Digital Health and Human Behaviour Project: Classifying depressed patients using diurnal activity patterns

Anonymous

Aalto University
School of Science

Contents

1 Introduction

Depression is a common mental disorder and a leading cause of disability worldwide (WHO 2021), and its prevalence will increase even more in the coming years (Garcia-Ceja et al. 2018). Depression can be associated with multiple subsequent health outcomes, such as future coronary events and cardiac deaths, other physical diseases including cancer and osteoporosis and severe psychiatric disorders such as schizophrenia (Kaseva et al. 2016). Therefore, depression can be seen as a big burden to the public healthcare as well as to individuals suffering from it.

The use of sensors to monitor personal health has also become more common in these days, and one of the most commonly measured metrics is activity (Garcia-Ceja et al. 2018). Many researches have been conducted about the connection between physical activity and depression, from different points of view. Some of the studies focus on physical activity as treatment for depression (e.g. Silveira et al. 2013), and others on how physical activity can prevent depression (e.g. Choi et al. 2019). However, not so many studies have studied the bidirectional connection between physical activity and depressive symptoms (Azevedo Da Silva et al. 2012.) Roshanaei-Moghaddam et al. (2009) describe the bidirectional association as follows: *"Many cross-sectional studies have reported that depressed patients are more sedentary. However, this association may be bidirectional: depression may lead to decreased levels of exercise due to low motivation and energy and decreased exercise may be a risk factor for depression."* Nevertheless, the association exists, but it is quite complex.

In addition to the physical activity, also the diurnal rhythms of the activity are related to the association between activity levels and depression (George et al. 2021). According to studies, the activity of non-depressed people is at its highest closer to the middle of the day, and its lower in early mornings and late evenings and nights. For depressed people, the activity rhythm peaks later in the evening and they are more active in the night time. (George et al. 2021; Garcia-Ceja et al. 2018.)

The aim of this project is to explore whether the diurnal activity patterns can be used in diagnosing depressed people, i.e. classifying depressed and non-depressed people. We use the activity data set including the actigraph watch measurements from 23 condition group subjects and 32 control group subjects. In the article that provided the used data, Garcia-Ceja et al. (2018) conduct same kind of classification, however, instead of diurnal rhythms, they use only single statistical values from each subject: the overall mean activity and the overall standard deviation of the mean activity. Jakobsen et al. (2020) use more advanced classification methods (Deep Neural Network, Convolutional Neural Network and Random Forest which is also used in this project) with the same data set, but again, they use the mean activity, standard deviation of the mean activity and proportion of zero activity as their features. George et al. (2021) use the diurnal rhythms in their study, but instead of classification they use tools from complexity science. Therefore, it can be said that this report brings something new to the existing literature.

The methods include exploring the diurnal activity patterns using descriptive analytics and based on that, feature extraction is conducted. Four different classification methods are used and evaluated with multiple metrics. Key result of the project is that the diurnal activity patterns can be used in this kind of classification, but more research on the topic is required.

The rest of this report is organized as follows. Chapter 2 presents the problem formulation in more detail. Chapter 3 describes the used dataset and Chapter 4 presents the used methods and their limitations. In Chapter 5, the results of analysis are presented and finally in Chapter 6 the results are discussed and conclusions made.

2 Problem Formulation

The research problem of this report is to evaluate whether the diurnal activity patterns can be used to classify subject to be either depressed or not. The diurnal rhythm means certain behavior occurring in specific times of the day, from day to day. The opposite of diurnal is nocturnal, which refers to rhythm in the night time, but in this report with diurnal rhythms we refer to rhythm during the whole 24 hours of the day.

The research question is **"Can the diurnal rhythms of activity levels be used in classifying depressed people, and what would the accuracy of such classification be?"** The aim is not to find optimal solution, but rather try different algorithms to see how well they work and what are their accuracies. The diurnal rhythms of each subject are measured as follows. The 24-hour period is divided into four intervals: morning, day, evening and night. The intervals can be of different lengths. The activity levels of an individual subject in each of the intervals is calculated by counting all the activities occurring in that time interval from the whole recording period and taking their mean. The mean values are then scaled to obtain better accuracy.

There is also practical motivation behind this kind of classification. National Institute for Health and Care Excellence (NICE) mention in their clinical guideline for recognition and management of depression in adults, that when assessing a person who may have depression, the assessment should not rely only on symptom count, but also the degree of functional impairment and disability associated with the possible depression should be taken into consideration. (NICE, 2009.) Therefore, it would be beneficial to use also activity-based assessment when making the depression diagnosis. The classification methods presented in this report could be used in making such assessment and it could support the other diagnosing methods.

The formulation of the problem doesn't take every aspect into consideration. In this analysis, the different reasons people might have behind their diurnal activity patterns and levels are not acknowledged. The diurnal activity patterns depend on other factors besides depression. For example, people can have temporary reasons for insomnia, such as stress, which affects the diurnal activity rhythms (more activity in the night time, maybe less activity during day). People also have their characteristic rhythm, "internal clock", which make their rhythm differ from other people. People working in shifts have good reasons for changes in the rhythm or for an unusual rhythm.

3 Dataset

The data consists of activity data collected from 23 depressed patients (condition group) and from 32 non-depressed contributors (control group). It was originally collected for the study of motor activity in schizophrenia and major depression in the study of Berle et al. (2010). However, Garcia-Ceja et al. (2018) provided the data without the patients with schizophrenia. The condition group include both unipolar and bipolar depressed patients and the control group consists of 23 hospital employees, 5 students and 4 former patients without current psychiatric symptoms.

The activity is monitored with an actigraph watch (Actiwatch, Cambridge Neurotechnology Ltd, England, model AW4). The watch uses a piezoelectric accelerometer that records the amount and duration of movement in all directions, as well as the integration of intensity. With sampling frequency of 32Hz, movements over 0.05 g are recorded and the corresponding voltage is stored as an activity count. Total activity counts were recorded in one minute intervals and the number of counts is proportional to the intensity of movement. (Garcia-Ceja et al. 2018.)

Different participants used the actigraph watch for different time periods, for which the statistics can be seen in Table 1. The data was gathered between years 2002 and 2006.

	Control group	Condition group
Mean	23.1	17.6
Std	8.5	3.9
Min	15	14
Max	47	30

Table 1: Statistics of number of collected days of activity

The activity data of collected for each subject over the certain time period, has three columns: timestamp (one minute intervals), date (date of measurement) and activity (activity measurement from the actigraph watch).

Data column	Data type	Description
timestamp	datetime (YYYY-MM-DD HH:MM)	The time of when the recording is done (one minute intervals)
date	date (YYYY-MM-DD)	Date when the recording is done
activity	INTEGER	Activity measurement from the actigraph watch

Table 2: Columns of the activity data and their explanations

Besides the activity data, the data set also includes separate scores.csv, which has for example information about the age, sex, education, work and other things of the subject. For the condition group, it also has clinical information. However, only the activity data measured by actigraph watch was considered in this analysis, so therefore the scores.csv is not described in detail.

4 Methods

In this section, the methods for conducting the analysis are described. First, the data and its characteristics were described and explored using visualizations. Based on the obtained observations, feature extraction was done in order to make the classification easier to compute as well to obtain better accuracy. Lastly, the methods for making the classification are described.

4.1 Descriptive and explorative analytics

It was required to do both subject and group level observations from the data. In Figure 1, there are examples of the activity for two subjects, one from the control group and one from the condition group.

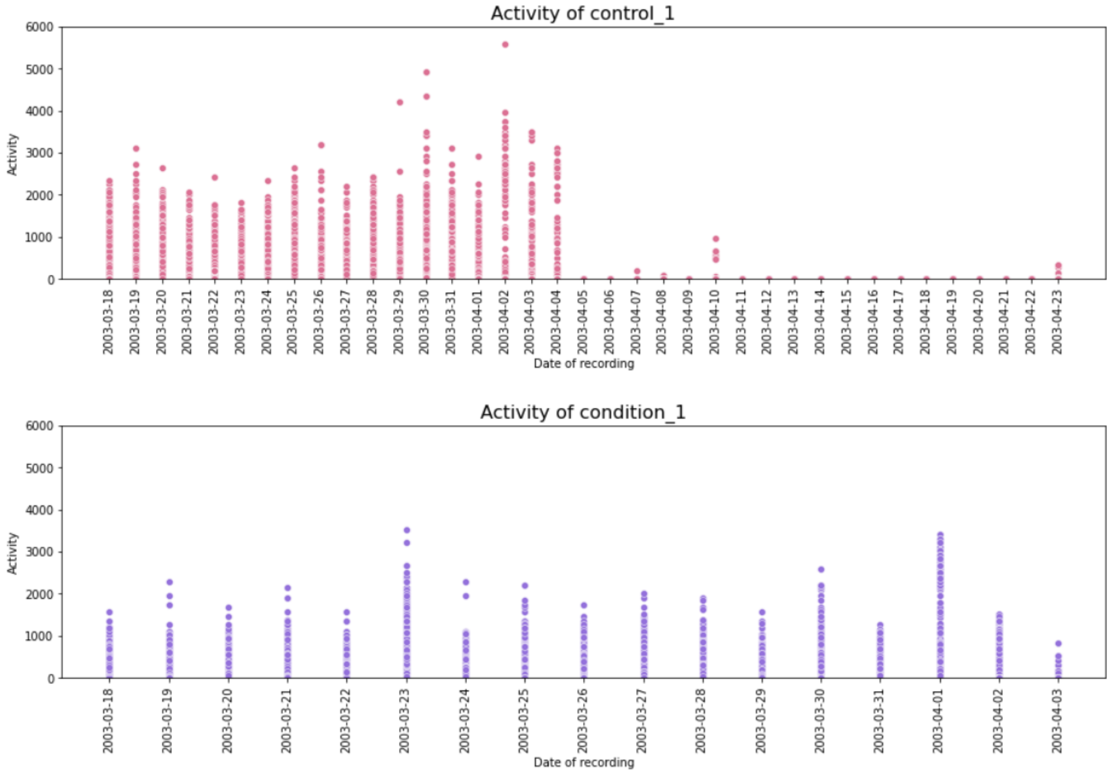

Figure 1: Example of activity records from condition and control group

From the Figure 1, it can be seen that the both subjects wore the actigraph watch in March - April, 2003. The subject belonging to control group had lot of activity in the first half of the recording period but for some reason, in the other half of the period the most of the activity is around zero. The subject belonging to condition group has quite even amounts of activity for each recording date, but their activity is in general lower than the control group subject in their first half of the recordings.

Because the activity was recorded from the subjects for different time periods, in this analysis we are not interested in the differences between days, even though there might be some interesting findings, for example differences between weekends and weekdays. Instead, we are interested in the diurnal activity patterns. In order to explore the differences in daily activity patterns of control and condition group, the activity recordings were put into hourly bins. For each subject, the mean activity for each hour was calculated. Therefore, we didn't have to consider the different lengths of time periods.

In Figure 2, the first plot shows the mean activity of each of the 24 hours for control group and the second plot shows the same for condition group. Each line represents the mean activity levels of one participant. The x-axis shows the hour bins, the times are presented in 24-hour time.

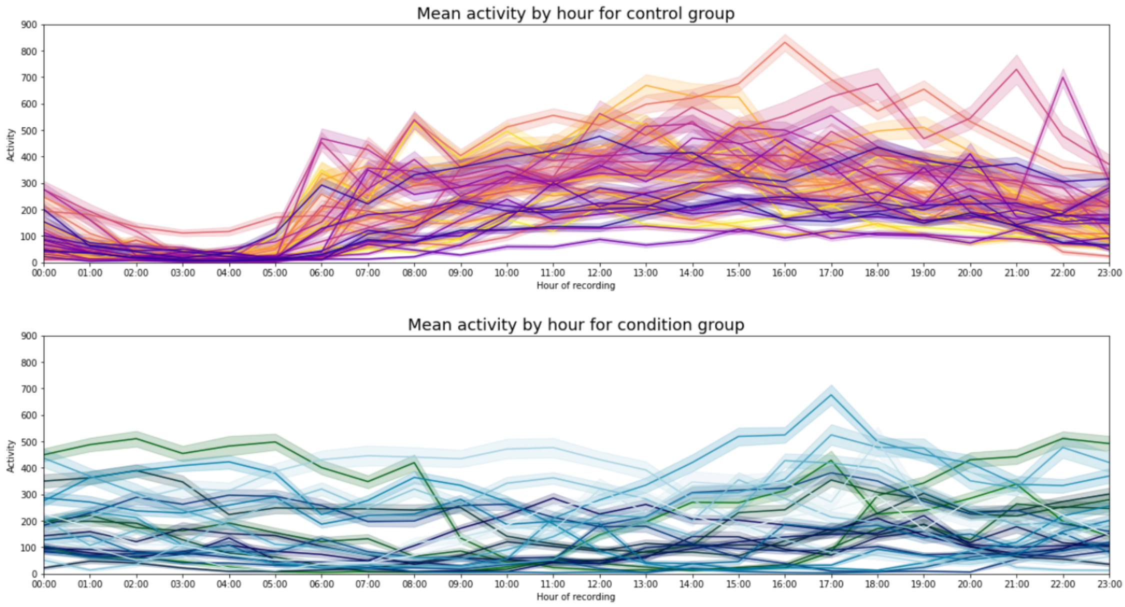

Figure 2: Mean activity in hourly bins for control and condition groups

From the Figure 2, it can be seen that most of the control group follow a similar diurnal pattern in their activity. In the night, between midnight and early morning hours, the activity is significantly lower (around zero for most of the subjects), whereas in the daytime there is clear growth in the activity. For the condition group, there is more variation in the diurnal patterns between subjects and there is no similar pattern as with the control group. In general, the mean activity during days in the condition group is also lower than in the control group.

As mentioned in the Chapter 1, according to the studies the activity of non-depressed people is at its highest closer to the middle of the day, and its lower in early mornings and late evenings and nights. For depressed people, the activity rhythm peaks later in the evening and they are more active in the night time. Depressed people have also lower activity levels in general. (George et al. 2021; Garcia-Ceja et al. 2018.) The observations obtained from the Figure 2 affirm the conclusions from the studies.

4.2 Feature extraction

After finding the differences in the daily activity patterns for condition and control group, we decided to use this information in our classification. Instead of using all the time stamps, the features could be extracted from the mean activities of certain time intervals.

In Figure 3, each subject is represented by their mean activity in the day time (between 6am and 6pm) and their mean activity in the night time (between 6pm and 6am). The times for the bins were decided roughly estimating which bins would show the difference between control and condition group.

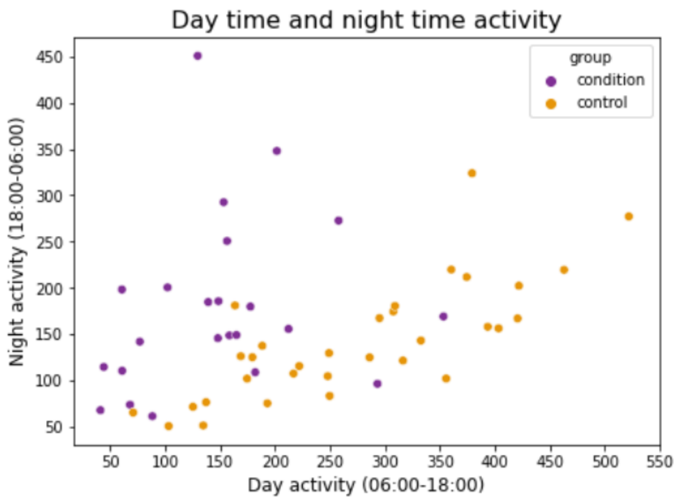

Figure 3: Day and night time activity for control and condition group

From the Figure 3, it can be concluded that control group has lower night activity but generally higher day activity than the condition group and vice versa. However, there is still some points that differ from their own group. To be able to classify the condition and control group even better, it was decided that instead of only dividing day and night activity, we will use four time intervals: morning, day, evening and night.

The four time intervals were chosen as follows. The night interval starts at midnight and lasts til 5am, the morning interval starts at 5am and ends at 9am, the day interval starts at 9am and ends at 4pm, and the evening interval starts at 4pm and ends at midnight. The interval are different lengths but they were determined based on the differences between plots in the Figure 2.

Finally, the mean activity level for each time bin was calculated for each participant. The activity level values were then normalized to range between 0 and 1 using min-max scaler. In Table 3 the final features used for the classification are shown.

Feature	Data type	Description
night activity	FLOAT	Activity between midnight and 5am
morning activity	FLOAT	Activity between 5am and 9am
day activity	FLOAT	Activity between 9am and 4pm
evening activity	FLOAT	Activity between 4pm and midnight
depressed	BINARY	1 if the observation is from the condition group, 0 otherwise

Table 3: Features used for classification

4.3 Classification

Four different classification methods were tested. The goal was not to find the optimal classifier or the optimal solution, but to explore whether classification would yield in sensible results, so the default methods were used. The four tested classifiers are K-Nearest-Neighbours (KNN), Random Forest, Decision Tree and Support Vector Classification (SVC). We used the functions provided by scikit-learn (sklearn), which is a Python library for machine learning.

Initially, the data was divided to training and test sets in order to evaluate the accuracy of classification. The size of the training sample was 41 and the size of the testing sample was 14. However, changing the random seed for the function that split the data to train and test sets affected the results a lot, so we decided to do cross-validation instead. In cross-validation, instead of dividing the data into training and test sets, the data is divided into k sets, where each of the sets is used as test set whereas the other sets act as the training data. In this task, five folds were used for the cross-validation. For each evaluation metric, the mean of the five folds is calculated as the result.

The first classification method used was K-Nearest Neighbours (KNN), the idea of which is that the object is assigned to a class based on the classes of its k nearest neighbours. The default settings of sklearn's KNeighborsClassifier function were used, which were 5 neighbours and Euclidean distance as the distance metric. Next classification algorithms were Random Forest and Decision Tree classifier. Random Forest classifier uses multiple decision trees and chooses the class by majority vote. The default settings of the sklearn functions were used. Lastly, Support Vector Classification was conducted. We used again the default settings, where the regularization parameter C is 1.0 and the kernel method is RBF.

Classification was evaluated using different metrics and the results are presented in Chapter 5.

4.4 Limitations

There are some limitations related to the methods described in this section. First of all, the used data set is relatively small, as there were only 55 observations all together. The data is also a bit unbalanced, because there are 23 positive cases (depressed) and 32 negative cases (not depressed). This may have an effect on the results, but we didn't want to leave out any of the cases because then the amount of used data would be even smaller. Also, in the real life settings, there are more non-depressed than depressed people, so it would be even useful to train the classifier with less depressed people if this analysis was done for real purposes.

Secondly, in the feature reduction part, the time intervals were determined based on solely the descriptive analytics. The intervals could have been determined by using also research literature as support. However, it was mentioned in the literature that the diurnal rhythms follow the same kind of pattern that we obtained in our analysis. Another idea is that we could have tested different intervals to see which yields to best results.

Thirdly, in the classification part, classification algorithms were chosen arbitrarily and the default setting were used. This part could also have better scientific support, however, the aim of this project was more to see whether classification could be used, not to find out which classification method is optimal.

5 Results

In this section, the results of the analysis are presented. For this dataset, Garcia-Ceja et al. (2018) listed different metrics recommended to use for the evaluation, from which we selected eight: true positive (TP), true negative (TN), false positive (FP), false negative (FN), precision, recall, accuracy and F1-score. TP is the number of correct classified positive samples, TN is the number of correct classified negative samples, FP is the number of negative samples wrongly classified as positive and FN the number of positive samples incorrectly classified. Precision expresses the fraction of true positives among all samples classified as positive. Recall presents the ratio of correctly classified relevant samples among all the relevant samples. F1-score is the harmonic mean of precision and recall, and accuracy represents the percentage of correctly classified samples. (Garcia-Ceja et al. 2018.)

The obtained results for each classification method can be seen in Table 4. The presented values are the means of the five different folds of cross-validation.

	KNN	Random Forest	Decision Tree	SVC
TN	6.0	5.8	5.4	5.8
FP	0.4	0.6	1.0	0.6
FN	1.0	1.4	0.8	1.2
TP	3.6	3.2	3.8	3.4
Precision	0.91	0.88	0.81	0.86
Recall	0.79	0.72	0.84	0.75
Accuracy	0.87	0.82	0.84	0.84
F1 score	0.83	0.74	0.81	0.79

Table 4: Results of metrics for each classifier

From the Table 4, it can be concluded that all the classification methods give a good accuracy. KNN gives the best accuracy, but other methods are not far behind. However, accuracy alone can be misleading especially for imbalanced datasets.

Looking at the other metrics, it seems that KNN, Random Forest and SVC tend to classify samples as negative (not depressed) more often. The amount of false positives is smaller than the amount of false negatives. This seems reasonable because there are more negative samples in the data. For Decision Tree, it's other way around, so it seems to classify samples falsely as positive more often. For this reason, the precision of KNN, Random Forest and SVC is higher than for Decision Tree, whereas the recall for Decision Tree is the highest. Both KNN and Decision Tree yield to almost the same F1-score. However, in this situation, the KNN should be chosen as the best classifier, because it's not desired outcome to label people falsely as depressed based on their activity (especially if this classification wouldn't be the only thing taken into consideration when making the diagnosis).

Main result of this analysis was that diurnal activity patterns can be used for classifying depressed and non-depressed patients. With limited amount of data, and without further optimization of the classifiers, even 87% accuracy can be obtained, whereas in the study by Garcia-Ceja et al. (2018), where they used only single statistical measures of the same data set, the best accuracy was only 73%. This result encourages for further research on this matter. Implications and ideas for future analysis are discussed in the Chapter 6.

6 Conclusion and discussion

This project explored the association between the diurnal pattern of physical activity levels and depression, and whether that could be used in classification. Previous studies had found support for the association between the diurnal physical activity levels and depression, but not many of them had tried to use that association in classification. The studies that used the association for classification, used only single statistical measures from the data, instead of looking at the activity levels in different times of the day. Therefore, the aim of this project was to fill that gap and see if accurate classification could be done using activity levels in different times of the day.

In the descriptive and explorative analytics, differences between the diurnal activity patterns between condition and control group were found using visualization. It was found that non-depressed participants have relatively similar diurnal activity pattern with each other, where they are least active in the night time, and the activity increases in the day time. For depressed participants, there was no similar visible pattern, but more variation between the subjects. In general, depressed participants were more active in the night time and less active in the day time compared to non-depressed participants. According to these findings, mean activity levels in four time intervals (morning, day, evening and night) were calculated for each subject and those were used as the features in classification.

Four different classification methods were used and their performance was evaluated using multiple metrics. K-Nearest-Neighbours provided the best accuracy (87%).

The main finding was however that the diurnal rhythms could potentially be used as features for classifying depressed patients. According to National Institute for Health and Care Excellence (2009), it would be reasonable to use also activity-based assessment as a tool in making a depression diagnostics, besides other symptoms. Therefore, it can be seen that this kind of classification could have an actual and potential purpose also in practice.

However, this analysis also has its limitations. The most fundamental problem is that the analysis doesn't take into consideration the different reasons people might have for their differing diurnal activity levels. For example, people who work in night shift will most likely have kind of different activity levels compared to normal and that doesn't mean they are depressed. For this reason, this kind of classification should never be used without background information.

Another main limitation is the small amount of data. It would be reasonable to continue the analysis presented in this report with a larger data sample, so that more reliable results could be obtained. Also more data would help with finding the optimal classification methods. As mentioned in the Chapter 4.4, the methods have some other limitations besides the amount of data. The data is also a bit unbalanced, which might affect the results. There is some arbitrariness in the methods as well, as the intervals for determining the used features were based on visual analysis and in the classification part the used algorithms were chosen arbitrarily and the default settings were used. Despite the multiple limitations, the yielded results seemed to be reasonable and support the findings made in the previous studies.

There are many ideas for future analysis to continue on this problem. First of all, the data included the scores.csv, which had some valuable information which was not utilized in this analysis. For example, it included the information about whether the condition group subject where unipolar depressed or bipolar depressed. The classification could be extended to classifying both unipolar and bipolar patients. The scores.csv also had other interesting features that might reveal something valuable when studied together with the activity levels.

Another thing that would be interesting topic to study in the future using this data set, is sleep analysis. For example Kaseva et al. (2019) found positive correlations between sleep problems and symptoms of depression, using self-administered questionnaires for measuring physical activity and sleep. The data set used in this report could give good addition to the self-administered questionnaires to bring a more objective viewpoint. Something that wasn't either fully utilized in this report is the other activity rhythms besides the daily rhythm. Taking a longer time period into consideration and trying to find some weekly or monthly patterns in activity might also lead to valuable discoveries.

In conclusion, the analysis managed to find answer to the research question and it was possible to do classification based on the diurnal rhythms of the subjects. This result is inspiring and encourages for further research on this matter, despite the limitations.

References

Azevedo Da Silva, M. Singh-Manoux, Archana. Brunner, E.J. Kaffashian, S. Shipley, M.J. Kivimäki, M. Nabi, H. (2012) Bidirectional association between physical activity and symptoms of anxiety and depression: the Whitehall II study. *Eur J Epidemiol*, 27(7), 537-546. doi: 10.1007/s10654-012-9692-8

Berle, J.O. Hauge, E.R. Oedegaard, K.J. Holsten, F. Fasmer, O.B. (2010) Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. *BMC Research Notes*, 3 (149). doi: 10.1186/1756-0500-3-149

Choi, K.W. Chen, C. Stein, M.B. Klimentidis, Y.C. Wang, M.-J. Koenen, K.C. Smoller, J.W. (2019) Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study. *JAMA Psychiatry*, 76(4): 399–408. doi:10.1001/jamapsychiatry.2018.4175

Garcia-Ceja, E. Riegler, M. Jakobsen, P. Tørresen, J. Nordgreen, T. Oedegaard, K. J. Bernt Fasmer, O. (2018) Depresjon: A Motor Activity Database of Depression Episodes in Unipolar and Bipolar Patients. In *Proceedings of 9th ACM Multimedia Systems Conference.* Amsterdam, Netherlands. pp. 473-479. doi:10.1145/3204949.3208125

George, S.V. Kunkels, Y.K. Booji, S. Wichers, M. (2021) Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed. *Scientific Reports*, 11. doi: 10.1038/s41598-021-92890-w

Jakobsen, P. Garcia-Ceja, E. Riegler, M. Antonsen Stabell, L. Nordgreen, T. Tørresen, J. Fasmer, O.B. Oedegaard, K.J. (2020) Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls *PLoS One*, 15(8). doi: 10.1371/journal.pone.0231995

Kaseva, K. Rosenström, T. Hintsa, T. Pulkki-Råback, L. Tammelin, T. Lipsanen, J. Yang, X. Hintsanen, M. Hakulinen, C. Pahkala, K. Hirvensalo, M. Hutri-Kähkönen, N. Raitakari, O.T. Keltikangas-Järvinen, L. (2016) Trajectories of Physical Activity Predict the Onset of Depressive Symptoms but Not Their Progression: A Prospective Cohort Study *Journal of Sports Medicine.* doi: 10.1155/2016/8947375

Kaseva, K. Dobewall, H. Yang, X. Pulkki-Råback, L. Lipsanen, J. Hintsa, T. Hintsanen, M. Puttonen, S. Hirvensalo, M. Elovainio, M. Raitakari, O. Tammelin, T. (2019) Physical Activity, Sleep, and Symptoms of Depression in Adults—Testing for Mediation *Medicine Science in Sports Exercise*, 51(6), 1162-1168. doi: 10.1249/MSS.0000000000001896

NICE (2009). Depression in adults: recognition and management. *Clinical guideline [CG90].* Available at: https://www.nice.org.uk/guidance/cg90 (visited 27.12.2021)

Roshanaei-Moghaddam, B. Katon, W.J. Russo, J. (2009) The longitudinal effects of depression on physical activity. *General Hospital Psychiatry*, 31(4), 306-315. doi: 10.1016/j.genhosppsych.2009.04.002

Silveira, H. Moraes, H. Oliveira, N. Coutinho, E.S.F. Laks, J. Deslandes, A. (2013) Physical Exercise and Clinically Depressed Patients: A Systematic Review and Meta-Analysis. *Neuropsychobiology*, 67(2), 61–68. doi:10.1159/000345160

WHO (2021) Depression. Available at: https://www.who.int/news-room/fact-sheets/detail/depression (visited 27.12.2021)