UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Investigation of daily motor activity in depression using a wearable actimetry unit

Brino Valentina 2054723
Fantinato Chiara 2039066
Furlan Camilla 2055273
Guarise Pietro 2055628

Academic Year 2021 – 2022

Date of submission 12/06/2022

Abstract

Altered motor activity has been shown to be associated with depression. In this report, we present possible biomarkers to distinguish between depressive and normal state based on motor activity measured with an actigraph watch (Cambridge Neurotechnology Ltd, England, model AW4). We analyzed sensor data collected from 23 depressed patients and 32 healthy controls. After a pre-processing step, some metrics have been extracted from values of total activity, weekday activity and weekend activity (mean, standard deviation, coefficient of variation). It has been demonstrated the existence of significant differences of these metrics across the two states and, furthermore, the significant negative correlation between them and a clinical assessment of depression severity based on the Montgomery-Asberg Depression Rating Scale (MADRS). To evaluate biomarkers' performance in distinguishing patients from controls, a ROC analysis has been performed and the resulting ROC AUC estimates varied from 0.63 to 0.76. These findings indicate scores derived from motor activity tracking as ecological and non-invasive biomarkers to detect depression. In this study it has been considered summary measures of motor activity across the whole observation period and not also its daily time-course.

Table of contents

1. Background

Since 2010 mental health related problems have been the leading cause of global disability. Depression is number one of the most frequent disorders: it affects more than 300 million of people and is mostly manifested in sadness, agitation, loss of weight, suicide thoughts, as well as sleep disturbance and motor-activity alterations. [1, 2, 3]

In recent years, several studies have investigated the association between changes in physical activity and depressive symptoms. Evidence suggests that the mental state is somehow measurable through recordings of motor-activity. Despite this, objective methods are rarely used in psychiatric clinical practice. Reported findings show that wearable devices, like actigraph watches, provide a quantitative method to track circadian rest-activity variations in a reliable, non-invasive and accessible way. In particular, a previous study shows that depressive state is associated with reduced daytime motor-activity, increased nighttime activity compared to healthy controls. [4]

The aim of the present analysis was to explore the usefulness of digital data to detect a depressed patient, evaluating motor-activity monitoring as a biomarker for depression.

Firstly, we attempted to identify any difference between healthy people and depressed patients in daily activity, or any derived score, and a potential association between the motor activity and the depression score severity. Lastly, we tried to provide a demonstration of the daily activity's performance in discriminating patients from healthy controls.

2. Material and Methods

2.1 Dataset

The dataset is made up of three components. The first one consists of the daily motor activity measured for 23 depressed patients. The second one consists of the daily motor activity measured for 32 healthy controls. The daily motor activity has been monitored using an actigraph watch (Cambridge Neurotechnology Ltd, England, model AW4) worn on the right wrist. The actiwatch measures the activity using a piezoelectric accelerometer, has a sampling frequency of 32 Hz and records only movements over 0.05 g. The produced voltage is stored as activity count and the number of counts is proportional to the intensity of the movement [6]. In the dataset the activity count is reported in counts per minute (CPM). The last component of the dataset consists in the assessment of depression severity for the condition group. The metric used for the evaluation has been the Montgomery-Asberg Depression Rating Scale (MADRS), which considers 10 of the most commonly occurring symptoms in depressive illness (apparent sadness, reported sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, suicidal thoughts) [5]. The MADRS score in the range 0-9 reflects no depressive symptoms, while from 10 to 60 the depression severity increases [6]. The MADRS scores are reported in the table score.csv, having the following columns: number (patient identifier), days (number of days of measurements[1]), gender (1 for female; 2 for male), age (age in age groups), melanch (1 for melancholia; 2 for no melancholia), afftype (1 for bipolar type II; 2 for unipolar depressive; 3 for bipolar type I), edu (education grouped in years), inpatient (1 for hospitalized subjects; 2 for outpatient), marriage (2 for single; 1 for married or cohabiting), work (1 for working or studying; 2 for unemployed/sick leave/pension), madrs1 (MADRS score when the measurement started), madrs2 (MADRS score when the measurement stopped). Number, days, gender and age are available also for controls. The table score.csv presents some missing data. For three depressed subjects the presence or absence of melancholia has not been recorded, while for another condition the years of education are missing. Furthermore, all variables are missing for healthy controls except for number, days, gender, and age. In Figure 1 missing data patterns are reported.

Figure 2 and Figure 3 show the composition of the dataset based on the information available in score.csv.

Figure 4 and Figure 5 show how the mean of daily motor activity is distributed across weekdays and hours of day in the condition and control groups respectively. Note that the condition group seems to

[1] The number of days of measurements reported in scores.csv could not be consistent with the actual number of recording days because participants could have worn the actigraph watch for a longer period than the one agreed.

be characterized by lower activity with respect to the control group, especially on the weekend (dayNumber 1 and 7).

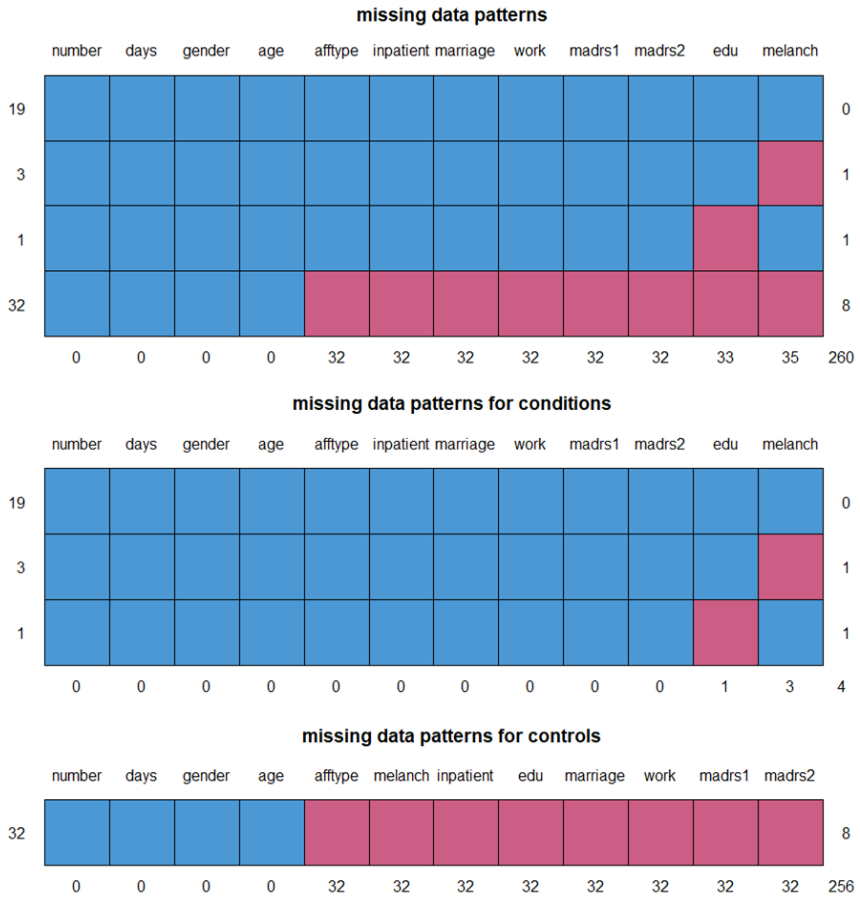

Figure 1. From the top to the bottom: the missing data patterns considering all the participants; the missing data patterns considering only the depressed subjects; the missing data patterns considering only the healthy controls.

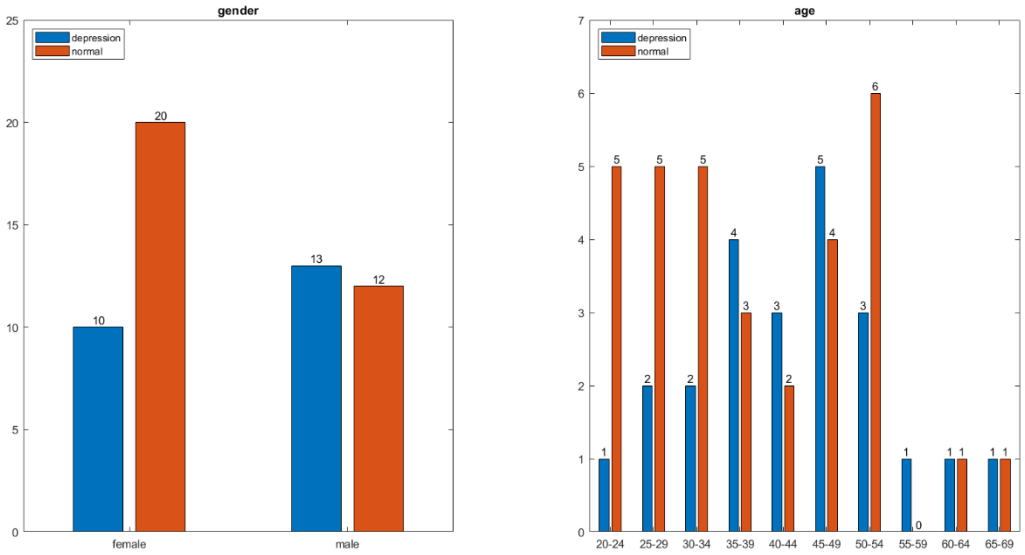

Figure 2. Gender and age across participants (55).

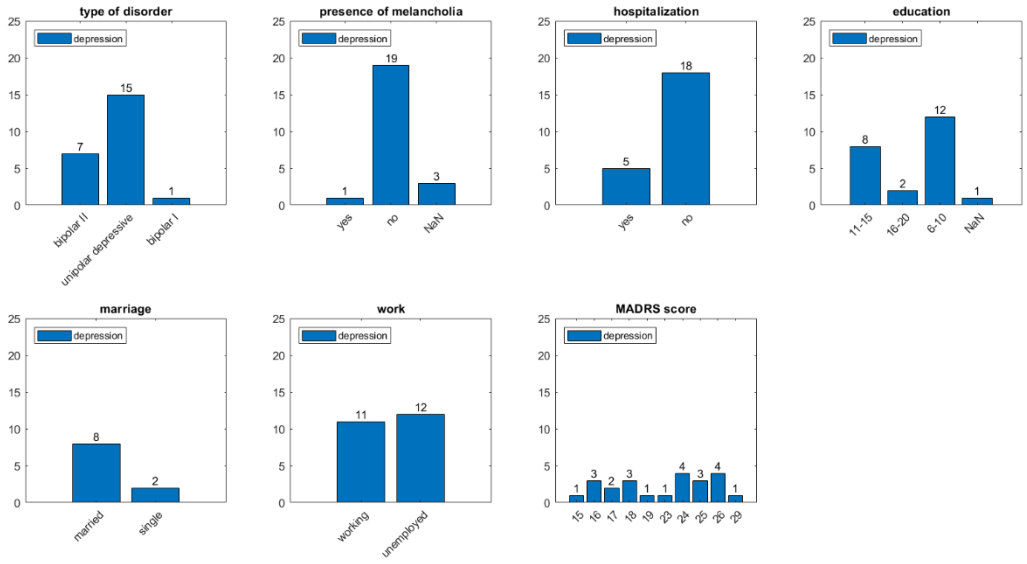

Figure 3. Type of disorder, presence of melancholia, hospitalization, education, marriage, work and MADRS score across depressed subjects (23). MADRS score has been obtained averaging madrs1 and madrs2.

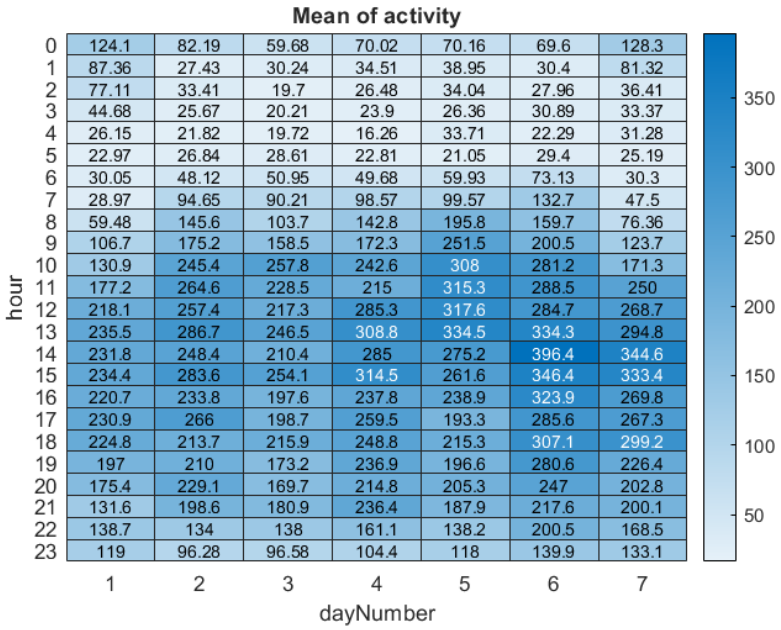

Figure 4. Heatmap of mean motor activity of condition group by weekdays and hour of days.

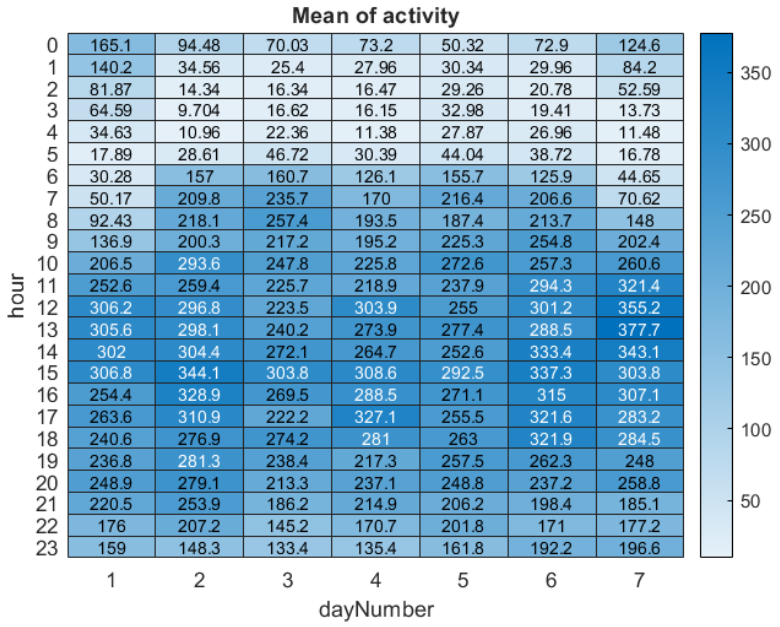

Figure 5. Heatmap of mean motor activity of control group by weekdays and hour of days.

2.2 Methods

For each participant, we selected only consecutive days with more than 10 hours of daily activity and we discarded all the others (which were supposed to be not enough informative for the analysis).

We decided to replace madrs1 and madrs2 missing data for all the healthy participants with integer values drawn from the discrete uniform distribution on the interval 0-9, according to MADRS cutoff point used to discriminate depressive from non-depressive symptoms [6]. We set a seed equal to 42 to guarantee reproducibility of the results. Moreover, we summarized the depression score severity in a single value (madrs), by computing the average between madrs1 and madrs2 for all the subjects.

From the cleaned data of both control and conditions groups, we assessed 10 potential biomarkers based on daily activity values: mean, standard deviation and coefficient of variation of total activity; mean, standard deviation and coefficient of variation of activity during the weekdays (from Monday to Friday); mean, standard deviation and coefficient of variation of activity during the weekend days (Saturday and Sunday); autocorrelation of the daily mean activity signal. We computed the latter as a binary biomarker, equal to 1 if the signal resulted to be autocorrelated and equal to 0 otherwise and its accuracy has been assessed.

For all other metrics, normal distribution of the biomarkers under examination was tested by Lilliefors test.

We performed statistical tests to verify statistical differences between controls and depressed patients in daily activity: we used t-test for normally distributed data and the Wilcoxon rank-sum test for non-normally distributed data (alpha = 0.05). We applied multiple comparison correction [7, 8] for

controlling the false discovery rate (FDR) and for estimating corrected significant p-values for each potential biomarker.

The relationship between each potential biomarker and depression score severity was investigated with a correlation analysis. Biomarkers' data and madrs of all participants were correlated by computing Pearson's correlations for normally distributed data and non-parametric Spearman's correlations for non-normally distributed data. We applied multiple comparison correction [7, 8] also on these results to evaluate statistical significance of the correlation coefficients.

To assess the statistical power of each biomarker to discriminate between depressive and non-depressive subjects, an analysis of Receiver Operating Characteristic (ROC) curves was performed. We used the ROC Area Under the Curve (AUC) as a performance metric.

Finally, we performed a sensitivity analysis investigating the association between our potential biomarkers and three covariates (days of measurements, gender and age). As mentioned above, we tested the correlation of each variable against all biomarkers by using Pearson's correlations for normally distributed data and non-parametric Spearman's correlations for non-normally distributed data. The resulting coefficients were corrected using the FDR approach [7, 8].

3. Results

Two-sample t-tests and non-parametric Wilcoxon rank-sum tests revealed that there are significative differences between healthy people and depressed patients for all biomarkers based on means, standard deviation and coefficient of variation (corrected p-values < 0.05 as shown in Figure 6). The accuracy computed for the autocorrelation of the daily mean activity was 0.53, suggesting that this biomarker was not helpful in distinguishing between normal and depressive state and therefore we decided to exclude it from further analyses.

The correlation analysis showed that motor activity is generally related to the depression score severity. The correlation coefficients were all negative and statistically significant for mean, standard deviation of total activity; mean, standard deviation of activity during the weekdays; mean and standard deviation of activity during the weekend days (Figure 7). We did not find significant associations between madrs scores and the coefficients of variation of activity across days.

The ROC curves for each biomarker are shown in Figure 8. ROC AUC estimates varied from 0.63 to 0.76 and the best performance was observed for the coefficient of variation of activity during the weekends, inferring that it seemed to distinguish better between conditions and controls.

The sensitivity analysis, performed for both groups of subjects, revealed that only the standard deviation of total activity and the standard deviation of activity during the weekdays were negatively correlated with the variable age (Figure 9), but with a borderline p-value (p-value = 0.47).

In contrast, none of the correlations between the biomarkers against days of measurements and gender covariates were statistically significant after correction for multiple comparisons.

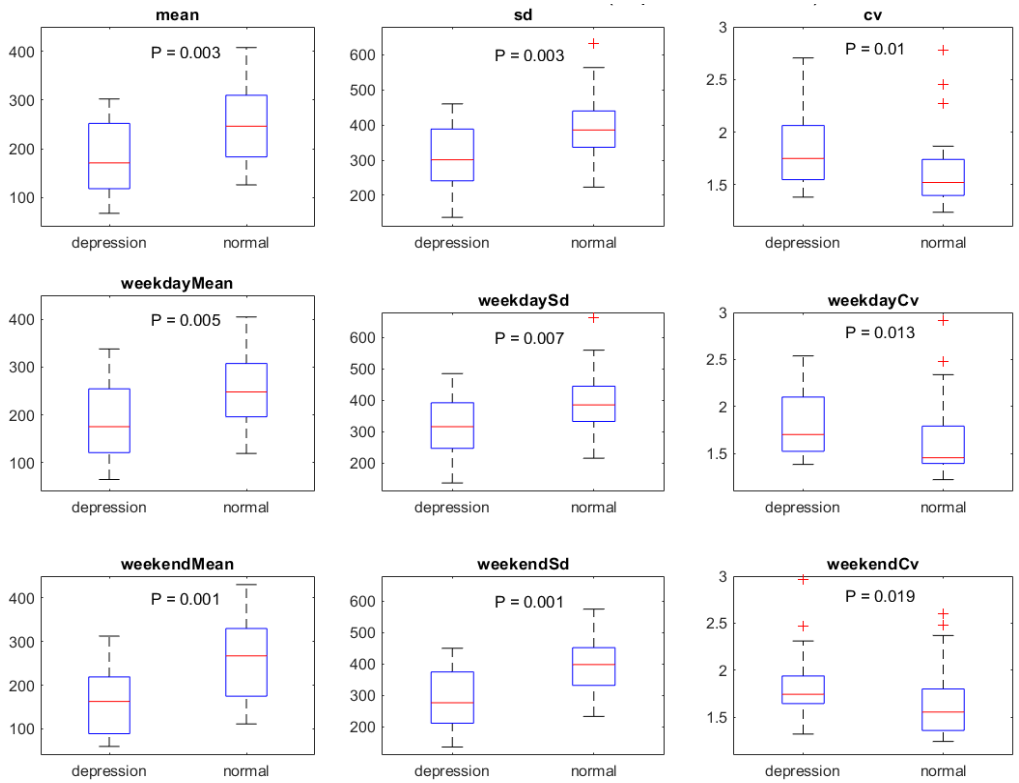

Figure 6. Differences between normal state and depressive state. Boxplots show biomarkers data grouped by the presence or absence of depression. Statistical significance of the differences between the two groups is proved by the correspondent p-values (all below 0.05), resulting from multiple comparison correction of the results obtained from statistical tests (t-test for normally-distributed data and Wilcoxon rank-sum test for non-normally distributed data).

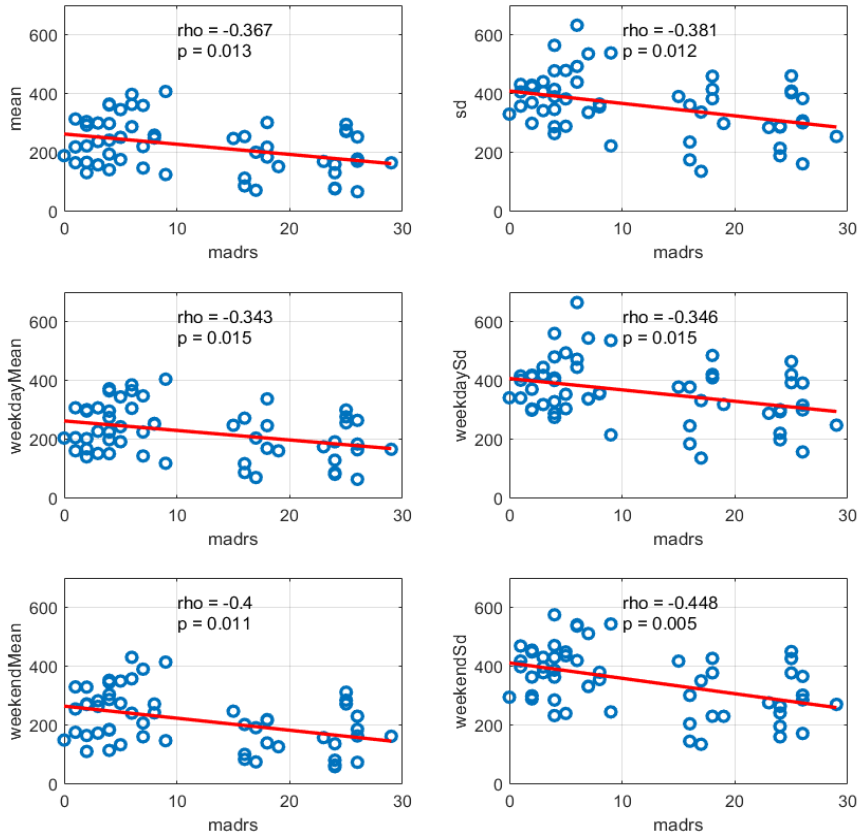

Figure 7. Significant correlations between biomarkers and madrs, with the respective p-values after multiple comparison correction.

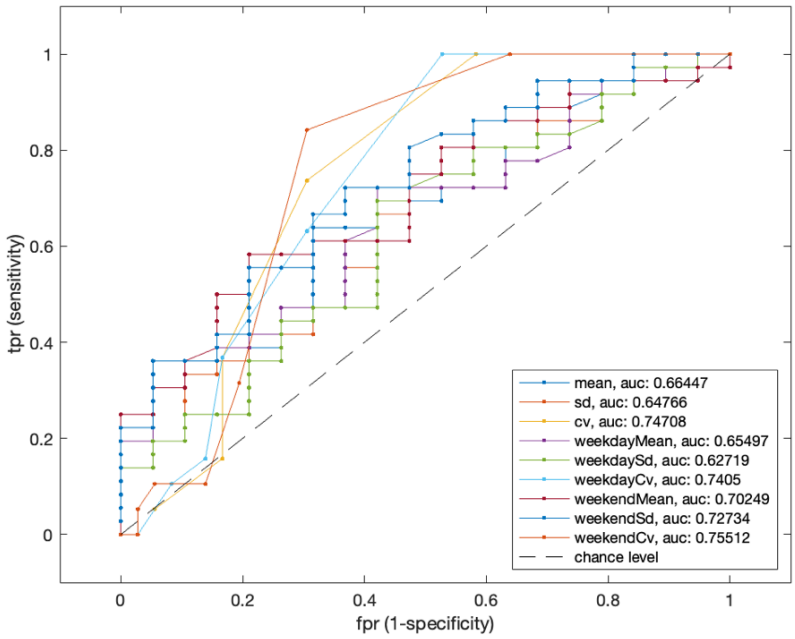

Figure 8. Receiver Operating Characteristic (ROC) curves for the classification of participants into depressive and non-depressive groups using biomarkers extracted from motor activity.

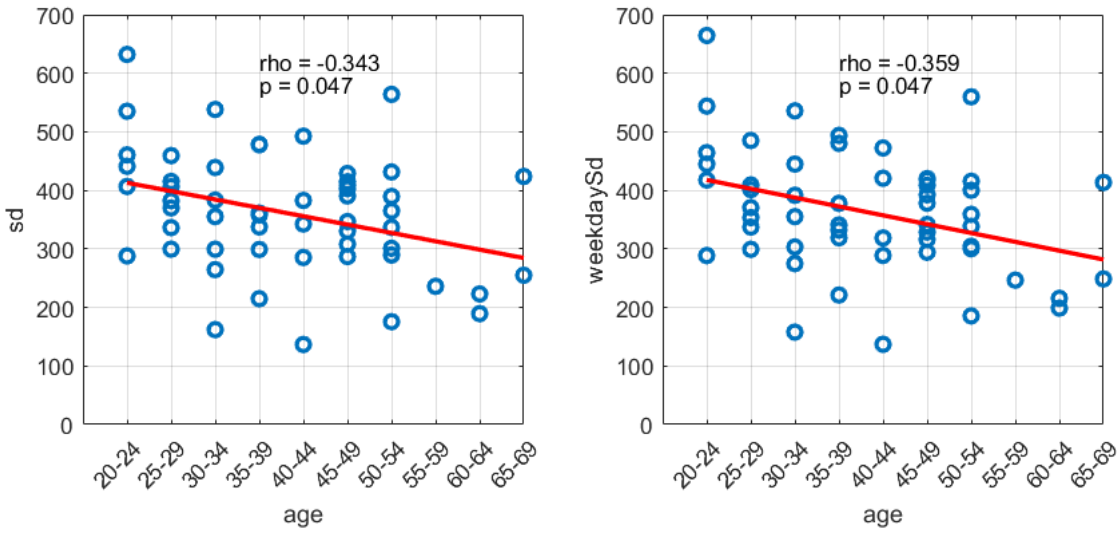

Figure 9. Significant correlations of biomarkers against age, with the respective p-values after multiple comparison correction.

4. Discussion

We computed possible biomarkers for depression based on motor activity, which present significant differences between depressed patients and healthy controls (mean, standard deviation and coefficient of variation of total activity; mean, standard deviation and coefficient of variation of activity during the weekdays; mean, standard deviation and coefficient of variation of activity during the weekend days) and, in particular, they are lower in the first ones than in the latter. Thus, we concluded that differences between depressive and normal state in motor activity there exist, as supported by the results of a previous study showing that depressive state is associated with reduced daytime motor-activity, increased nighttime activity compared to healthy controls [4]. We decided to distinguish between weekdays and weekends activity based on what we observed looking at Figure 4 and Figure 5: the condition group seems to be characterized by lower activity with respect to the control group, especially in the weekend. Beyond the biomarkers presenting differences between depressive and normal state, we considered a further possible biomarker for depression based on the autocorrelation of the daily mean motor activity signal, in line with our hypothesis that the pattern of daily mean motor activity in terms of autocorrelation could be different between conditions and controls. Unfortunately, we noticed that this metric would not be helpful for our classification problem (accuracy = 0.53), thus we didn't consider it anymore.

The correlation analysis revealed a significant negative correlation between MADRS score and all our biomarkers other than the coefficient-of-variation-based ones. This finding suggests that these metrics could be capable of distinguishing between depressive and normal state. To compute these correlations, we imputed the MADRS scores of healthy controls, which were missing, with values drawn from the discrete uniform distribution on the interval 0-9, since on the MADRS scale this interval is related to absence of depressive symptoms. We took this decision because we assumed that MADRS scores for controls had not been collected since it was obvious that their value would have been within the 0-9 interval. This analysis also shows that there were no correlations between biomarkers and days, gender and age other than for the standard deviation of total activity and the standard deviation of activity during the weekdays which resulted to be negatively correlated with the variable age, but with a borderline p-value (p-value = 0.047). Thus, we concluded with the absence of correlation between the considered covariates and the biomarkers. We considered in the correlation analysis only the variables without missing data for both the depressive and normal state because we were interested in including both the states in our computations.

Being the AUCs of our biomarkers all greater than 0.5, we demonstrated that they are all capable of distinguishing patients from controls better than what a chance classifier could do. In particular, we

noticed that all the coefficient-of-variation-based metrics and the weekend metrics seem to classify better than the others.

A limitation of our study lies in the fact that we considered summary measures of the motor activity across the whole observation period and not also its daily time-course. In this way we could have wasted some useful information.

Further studies based on a test-rested dataset are needed to enforce the reliability and reproducibility of our biomarkers. Finally, collecting values for covariates such as work and marriage also for healthy controls could provide further suggestions on what could be helpful, beyond the motor activity, to distinguish patients from controls, in order to build a more reliable model to predict depression state.

References

[1] Mark Olfson, Benjamin G Druss, and Steven C Marcus. 2015. Trends in mental health care among children and adolescents. New England Journal of Medicine 372, 21 (2015), 2029–2038.

[2] Guilherme V Polanczyk, Giovanni A Salum, Luisa S Sugaya, Arthur Caye, and Luis A Rohde. 2015. Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. Journal of Child Psychology and Psychiatry 56, 3 (2015), 345–365.

[3] Jean M Twenge. 2015. Time period and birth cohort differences in depressive symptoms in the US, 1982–2013. Social Indicators Research 121, 2 (2015), 437–454.

[4] Christopher Burton, Brian McKinstry, Aurora Szentagotai Tătar, Antoni Serrano- Blanco, Claudia Pagliari, and Maria Wolters. 2013. Activity monitoring in patients with depression: a systematic review. Journal of affective disorders 145, 1 (2013).

[5] Stuart A. Montgomery and MARIE Asberg. 1979. A new depression scale designed to be sensitive to change. The British journal of psychiatry 134, 4 (1979), 382–389.

[6] Enrique Garcia-Ceja, Michael Riegler, Petter Jakobsen, Jim Tørresen, Tine Nordgreen, Ketil J. Oedegaard, and Ole Bernt Fasmer. 2018. Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients. In Proceedings of the 9th ACM Multimedia Systems Conference (MMSys '18). Association for Computing Machinery, New York, NY, USA, 472–477.

[7] Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological). 57(1), 289-300.

[8] Benjamini, Y. & Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics. 29(4), 1165-1188.