.63	0.74	0.69	0.69	**0.37**	15	23	9	8
Neural Network	Random oversampling	0.61	0.69	0.65	0.58	0.71	0.66	0.65	0.29	14	22	10	9
	SMOTE	0.65	0.69	0.67	0.60	0.73	0.68	0.67	0.34	15	22	10	8
Weighted Convolutional	Baseline	0.22	0.88	0.60	0.56	0.61	0.59	0.6	0.12	5	28	4	18
Neural Network	No oversampling	**0.70**	**0.69**	**0.69**	0.62	0.76	0.70	0.69	**0.38**	16	22	10	7

TP: True Positives (condition cases classified correctly as labeled).

FN: False Negatives (condition cases misclassified as control cases).

TN: True Negatives (control cases classified correctly as labeled).

FP: False Positives (controls cases misclassified as condition cases).

Sensitivity: True Positive Rate; TP / (TP + FN). Specificity: True Negative Rate; TN / (TN + FP). Weighted Recall: (Sensitivity x (TP + FN)) + (Specificity x (TN + FP)) / (TP + FN + TN + FP). PPV: Positive Predictive Value; TP / (TP + FP). NPV: Negative Predictive Value: TN / (TN + FN). Weighted Precision: (PPV x (TP + FN)) + (NPV x (TN + FP)) / (TP + FN + TN + FP). Accuracy: (TP + TN) / (TP + TN + FP + FN). MCC: Matthews Correlation Coefficient; ((TP x TN) – (FP x FN)) / $\sqrt{}$ ((TP + FP) x (TP + FN) x (TN + FP) x (TN + FN)).

The interpretability analysis of the Random Forest classifier is presented in a partial dependence plot for each analyzed feature (fig. 1). Regarding the mean activity level and standard deviation of mean activity, the overall tendency was decreasing values associated with a condition classification. The trend differs for the proportion of zero activity, where increasing percentage was associated with condition predicted as outcome. Similar overall tendencies were statistical observable between the groups (tab. 1), as the depressed patients were significantly lower in mean activity ($p < 0.001$) and SD of mean activity ($p < 0.001$), and had elevated ratios of minutes with an activity level of zero ($p = 0.010$) compared to controls.

Fig. 1. Model interpretability analysis (1st run): Partial Dependence Plots (PDPs) of the Random Forest classification. The x-axis represents the feature value whereas the y-axis is the models output value.

As an attempt to capture the quintessence of the misclassified groups, the false cases were commonly identified as misclassifications in the four previous mentioned ML algorithms; weighted CNN and DNN (no oversampling), DNN (random oversampling) and Random Forest (SMOTE). Six conditions were constantly classified falsely (FN) in all four outcomes. There were no significant differences (t-test) when comparing FN and the correctly classified depressions (TP) on MADRS scores. Fewer controls were commonly misclassified (n = 4). For that reason, the false positives group (FP) consisted of all controls misclassified in at least three out of four predicted outcomes (n = 7). When comparing all four outcome groups, FN, FP, TP and true negative controls (TN), we found ANOVA significant group differences for all the analyzed statistical features, mean activity ($p < 0.001$), SD mean activity ($p < 0.001$) and portions of zeros ($p = 0.007$) (tab. 3). There were no significant group differences (ANOVA) for age composition and the number of days the participants wore the Actigraph.

Table 3. Characteristics of classification results by predicted condition from the 1st Machine Learning run.

	Predicted groups				
	True Positives	**False Negatives**	**True Negatives**	**False Positives**	**ANOVA[1]**
Label	Condition	Condition	Control	Control	
N	17	6	25	7	
Mean Activity	156.71 ± 64.46*/ϕ	284.53 ± 37.34ϕ	313.31 ± 68.08*/^	191.16 ± 42.94^	F(3,51) = 24.21, ***p < 0.001***
SD[2]	266.91 ± 86.62*/ϕϕ	395.85 ± 40.93ϕϕ	433.63 ± 91.29*/^^	303.23 ± 51.92^^	F(3,51) = 15.44, ***p < 0.001***

Proportion of Zeros[3]	0.420 ± 0.146**	0.287 ± 0.143	0.295 ± 0.064**	0.312 ± 0.147	F(3,51) = 4.58, ***p = 0.007***

All data are given as mean ± standard derivation.

[1]One-way ANOVA, significance level $p < 0.05$

[2]Standard Derivation of mean activity

[3]Ratio of minutes with an activity level of zero

Post hoc Bonferroni tests (significance level $p < 0.05$):

*$p < 0.001$ - TP compared to TN

**$p = 0.006$ - TP compared to TN

ϕ$p < 0.001$ - FN compared to TP

ϕϕ$p = 0.011$ - FN compared to TP

^$p < 0.001$ - FP compared to TN

^^$p = 0.003$ - FP compared to TN

As shown in table 3 the TP conditions have Bonferroni significantly reduced mean activity compared to both FN conditions ($p < 0.001$) and TN controls ($p < 0.001$). The TP conditions had also Bonferroni significant decreased SD compared to FN conditions ($p = 0.011$) and TN controls ($p < 0.001$). In addition, the portions of minutes with an activity level of zero were significantly higher for TP compared to TN ($p = 0.006$). There were no Bonferroni significant differences between the misclassified depressions (FN) and the two control groups (TN + FP), but the misclassified control group (FP) had significantly reduced mean activity ($p < 0.001$) and SD ($p = 0.003$) compared to TN controls.

For the second ML runs, the six patients identified as the FN condition group of the first ML runs were omitted from the analysis. This time the condition group consisted of 7 females and 10 males, with an average age of 44.8 ± 10.6 years. Eleven were diagnosed with unipolar depression and six with bipolar disorder (tab. 4).

Table 4. Characteristics of depressed patients and healthy controls analyzed in the 2nd Machine Learning run.

	Depressed patients	Healthy Controls	t-test*
Label	Condition	Control	
Days[1]	215	402	
N	17	32	
Gender (male/female)	10 / 7	12 / 20	
Age	44.8 ± 10.6	38.2 ± 13.0	p = 0.077
Days used Actigraph	12.7 ± 3.0	12.6 ± 2.3	p = 0.913
Diagnosis (unipolar/bipolar)	11 / 6		
MADRS at start	23.3 ± 4.6		
MADRS at end	20.2 ± 4.8		
Extracted Statistical Features:			
Mean Activity	156.71 ± 64.46	486.59 ± 81.10	***p < 0.001***
SD[2]	266.91 ± 86.62.0	405.10 ± 99.87	***p < 0.001***
Proportion of Zeros[3]	0.420 ± 0.146	0.299 ± 0.086	***p = 0.001***

All data are given as mean ± standard derivation, if not otherwise specified

*Independent Samples T-Test with Levene's Test of Equality of Variances, significance level $p < 0.05$.

[1]Total number of days collected motor activity with Actigraph.

[2]Standard Derivation of mean activity

[3]Ratio of minutes with an activity level of zero

Random oversampling DNN performed unmatched with an overall weighted accuracy (MCC) of 0.65, a true positive rate of 0.82 (sensitivity), a true negative rate of 0.84 (specificity) and a weighted recall of 0.84. Weighted DNN without oversampling performed with an MCC of 0.62, a sensitivity of 0.82, a specificity of 0.81 and weighted recall of 0.82. These two deep

neural network approaches performed a cut above the rest, and the shared negative predictive values (NPV) of 0.90 indicates a limited number of depressions incorrectly classified as controls. The best performing Random Forest approach (SMOTE) achieved a MCC of 0.53 and a weighted recall of 0.78. Weighted CNN performed with a MCC of 0.32, a sensitivity of 0.82 and a specificity of 0.50. The impact of class balancing techniques on the various algorithms performances were comparable to the results established by the first ML run (tab. 5).

Table 5. Machine Learning classification results (2nd run) in motor activity time series of motor retarded depressed patients (n = 17) and healthy controls (n = 32).

Machine Learning Approach	Class Balancing Technique	Classification results by label											
		Sensitivity	Specificity	Weighted Recall	PPV	NPV	Weighted Precision	Accuracy	MCC	TP	TN	FP	FN
	Baseline	0.06	0.81	0.55	0.14	0.62	0.45	0.55	-0.17	1	26	6	16
Random Forest	No oversampling	0.53	0.84	0.73	0.64	0.77	0.73	0.73	0.39	9	27	5	8
	Random oversampling	0.59	0.81	0.73	0.63	0.79	0.73	0.73	0.41	10	26	6	7
	SMOTE	0.76	0.78	0.78	0.65	0.86	0.79	0.78	0.53	13	25	7	4
	Baseline	0.12	0.91	0.63	0.40	0.66	0.57	0.63	0.04	2	29	3	15
Deep	No oversampling	0.59	0.88	0.78	0.71	0.80	0.77	0.78	0.49	10	28	4	7
Neural Network	Random oversampling	**0.82**	**0.84**	**0.84**	0.74	**0.90**	0.84	0.84	**0.65**	14	27	**5**	**3**
	SMOTE	0.76	0.78	0.78	0.65	0.86	0.79	0.78	0.53	13	25	7	4
	Baseline	0.18	0.78	0.57	0.30	0.64	0.52	0.57	-0.05	3	25	7	14
Weighted Deep	No oversampling	**0.82**	**0.81**	**0.82**	0.70	**0.90**	0.83	0.82	**0.62**	14	26	6	3
Neural Network	Random oversampling	0.88	0.66	0.73	0.58	0.91	0.80	0.73	0.51	15	21	11	2
	SMOTE	0.82	0.66	0.71	0.56	0.88	0.77	0.71	0.46	14	21	11	3
Weighted Convolutional	Baseline	0.06	0.97	0.65	0.50	0.66	0.60	0.65	0.07	1	31	1	16
Neural Network	No oversampling	0.82	0.50	0.61	0.47	0.84	0.71	0.61	0.32	14	16	16	3

TP: True Positives (condition cases classified correctly as labeled).

FN: False Negatives (condition cases misclassified as control cases).

TN: True Negatives (control cases classified correctly as labeled).

338 FP: False Positives (controls cases misclassified as condition cases).
339 Sensitivity: True Positive Rate; TP / (TP + FN). Specificity: True Negative Rate; TN / (TN +
340 FP). Weighted Recall: (Sensitivity x (TP + FN)) + (Specificity x (TN + FP)) / (TP + FN + TN
341 + FP). PPV: Positive Predictive Value; TP / (TP + FP). NPV: Negative Predictive Value: TN
342 / (TN + FN). Weighted Precision: (PPV x (TP + FN)) + (NPV x (TN + FP)) / (TP + FN + TN
343 + FP). Accuracy: (TP + TN) / (TP + TN + FP + FN). MCC: Matthews Correlation
344 Coefficient; ((TP x TN) – (FP x FN)) / √ ((TP + FP) x (TP + FN) x (TN + FP) x (TN + FN)).

345

346 To illustrate the characteristics of the misclassified condition and control groups of the second
347 ML runs, we looked to the common misclassifications of the two superior predicting
348 algorithms. Random oversampling DNN misclassified three conditions and five controls, and
349 all were correspondingly misclassifications of weighted DNN. Also for Random Forrest with
350 SMOTE these eight misclassifications were mutual. When comparing the four outcome
351 groups (TP, FN, TN and FP) we found statistically significant ANOVA differences ($p <$
352 0.001) for all the analyzed statistical features (tab. 6). There were no significant differences
353 for either mean age or days the participants used the Actigraph between groups, and no
354 significant differences (t-test) when comparing the MADRS scores of the TP and FN
355 condition groups. The FN group consisted of one male and two females, one diagnosed with
356 bipolar depression and two with unipolar. The FP group included two males and three
357 females.

358

	Predicted groups				
	True Positives	**False Negatives**	**True Negatives**	**False Positives**	**ANOVA[1]**

Label	Condition	Condition	Control	Control	
N	14	3	27	5	
Mean Activity	136.78 ± 50.18*/ϕ	249.69 ± 33.55ϕ	308.30 ± 67.98*/^	169.35 ± 23.81^	F(3,45) = 28.61, ***p < 0.001***
SD[2]	239.59 ± 64.38*/ϕϕ	394.36 ± 59.16ϕϕ	428.12 ± 90.12*/^^	280.81 ± 40.18^^	F(3,45) = 19.54, ***p < 0.001***
Proportion of Zeros[3]	0.454 ± 0.107*/ϕϕϕ	0.262 ± 0.227ϕϕϕ	0.286 ± 0.085*	0.364 ± 0.059	F(3,45) = 9.22, ***p < 0.001***

Table 6. Characteristics of classification results by predicted condition from the second Machine Learning run.

All data are given as mean ± standard derivation.

[1]One-way ANOVA, significance level $p < 0.05$

[2]Standard Derivation of mean activity

[3]Ratio of minutes with an activity level of zero

Post hoc Bonferroni tests (significance level $p < 0.05$):

*$p < 0.001$ - TP compared to TN

ϕ$p = 0.026$ - FN compared to TP

ϕϕ$p = 0.020$ - FN compared to TP

ϕϕϕ$p = 0.027$ - FN compared to TP

^$p < 0.001$ - FP compared to TN

^^$p = 0.002$ - FP compared to TN

As shown in table 6, the TP conditions presents Bonferroni significantly reduced mean activity compared to FN conditions ($p = 0.026$) and TN controls ($p < 0.001$). The FP controls presents significantly reduced mean activity levels compared to TN controls ($p < 0.001$). The SD of mean activity is significantly reduced for TP conditions compared to FN conditions ($p = 0.020$) and TN controls ($p < 0.001$). SD is furthermore significantly reduced for FP controls compared to TN controls ($p = 0.002$). For the proportion of minutes with an activity level of zero, TP conditions presents significantly increased portions compared to FN conditions (p =

0.027) and TN controls ($p < 0.001$). There were no Bonferroni significant differences between the misclassified depressions (FN) and the two control groups (TN + FP), as well as between the misclassified controls (FP) and the two condition groups (TP + FN).

Figure 2 presents the interpretability analysis of the second runs' Random Forest classifier, illustrating the features' behavior in the algorithm and their impact on the decisions. The overall tendencies observed in the plots look virtually indistinguishable to the PDPs of the first ML run, although the trends look somewhat stronger. This is as expected, due to the identical data analyzed with a reduced condition group. Similar to the observed significant group differences presented in table 4, decreasing values of mean activity and standard deviation of mean activity, as well as increasing proportion of zero activity associates with the condition state classification. Furthermore, these trends reflect the significant differences between the four groups presented in table 6.

Fig. 2. Model interpretability analysis (2nd run): Partial Dependence Plots (PDPs) of the Random Forest classification. The x-axis represents the feature value whereas the y-axis is the models output value.

Discussion

In our quest to answer the objective of the study, we have tested three different machine learning classifiers analyzing an imbalanced dataset with a larger control group than condition group. ML´s ability to discriminate between depressed patients and healthy controls seems generally promising, as our results are substantially above both random and the baseline predictions. According to our experiment, the Deep Neural Network algorithm seems to be the preeminent ML approach to discriminate between depressed patients and healthy controls

in motor activity data. The most optimistic overall result were attained by unweighted DNN with the random oversampling technique, classifying 82 % of the depressed patients correctly and 84 % of the controls. Weighted DNN without oversampling techniques also coped with classifying 82 % of the depressions correctly, but achieved only to classify 81% of controls correctly. However regardless of the outcome, all machine learning approaches utilized in this experiment achieved better results than what was found in a previous study employing a nonlinear discriminant function statistical analysis to differentiate between mood states in 24 hours motor activity recordings of bipolar inpatients (14). This method managed to classify manic and mixed states rather accurately, but 42 % of the depressions misclassified as manic. The misclassification is explained by the fact that depressed patients appear to be a complex and varied group, as some patients presents manic-like motor activity patterns in their depressive state. Therefore, it is more appropriate to compare this study's results with the results of our experiment's first ML run. In the second ML run, the depressed patients misclassified in the first ML run were omitted from analyzes, as their motor activity patterns looked more analogous to the correctly classified controls, and seems to have contributed only as noisy confusion in the first analysis. Regarding the manic-like patterns, it is previously demonstrated that the motor activity patterns of mania look more similar to those of healthy controls than depressive patients (2).

A study by Krane-Gartiser and colleagues (4) investigated group differences between unipolar depressed inpatients with and without motor retardation in 24 hours of motor activity recordings. Depressions without motor retardation were found to have increased mean activity levels; reduced variability compared to the motor retarded patients, as well as augmented complexity in activity patterns. The left out patients of this experiment had significantly increased mean activity compared to the patients involved in the second analysis. The variability measure reported by this previous study is the coefficient of variation (CV) (4), a

ratio estimated from SD divided by the mean. Both CV and SD are estimates of variance, expressing the stability of the mean in a time series (48). Based on the numbers presented in Table 3, the excluded patient group (FN) appears to have reduced CV compared to the others patients (TP). Furthermore, previous studies of the complete current study population have identified increased daytime variability in activity levels for the depressed patients compared to the controls, as well as reduced complexity in daytime activity for the depressed patients (13). Sample entropy, the nonlinear index of complexity, was not estimated for this experiment. According to experiences within this research group, it is problematic to estimate sample entropy from time series containing longer durations of zero activity, like 24-hour sequences. All series become distinctly ordered, and the ability to differentiate between groups turn out to be significantly reduced. Consequently, sample entropy is applicable only for ultradian periods of more or less continuous activity. Nevertheless, overall it is a reasonable assumption that the left out patients of this experiment are quite similar to the group of unipolar patients without motor retardation reported on by Krane-Gartiser et al.

We have investigated a heterogeneous patient group consisting of unipolar and bipolar depressed inpatients and outpatients. No differences between depressed inpatients and outpatients have previously been identified (49). According to previous studies, unipolar and bipolar depressions seem to resemble each other in 24-hour motor activity (4, 50). On the other hand, psychomotor restlessness appears associated with bipolar depression, and the feature seems to differentiate bipolar disorder from unipolar depressions (51). In our sample, this does not seem to be the case as the ratio of unipolar and bipolar cases equally distributes in both the analyzed and left out patients groups. Also, the proportion of omitted depressions without motor retardation harmonize with existing evidence as agitated energy seems present in approximately one out of five depressions (8). Consequently, our promising ML results

453 then only derivate from comparing the activity patterns of motor retarded depressed patients
454 to controls.

455 Previous studies on accelerometer data and mood disorders have advised against applying ML
456 in smaller samples, mainly due to the risk of overfitting producing untrustworthy results (14).
457 Overfitting is a phenomenon occurring when an ML algorithm trains itself to perfect
458 predictions on a specific dataset, but then predicts poorly on new data due to systematic bias
459 incorporated in the judgement model. In our experiments, we applied the Leave-One-User-
460 Out validation strategy to minimize the risk of overfitting as recommended. In addition, our
461 findings are in line with existing knowledge. Therefore, our results may be regarded as
462 credible.

463 The most optimistic individual sensitivity and specificity results accomplished by DNN in the
464 second ML run represent most likely the ML algorithms' tendency to favor the majority class.
465 So when DNN without oversampling achieved a specificity (true negative rate) of 0.88 by
466 analyzing the unprocessed and imbalanced datasets, the sensitivity was only 0.59. The both
467 weighted and random oversampling DNN approach achieved the most optimistic true positive
468 rate (sensitivity) of 0.88, however, with a paltry specificity of 0.66. The two most optimistic
469 discriminating DNN approaches applied, either random oversampling or weighting to deal
470 with the imbalance problem. Still, the weighted precisions of 0.84 and 0.83 indicate a
471 substantial number of misclassifications, probably related to data overlapping among the
472 falsely classified classes demonstrated in table 6. Combined, these misclassified subjects
473 establish a gray area of intersecting activation patters, probably partly related to individual
474 differences within the groups, as some people are natural slackers and others are born highly
475 active. Anyhow, little is known about the control group beyond age, gender and the absence
476 of a history of either affective or psychotic disorders. Physical properties such as older age
477 and higher body mass index have previously been found to affect mean motor activity (14).

We did not find significant age differences between the outcome groups in our experiments, and there was no information on body mass index in the dataset. For gender, differences in activation have not been identified in previous studies (14). Furthermore, there was no information on external influences like seasonal time of year when the data was collected, but this has previously been considered not relevant for the evaluation of motor activity recordings (49). On the other hand, at higher latitudes with significant seasonal change in solar insolation, hours of daylight and social rhythms, one should expect a seasonal expression in motor activity (52). The current data was collected at latitude 60.4 N, a position associated with substantial seasonal change in natural light and length of day. Various psychiatric and somatic drugs may also influence motor activity patterns (14), but besides lack of medical data on the controls, the sample size is too small for such sub analyzes.

Beside the small sample size, the main limitation of the study was the comparison between depressed patients and healthy controls. Firstly, this may have introduced errors and noise into results as individualistic divergent activity levels might have affected group differences. Secondly, although intra-individualistic compartments suggests that the bipolar manic state is associated with increased mean activity levels, no compelling evidence has been found when comparing manic patients to healthy controls (2). Thirdly, euthymic individuals with bipolar disorders have been found to have generally reduced mean motor activity, prolonged sleep and reduced sleep quality compared to healthy controls (53). Therefore, a more appropriate approach would have been to compare a group of depressed patients to themselves in the euthymic state, as to identify actual alterations in motor activity related to changes in mood state (54). To our knowledge, no such dataset exists.

In conclusion, this study has illustrated the promising abilities of various machine learning algorithms to discriminate between depressed patients and healthy controls in motor activity time series. Furthermore, that the machine learning´s ways of finding hidden patterns in the

503 data correlates with existing knowledge from previous studies employing linear and nonlinear
504 statistical methods in motor activity. In our experiment, Deep Neural Network performed
505 preeminent in discriminating between conditions and controls. Finally, we have enlightened
506 the heterogeneity within depressed patients, as it is recognizable in motor activity
507 measurements.

508 Acknowledgements

509 This publication is part of the INTROducing Mental health through Adaptive Technology
510 (INTROMAT) project.

511 References

512 1. Bauer M, Andreassen OA, Geddes JR, Kessing LV, Lewitzka U, Schulze TG, et al.
513 Areas of uncertainties and unmet needs in bipolar disorders: clinical and research
514 perspectives. 2018.
515 2. Scott J, Murray G, Henry C, Morken G, Scott E, Angst J, et al. Activation in bipolar
516 disorders: A systematic review. JAMA Psychiatry. 2017;74(2):189-96.
517 3. Burton C, McKinstry B, Szentagotai Tătar A, Serrano-Blanco A, Pagliari C, Wolters
518 M. Activity monitoring in patients with depression: A systematic review. Journal of Affective
519 Disorders. 2013;145(1):21-8.
520 4. Krane-Gartiser K, Henriksen T, Vaaler AE, Fasmer OB, Morken G. Actigraphically
521 assessed activity in unipolar depression: a comparison of inpatients with and without motor
522 retardation. Journal of clinical psychiatry. 2015;76(9):1181-7.
523 5. Faedda GL, Marangoni C, Reginaldi D. Depressive mixed states: A reappraisal of
524 Koukopoulos′criteria. Journal of Affective Disorders. 2015;176:18-23.
525 6. Sabelli HC, Carlson-Sabelli L, Javaid JI. The thermodynamics of bipolarity: A
526 bifurcation model of bipolar illness and bipolar character and its psychotherapeutic
527 applications. Psychiatry. 1990;53(4):346-68.
528 7. Dilsaver SC, Chen YR, Shoaib AM, Swann AC. Phenomenology of mania: evidence
529 for distinct depressed, dysphoric, and euphoric presentations. American Journal of Psychiatry.
530 1999;156(3):426-30.
531 8. Tondo L, Vázquez GH, Pinna M, Vaccotto PA, Baldessarini RJ. Characteristics of
532 depressive and bipolar disorder patients with mixed features. Acta Psychiat Scand.
533 2018;138(3):243-52.
534 9. Bourguignon C, Storch K-F. Control of Rest:Activity by a Dopaminergic Ultradian
535 Oscillator and the Circadian Clock. 2017;8(614).
536 10. Alloy LB, Ng TH, Titone MK, Boland EM. Circadian Rhythm Dysregulation in
537 Bipolar Spectrum Disorders. Current Psychiatry Reports. 2017;19(4):21.
538 11. Glass L, Kaplan D. Time series analysis of complex dynamics in physiology and
539 medicine. Medical progress through technology. 1993;19:115-.

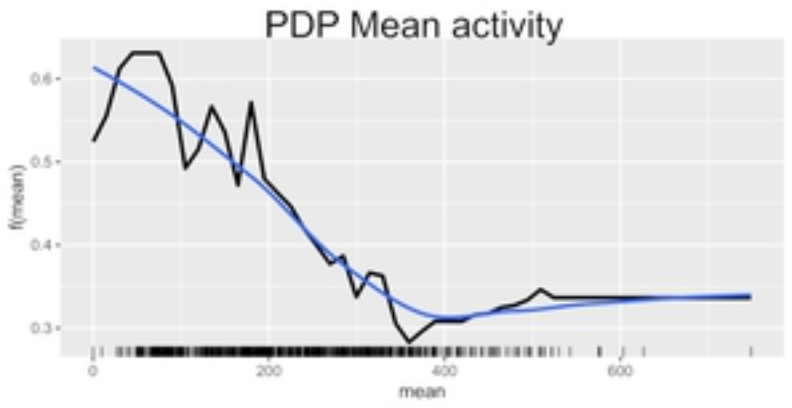
12. Fasmer EE, Fasmer OB, Berle JØ, Oedegaard KJ, Hauge ER. Graph theory applied to the analysis of motor activity in patients with schizophrenia and depression. PloS one. 2018;13(4):e0194791.
13. Hauge ER, Berle JØ, Oedegaard KJ, Holsten F, Fasmer OB. Nonlinear analysis of motor activity shows differences between schizophrenia and depression: a study using Fourier analysis and sample entropy. PloS one. 2011;6(1):e16291.
14. Scott J, Vaaler AE, Fasmer OB, Morken G, Krane-Gartiser K. 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. 2017;5(1):5.
15. Pathak J, Wikner A, Fussell R, Chandra S, Hunt BR, Girvan M, et al. Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2018;28(4):041101.
16. Wan ZY, Vlachas P, Koumoutsakos P, Sapsis T. Data-assisted reduced-order modeling of extreme events in complex dynamical systems. PloS one. 2018;13(5):e0197704.
17. Valenza G, Nardelli M, Lanatà A, Gentili C, Bertschy G, Paradiso R, et al. Wearable Monitoring for Mood Recognition in Bipolar Disorder Based on History-Dependent Long-Term Heart Rate Variability Analysis. IEEE Journal of Biomedical and Health Informatics. 2014;18(5):1625-35.
18. Chang H-A, Chang C-C, Tzeng N-S, Kuo TBJ, Lu R-B, Huang S-Y. Heart rate variability in unmedicated patients with bipolar disorder in the manic phase. Psychiatry and Clinical Neurosciences. 2014;68(9):674-82.
19. Fasmer OB, Liao H, Huang Y, Berle JØ, Wu J, Oedegaard KJ, et al. A naturalistic study of the effect of acupuncture on heart-rate variability. Journal of acupuncture and meridian studies. 2012;5(1):15-20.
20. Kolosnjaji B, Eckert C, editors. Neural network-based user-independent physical activity recognition for mobile devices. International Conference on Intelligent Data Engineering and Automated Learning; 2015: Springer.
21. Inoue M, Inoue S, Nishida T. Deep recurrent neural network for mobile human activity recognition with high throughput. Artificial Life and Robotics. 2018;23(2):173-85.
22. Ignatov A. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. Applied Soft Computing. 2018;62:915-22.
23. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems. 2017;28(10):2222-32.
24. Marblestone AH, Wayne G, Kording KP. Toward an integration of deep learning and neuroscience. Frontiers in Computational Neuroscience. 2016;10:94.
25. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA. 2939778: ACM; 2016. p. 1135-44.
26. Bzdok D, Ioannidis JPA. Exploration, Inference, and Prediction in Neuroscience and Biomedicine. Trends in Neurosciences. 2019.
27. Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016;19(3):404-13.
28. Liaw A, Wiener M. Classification and regression by randomForest. R news. 2002;2(3):18-22.
29. Ahmad MW, Mourshed M, Rezgui Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings. 2017;147:77-89.

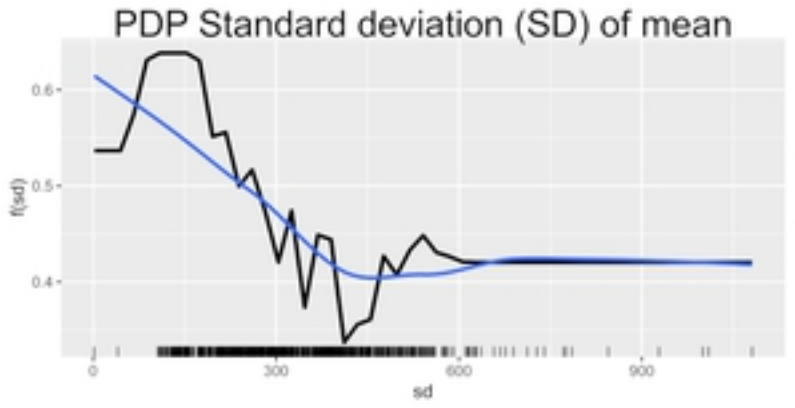
30. Molnar C. Interpretable machine learning: A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book2018. Available from: https://leanpub.com/interpretable-machine-learning.
31. Berle JO, Hauge ER, Oedegaard KJ, Holsten F, Fasmer OB. Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. BMC Research Notes. 2010;3:149-.
32. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th ed.). Washington DC: American Psychiatric Press; 1994.
33. Montgomery S, Asberg A. A new depression scale designed to be sensitive to change. Br J Psychiat. 1979;134.
34. Garcia-Ceja E, Riegler M, Jakobsen P, Tørresen J, Nordgreen T, Oedegaard KJ, et al., editors. Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients. Proceedings of the 9th ACM Multimedia Systems Conference; 2018. https://doi.org/10.1145/3204949.3208125: ACM.
35. Garcia-Ceja E, Riegler M, Jakobsen P, Torresen J, Nordgreen T, Oedegaard KJ, et al., editors. Motor Activity Based Classification of Depression in Unipolar and Bipolar Patients. 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS); 2018: IEEE.
36. Garcia-Ceja E, Osmani V, Mayora O. Automatic Stress Detection in Working Environments From Smartphones' Accelerometer Data: A First Step. IEEE Journal of Biomedical and Health Informatics. 2016;20(4):1053-60.
37. Riegler M, Lux M, Griwodz C, Spampinato C, de Lange T, Eskeland SL, et al., editors. Multimedia and medicine: Teammates for better disease detection and survival. Proceedings of the 24th ACM international conference on Multimedia; 2016: ACM.
38. Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering. 2006;30(1):25-36.
39. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002;16:321-57.
40. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.
41. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278-324.
42. Nabih-Ali M, El-Dahshan ELSA, Yahia AS. A review of intelligent systems for heart sound signal analysis. Journal of Medical Engineering & Technology. 2017;41(7):553-63.
43. Garcia-Ceja E, Brena R, Carrasco-Jimenez J, Garrido L. Long-term activity recognition from wristwatch accelerometer data. Sensors 2014;14(12):22500-24.
44. Pepe MS. The statistical evaluation of medical tests for classification and prediction: Medicine; 2003.
45. Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PloS one. 2017;12(6):e0177678.
46. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001:1189-232.
47. Greenwell BM. pdp: an R Package for constructing partial dependence plots. The R Journal. 2017;9(1):421-36.
48. Fleiss JL, Bigger Jr JT, Rolnitzky LM. The correlation between heart period variability and mean period length. Statistics in Medicine. 1992;11(1):125-9.
49. Krane-Gartiser K, Henriksen TEG, Morken G, Vaaler A, Fasmer OB. Actigraphic assessment of motor activity in acutely admitted inpatients with bipolar disorder. PloS one. 2014;9(2):e89574.
50. Krane-Gartiser K, Vaaler AE, Fasmer OB, Sørensen K, Morken G, Scott J. Variability of activity patterns across mood disorders and time of day. BMC psychiatry. 2017;17(1):404-.

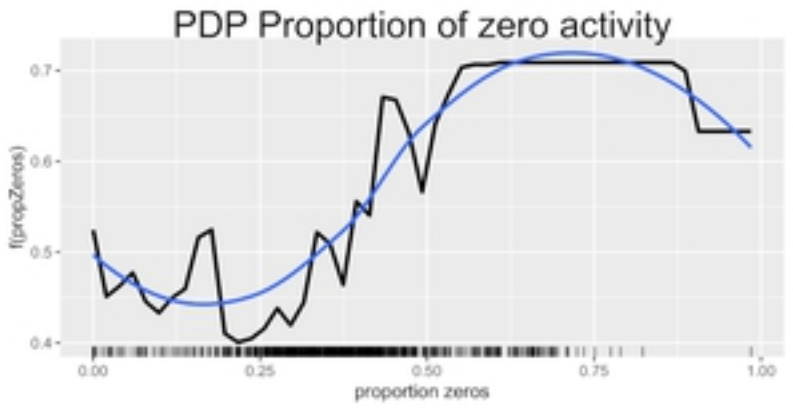
51. Oedegaard KJ, Neckelmann D, Fasmer OB. Type A behaviour differentiates bipolar II from unipolar depressed patients. Journal of Affective Disorders. 2006;90(1):7-13.
52. Pirkola S, Eriksen HA, Partonen T, Kieseppä T, Veijola J, Jääskeläinen E, et al. Seasonal variation in affective and other clinical symptoms among high-risk families for bipolar disorders in an Arctic population. International journal of circumpolar health. 2015;74(1):29671.
53. De Crescenzo F, Economou A, Sharpley AL, Gormez A, Quested DJ. Actigraphic features of bipolar disorder: A systematic review and meta-analysis. Sleep Med Rev. 2017;33:58-69.
54. Krane-Gartiser K, Asheim A, Fasmer OB, Morken G, Vaaler AE, Scott J. Actigraphy as an objective intra-individual marker of activity patterns in acute-phase bipolar disorder: a case series. International journal of bipolar disorders. 2018;6(1):8-.


PDP Mean activity
f(mean)
mean
PDP Standard deviation (SD) of mean activity
f(sd)
sd
PDP Proportion of zero activity
f(prop2zeros)
proportion zeros

Figure


PDP Mean activity
f(mean)
0.6
0.5
0.4
0.3
0
200
400
600
mean


PDP Standard deviation (SD) of mean
f(sd)
0.6
0.5
0.4
0
300
600
900
sd


PDP Proportion of zero activity
f(propZeros)
0.7
0.6
0.5
0.4
0.00
0.25
0.50
0.75
1.00
proportion zeros

Figure