DEPRESJON: A Motor Activity Database of Depression Episodes in Unipolar and Bipolar Patients

Enrique Garcia-Ceja
Department of Informatics, University of Oslo, Norway.
enriqug@ifi.uio.no

Michael Riegler
Department of Informatics, University of Oslo. Simula Metropolitan Center for Digital Engineering, Norway.
michael@simula.no

Petter Jakobsen
Division of Psychiatry, Haukeland University Hospital. Department of Clinical Medicine, University of Bergen, Norway.
petter.jakobsen@helse-bergen.no

Jim Tørresen
Department of Informatics, University of Oslo, Norway.
jimtoer@ifi.uio.no

Tine Nordgreen
Division of Psychiatry, Haukeland University Hospital. Department of Clinical Psychology, Faculty of Psychology, University of Bergen, Norway.
Tine.Nordgreen@uib.no

Ketil J. Oedegaard
K.G. Jebsen Centre for Neuropsychiatry and NORMENT. Division of Psychiatry, Haukeland University Hospital. Department of Clinical Medicine, University of Bergen, Norway.
Ketil.Odegaard@uib.no

Ole Bernt Fasmer
Department of Clinical Medicine, Section for Psychiatry, Faculty of Medicine and Dentistry, University of Bergen. Division of Psychiatry, Haukeland University Hospital. K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, Bergen, Norway.
Ole.Fasmer@uib.no

ABSTRACT

Wearable sensors measuring different parts of people's activity are a common technology nowadays. In research, data collected using these devices also draws attention. Nevertheless, datasets containing sensor data in the field of medicine are rare. Often, data is non-public and only results are published. This makes it hard for other researchers to reproduce and compare results or even collaborate. In this paper we present a unique dataset containing sensor data collected from patients suffering from depression. The dataset contains motor activity recordings of 23 unipolar and bipolar depressed patients and 32 healthy controls. For each patient we provide sensor data over several days of continuous measuring and also some demographic data. The severity of the patients' depressive state was labeled using ratings done by medical experts on the Montgomery-Asberg Depression Rating Scale (MADRS). In this respect, the here presented dataset can be useful to explore and understand the association between depression and motor activity better. By making this dataset available, we invite and enable interested researchers the possibility to tackle this challenging and important societal problem.

CCS CONCEPTS

• **Human-centered computing** → **Mobile devices**; • **Applied computing** → **Health informatics**; • **Computing methodologies** → *Classification and regression trees*;

KEYWORDS

Depression, Bipolar Disorder, Depressive Disorder, Major, Motor Activity, Machine Learning, Artificial Intelligence, Dataset

MMSys'18, June 12–15, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5192-8/18/06…$15.00
https://doi.org/10.1145/3204949.3208125

ACM Reference format:
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 9th ACM Multimedia Systems Conference, Amsterdam, Netherlands, June 12–15, 2018 (MMSys'18)*, 6 pages.
https://doi.org/10.1145/3204949.3208125

1 INTRODUCTION

The use of on body sensors to monitor personal health has become quite normal these days. Modern people are collecting vast amounts of data every day, for purposes such as increasing quality of life, supervise their fitness levels, or even to change bad habits. This data holds a lot of potential besides measuring the quantity of daily steps or calories burned, since continuous recordings of heart rate and activity levels usually are collected. There is an increasing awareness in the field of psychiatry on how these activity data relates to various mental health related issues such as changes in mood, personality, inability to cope with daily problems or stress and withdrawal from friends and activities [22, 24].

Since 2010 mental health related problems are the main cause for years lived with disability worldwide. Depression is number one of the most frequent disorders and the current trend is indicating that the prevalence will increase even more in the coming years [23, 26, 30]. Dealing with depression can be demanding since it can create physically, economically and emotionally problems often leading to problems with work and sick leaves [32].

Mental health problems are related to disturbance in internal biological systems[17]. These are complex systems and, because relations between the sensor data and the mood are not well understood yet, changes within these systems are difficult to detect.

Research indicates that early warning signals occur in critical transition periods preceding abrupt noticeable changes of state, and is usually indicated by a phenomena called critical slowing down [28]. Critical slowing down indicates that the system becomes slower and slower in recovering from small disturbances, i.e., a reduced ability to restore itself to its original condition [2]. Depression and bipolar disorder are episodic mood disorders, where the pathologic state and the healthy state might be understood as representing different stable states separated by sudden changes [6].

In context to this; the state of biological systems are somehow measurable through recordings of motor-activity. Evidence indicates that a depressive state is associated with reduced daytime motor-activity, as well as increased nighttime activity when comparing to healthy controls [5]. Reduced motor-activity is likewise reported in bipolar depressions, besides increased variability in activity levels compared to others [29].

Activity and movement measurements have become an emerging topic in the field of mental health. Several studies use sensors to measure patients movements over time and connect them to diagnosis or self reports [4, 27]. Usually, in these studies the data is analyzed using standard linear and nonlinear statistical methods. Reported findings include increased autocorrelations and variances as indicators of a critical slowing down [28], and increased skewness is also observed [8]. As one can easily see such data also holds potential for machine learning applications which is used more and more in the context of psychiatry and psychology [7, 12, 15, 19].

Databases of this kind come with two difficult problems, (i) data in the medical field is often protected and difficult to access which makes it hard to compare results between different approaches, (ii) data often contains a small amount of positive examples but much more negative ones (episodes are usually not the norm and its much easier to collect normal data compared to relevant cases).

With the here presented dataset, called Depresjon after the Norwegian word for depression, we try to tackle these two challenges by releasing a completely open dataset that can be used for research purposes and includes a relevant amount of depression cases compared to control participants. The main contributions of this paper are therefore:

- (i) A novel, open dataset containing sensor data of patients with depression and control patients.
- (ii) A large amount of depressed and non depressed patients in the dataset.
- (iii) A baseline evaluation with machine learning algorithms for classifying depressed v.s. nondepressed days, including suggestions for evaluation metrics to use in future research.

The reminder of the paper is describing the process of collecting the data and the data itself. In addition to that, a baseline evaluation is presented including suggestions for metrics to use.

2 DATA COLLECTION

The here presented dataset was originally collected for the study of motor activity in schizophrenia and major depression [4]. Motor activity was monitored with an actigraph watch worn at the right wrist (Actiwatch, Cambridge Neurotechnology Ltd, England, model AW4). The actigraph watch measures activity by using a piezoelectric accelerometer that is programmed to record the integration of intensity, amount and duration of movement in all directions. The sampling frequency is 32Hz and movements over 0.05 g are recorded. A corresponding voltage is produced and is stored as an activity count in the memory unit of the actigraph watch. The number of counts is proportional to the intensity of the movement. The right wrist was chosen to make the procedure more convenient for the participants, since most of them have their watches around the left wrist, and it is cumbersome to have two devices on the same arm. Total activity counts were continuously recorded in one minute intervals.

3 DATASET DETAILS

The dataset consists of actigraphy data collected from 23 unipolar and bipolar depressed patients (condition group). Figure 1 gives an example of the data plotted for one patient during 24 hours. Five subjects were hospitalized during their data collection period, and 18 were outpatients. The severity level of the ongoing depression was rated by a clinician on the Montgomery-Asberg Depression Rating Scale [18] at the beginning and conclusion of the motor-activity recordings. In addition, the dataset contains actigraphy data from 32 non-depressed contributors (control group), consisting of 23 hospital employees, 5 students and 4 former patients without current psychiatric symptoms.

The dataset can be accessed via:

http://datasets.simula.no/depresjon/ or directly downloaded from **https://doi.org/10.5281/zenodo.1219550** and contains the following: Two folders, whereas one contains the data for the controls and one for the condition group. For each patient we provide a csv file containing the actigraph data collected over time. The columns are: timestamp (one minute intervals), date (date of measurement), activity (activity measurement from the actigraph

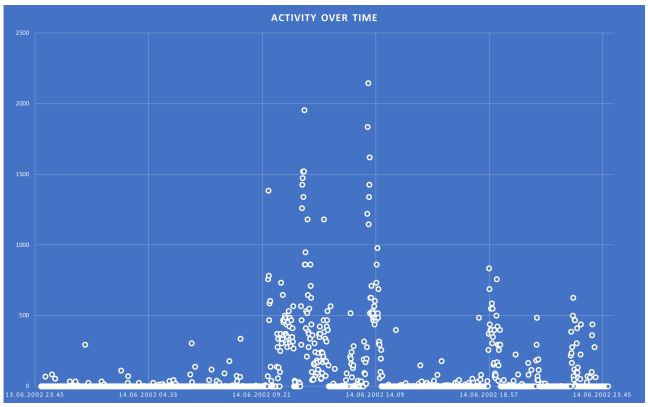

Figure 1: Example of activity data measured for one patient covering the duration of 24 hours.

watch). In addition, we also provide the MADRS scores in the file *scores.csv*. It contains the following columns; number (patient identifier), days (number of days of measurements), gender (1 or 2 for female or male), age (age in age groups), afftype (1: bipolar II, 2: unipolar depressive, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), inpatient (1: inpatient, 2: outpatient), edu (education grouped in years), marriage (1: married or cohabiting, 2: single), work (1: working or studying, 2: unemployed/sick leave/pension), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped).

3.1 Medical background

Depression is a mental disorder characterized by a feeling of emptiness or sadness, anxiety, sleep disturbance and general loss of initiative and interest in activities [21]. Symptoms like reduced energy, concentration problems, the feeling of worthlessness or guiltiness and various degree of suicidality might also be present, as well as delusions and hallucinations. The severity of a depression is determined by the amount of symptoms, their seriousness and duration, as well as the effect on social and occupational function [25]. Depression is associated with disrupted biological rhythms caused by environmental disturbance like seasonal change in natural light, disturbed daytime rhythms due to for instance shiftwork or longitude traveling, besides triggered by persistent daily rhythms deviating from the natural daylight cycle [1, 3]. In addition, the appearance of depressive symptoms relates to physical health issues, medical side effects, social factors, besides alcohol and substance abuse [21].

Depressions are also present in Bipolar disorder, a severe mental disorder characterized by intense mood fluctuations, as a counterpoint to the manic state which is associated with increased activity, reduced sleep, impulsivity, goal-directed actions and inflated self-esteem [10]. The main difference between bipolar disorder and unipolar depression is that mania is not present in the latter. Depression and bipolar disorder are genetic disorders, and can best be understood as an internal vulnerability to external circumstances disturbing the biological state [14]. But health issues, abuse, social features, life events, as well as other environmental circumstances might also cause symptoms of depression in the general population, e.g. for most people not associated with the genetic profile of bipolar disorder and unipolar depression. The lifetime prevalence of depressions is about 15% [25], but the incidences of episodes with a severity level not qualifying for a depressive diagnosis are far more prevalent [16]. It is well established that depression is characterized by altered motor activity [5], and that actigraph recordings of motor activity is an objective method for observing mood [29].

Alternative approaches for objectively detecting depression could be registration of heart rate [31] and voice recordings [33]. However, these methods are hardly studied in depression, probably because collection of such data is a far more complicated and challenging task than using a simple wrist worn actigraph to accumulate motor activity data.

3.1.1 Depression diagnosis. The Montgomery-Asberg Depression Rating Scale (MADRS) is used to grade the current severity of an ongoing depression [18]. Clinicians rate ten items relevant for depression based on observation and conversation with the patient, and the sum score (0-60) state the severity of the depression. Scores below 10 are classified as absence of depressive symptoms [9], and scores above 30 indicate a severe depressive state [20].

3.2 Dataset Overview

The actigraph devices were used by the study participants for an average of 12.6 days in the control and condition groups. See Table 1 for details. The total number of collected days was 693 comprising 402 days in the control group and 291 in the condition group (Figure 2). Note that the actigraph files might contain more days but only the first n days were considered in our analysis. Where n is the number of days reported in the *days* column from the *scores.csv* file. Figures 3 and 4 show heatmaps of the average activity level by weekday and hour of day for the control and condition group, respectively. The values were normalized across both groups to make them comparable. Clearly, the condition group presents less activity level, specially during weekends.

Table 1: Statistics of number of collected days by group.

	Control group	Condition group
Mean	12.6	12.6
Sd.	2.3	2.7
Max	20	18
Min	8	5

4 APPLICATIONS OF THE DATASET

Our vision is that the available data may eventually help researchers to develop systems capable of automatically detecting depression states based on sensor data. This dataset can be suitable (but not limited to) for the following applications.

- Use machine learning for depression states classification.
- MADRS score prediction based on motor activity data.
- Sleep pattern analysis of depressed v.s. nondepressed participants.

This dataset can be used as the basis for evaluating different machine learning methods and approaches such as: cost-sensitive

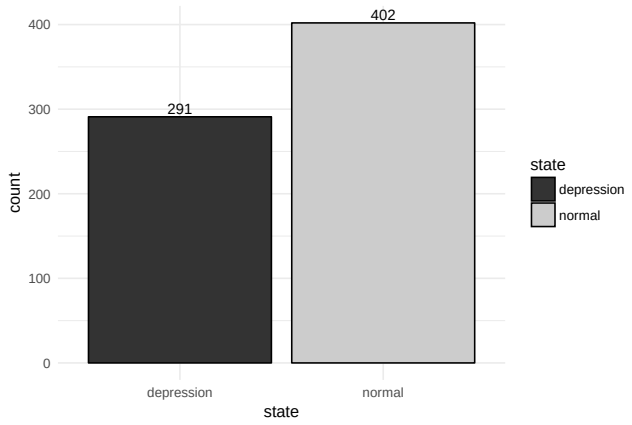

Figure 2: Distribution of days by state.

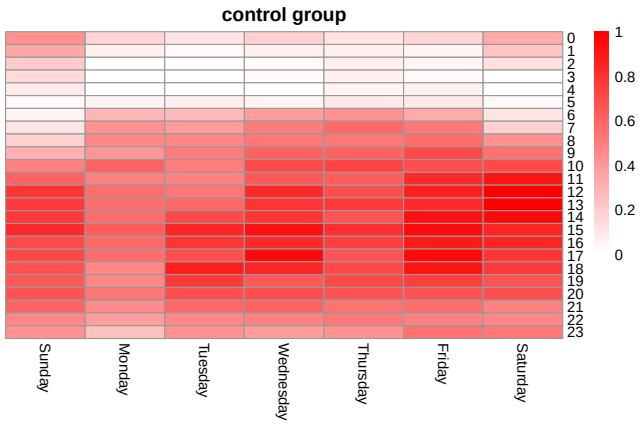

Figure 3: Heatmap of activity levels of control group by weekdays and hour of day.

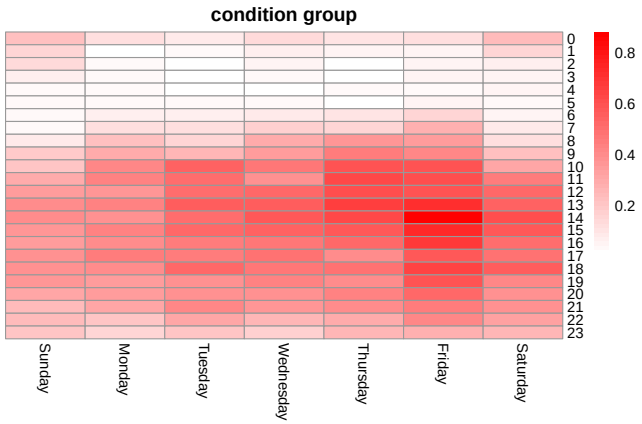

Figure 4: Heatmap of activity levels of condition group by weekdays and hour of day.

classification [13] and oversampling techniques for imbalanced class problems [11]. This dataset is also suitable for comparing different machine learning classification approaches such as feature based and deep learning based methods like convolutional neural networks and recurrent neural networks for time series.

5 SUGGESTED METRICS

Several different metrics can be used to evaluate classification algorithms. Many of these metrics come with different names if used in a medical context compared to for example information retrieval.

In this section a number of metrics are suggested for the presented dataset. A well chosen set of metrics depicting different aspects of the performance is the basis for a qualitative good evaluation. For the medical dataset it is also recommended to apply weighting of the metrics by the number of samples in the respective classes and report the weighted average, since it often occurs that classes in the medical field are imbalanced. Furthermore, to make comparison more fair it is advisable to present as many metrics as possible. Table 2 contains all suggested metrics and a short description.

6 BASELINE PERFORMANCE

In order to provide a baseline for a classification task (depressed v.s. nondepressed days), we tested different machine learning classification algorithms including: Nearest Neighbors, Linear kernel Support Vector Machine (SVM), Radial Basis Function kernel (RBF) SVM, Gaussian Process, Decision Tree, Random Forest, Neural Network, AdaBoost, Naive Bayes, Quadratic Discriminant Analysis (QDA)

Table 2: Recommended metrics for evaluation of the presented dataset. True positive (TP) number of correct classified positive samples, true negative (TN) number of correct classified negative samples, false positive (FP) number of negative samples wrongly classified as positive, false negative (FN) number of positive samples incorrectly classified as negative.

Metric	Description
Precision (PREC)	Depicts the fraction of true positives among those classified as positives. It is also called positive predictive value.
Recall/Sensitivity (REC/SEN)	This metric is the ratio of correctly classified relevant samples among all relevant samples in the dataset.
Accuracy (ACC)	Represents the percentage of correctly classified positive and negative samples. Can be misleading for imbalanced datasets and should be interpreted in combination with other metrics.
Specificity (SPEC)	SPEC, also called true negative rate depicts the classifiers performance in terms of correctly classified negative samples.
Matthews correlation coefficient (MCC)	MCC is a balanced measure which takes into account TP, FP, TN and FN. It can show the classifiers performance even if the classes are imbalanced.
F1-score (F1)	Harmonic mean of precision and recall.

Table 3: Weighted average classification performance (10-folded cross validation) reporting the metrics presented above. The best performing classifier is bold.

Classifier	Class	PREC	REC	ACC	SPEC	MCC	F1
Nearest Neighbors	depressed	0.395	0.705	0.675	0.669	0.318	0.5
Nearest Neighbors	nondepressed	0.878	0.669	0.675	0.705	0.318	0.758
Nearest Neighbors	weighted average	**0.752**	0.678	0.675	0.696	0.318	0.691
Linear SVM	depressed	0.577	0.721	0.727	0.734	0.433	0.638
Linear SVM	nondepressed	0.836	0.734	0.727	0.721	0.433	0.78
Linear SVM	**weighted average**	0.735	**0.729**	**0.727**	0.726	**0.433**	**0.724**
RBF SVM	depressed	0.546	0.732	0.724	0.724	0.426	0.622
RBF SVM	nondepressed	0.853	0.724	0.724	0.732	0.426	0.782
RBF SVM	weighted average	0.733	0.727	0.724	**0.729**	0.426	0.719
Gaussian Process	depressed	0.543	0.733	0.723	0.723	0.424	0.619
Gaussian Process	nondepressed	0.853	0.723	0.723	0.733	0.424	0.781
Gaussian Process	weighted average	0.732	0.727	0.723	0.729	0.424	0.718
Decision Tree	depressed	0.561	0.689	0.707	0.72	0.391	0.615
Decision Tree	nondepressed	0.813	0.72	0.707	0.689	0.391	0.763
Decision Tree	weighted average	0.711	0.707	0.707	0.702	0.391	0.702
Random Forest	depressed	0.509	0.702	0.7	0.704	0.375	0.585
Random Forest	nondepressed	0.838	0.704	0.7	0.702	0.375	0.764
Random Forest	weighted average	0.738	0.703	0.7	0.703	0.375	0.709
Neural Net	depressed	0.557	0.711	0.719	0.724	0.413	0.621
Neural Net	nondepressed	0.836	0.724	0.719	0.711	0.413	0.775
Neural Net	weighted average	0.727	0.719	0.719	0.716	0.413	0.715
AdaBoost	depressed	0.523	0.706	0.706	0.71	0.387	0.595
AdaBoost	nondepressed	0.838	0.71	0.706	0.706	0.387	0.767
AdaBoost	weighted average	0.733	0.709	0.706	0.707	0.387	0.71
Naive Bayes	depressed	0.663	0.63	0.694	0.747	0.379	0.645
Naive Bayes	nondepressed	0.716	0.747	0.694	0.63	0.379	0.731
Naive Bayes	weighted average	0.69	0.69	0.694	0.688	0.379	0.688
QDA	depressed	0.694	0.634	0.7	0.761	0.397	0.66
QDA	nondepressed	0.704	0.761	0.7	0.634	0.397	0.73
QDA	weighted average	0.699	0.697	0.7	0.699	0.397	0.695
ZeroR baseline	depressed	0.000	0.000	0.580	1.000	0.000	0.000
ZeroR baseline	nondepressed	0.580	1.000	0.580	0.000	0.000	0.734
ZeroR baseline	weighted average	0.337	0.580	0.580	0.580	0.000	0.426

and ZeroR (which is basically the majority class baseline). Each day was characterized by a feature vector which was computed by extracting a set of features on a per day basis from the activity level. The extracted features were the mean activity level, the corresponding standard deviation and the percentage of events with no activity i.e, activity level = 0. The features were normalized between 0 and 1 before the classification. The evaluation was performed using 10-fold cross validation and Table 3 provides an overview of the obtained results. Here, we can see that Linear SVM obtained the best overall weighted recall, accuracy, MCC and F1-score (bold in the table). In general, all methods tested can outperform the ZeroR baseline (basically majority class baseline reported in the last row in Table 3) but have problems to detect depression. On the contrary, detecting nondepressed seems not a big problem for most of the classifiers. Several methods reach a precision above 0.8 for the nondepressed class. This is also a good example why it is important to look at metrics beyond accuracy, precision and recall. MMC is in these cases a better and more accurate indicator for the real classification performance, which is in our baseline experiments rather low. A MCC close to 1 would be preferable. The best results in terms of classification precision for the depressed class were obtained by Naive Bayes and QDA with a precision above 0.65. The best recall for the depressed class is achieved by Gaussian Process with 0.733 which at the same time sacrifices precision for it (0.543).

All in all, it can be concluded that the here presented baseline classification results still hold a lot of improvement potential. The main reason therefor are most probably the simple statistical features used for the classification. Compressing the information into single statistical values removes information that could contain valuable hints for the classification task. Apart from that, the here presented baseline does not take the flow of time (time series) into account, which also can contain important hints. Nevertheless, a more detailed analysis is out of scope for this paper but especially for the depressed class an improvement should be possible by using more sophisticated methods such as time series analysis for the sensor data using deep learning or a sliding window approach with the addition of more discriminative features.

7 CONCLUSION

Reproducibility and comparability of results is an important factor of high quality research. In this paper we presented a dataset in the field of depression analysis enabling reproducibility and comparability, which makes it unique in the field of computer science and psychology. The dataset has been collected from real patients wearing actigraphy watches. Furthermore, high quality labels were created through the assessment of the patients via medical experts while they were wearing the sensors. In addition, we also provided a baseline evaluation and suggested metrics. Datasets in the medical field are rare and such a dataset as the here presented one can make multi-disciplinary research possible in order to improve care for the patients all over the world.

ACKNOWLEDGEMENTS

This publication is part of the INTROducing Mental health through Adaptive Technology (INTROMAT) project, funded by the Norwegian Research Council (259293/o70)

REFERENCES

[1] Lauren B Alloy, Tommy H Ng, Madison K Titone, and Elaine M Boland. 2017. Circadian rhythm dysregulation in bipolar spectrum disorders. *Current psychiatry reports* 19, 4 (2017), 21.
[2] Atiyeh Bayani, Fatemeh Hadaeghi, Sajad Jafari, and Greg Murray. 2017. Critical slowing down as an early warning of transitions in episodes of bipolar disorder: A simulation study based on a computational model of circadian activity rhythms. *Chronobiology international* (2017).
[3] William Bechtel. 2015. Circadian rhythms and mood disorders: are the phenomena and mechanisms causally related? *Frontiers in psychiatry* 6 (2015), 118.
[4] Jan O Berle, Erik R Hauge, Ketil J Oedegaard, Fred Holsten, and Ole B Fasmer. 2010. Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. *BMC research notes* 3, 1 (2010), 149.
[5] 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).
[6] Ole B Fasmer, Hagop S Akiskal, John R Kelsoe, and Ketil J Oedegaard. 2009. Clinical and pathophysiological relations between migraine and mood disorders. *Current Psychiatry Reviews* 5, 2 (2009), 93–109.
[7] Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. 2016. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE journal of biomedical and health informatics* 20, 4 (2016), 1053–1060.
[8] Vishwesha Guttal and Ciriyam Jayaprakash. 2008. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecology letters* 11, 5 (2008), 450–460.
[9] CJ Hawley, TM Gale, and T Sivakumaran. 2002. Defining remission by cut off score on the MADRS: selecting the optimal value. *Journal of affective disorders* 72, 2 (2002), 177–184.
[10] RM Hirschfeld. 2014. Differential diagnosis of bipolar disorder and major depressive disorder. *Journal of affective disorders* 169 (2014), S12–S16.

[11] Nathalie Japkowicz and Shaju Stephen. 2002. The class imbalance problem: A systematic study. *Intelligent data analysis* 6, 5 (2002), 429–449.
[12] N Keshan, PV Parimi, and Isabelle Bichindaritz. 2015. Machine learning for stress detection from ECG signals in automobile drivers. In *Big Data (Big Data), 2015 IEEE International Conference on*. IEEE, 2661–2669.
[13] Sotiris Kotsiantis, Dimitris Kanellopoulos, Panayiotis Pintelas, and others. 2006. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering* 30, 1 (2006), 25–36.
[14] Dominic Landgraf, Michael J McCarthy, and David K Welsh. 2014. The role of the circadian clock in animal models of mood disorders. *Behavioral neuroscience* 128, 3 (2014), 344.
[15] L. B. Leng, L. B. Giin, and W. Y. Chung. 2015. Wearable driver drowsiness detection system based on biomedical and motion sensors. In *2015 IEEE SENSORS*. 1–4. https://doi.org/10.1109/ICSENS.2015.7370355
[16] Peter M Lewinsohn, Ari Solomon, John R Seeley, and Antonette Zeiss. 2000. Clinical implications of "subthreshold" depressive symptoms. *Journal of abnormal psychology* 109, 2 (2000), 345.
[17] Eva M Marco, Elena Velarde, Ricardo Llorente, and Giovanni Laviola. 2016. Disrupted circadian rhythm as a common player in developmental models of neuropsychiatric disorders. In *Neurotoxin Modeling of Brain Disorders — Life-long Outcomes in Behavioral Teratology*. Springer, 155–181.
[18] 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.
[19] Oscar Martinez Mozos, Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, Radu Dobrescu, and Jose Manuel Ferrandez. 2017. Stress detection using wearable physiological and sociometric sensors. *International Journal of Neural Systems* 27, 02 (2017), 1650041.
[20] Matthias J Müller, Hubertus Himmerich, Barbara Kienzle, and Armin Szegedi. 2003. Differentiating moderate and severe depression using the Montgomery–Åsberg depression rating scale (MADRS). *Journal of affective disorders* (2003).
[21] NICE. 2009. National Institute for Health and Clinical Excellence. Depression in adults: recognition and management. NICE guideline CG90. https://www.nice.org.uk/guidance/cg90. (2009). [last visited, February 14, 2018].
[22] The National Institute of Mental Health IRC. 2018. Definitions of the RDoC Domains and Constructs. (2018). https://www.nimh.nih.gov/research-priorities/rdoc/definitions-of-the-rdoc-domains-and-constructs.shtml
[23] 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.
[24] Frank J Penedo and Jason R Dahn. 2005. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Current opinion in psychiatry* 18, 2 (2005), 189–193.
[25] Stephen Pilling, Ian Anderson, David Goldberg, Nicholas Meader, Clare Taylor, Two Guideline Development Groups, and others. 2009. Depression in adults, including those with a chronic physical health problem: summary of NICE guidance. *BMJ* 339, 10.1136 (2009).
[26] 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.
[27] Nadja Razavi, Helge Horn, Philipp Koschorke, Simone Hügli, Oliver Höfle, Thomas Müller, Werner Strik, and Sebastian Walther. 2011. Measuring motor activity in major depression: the association between the Hamilton Depression Rating Scale and actigraphy. *Psychiatry research* 190, 2 (2011), 212–216.
[28] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. 2009. Early-warning signals for critical transitions. *Nature* 461, 7260 (2009), 53.
[29] Jan Scott, Greg Murray, Chantal Henry, Gunnar Morken, Elizabeth Scott, Jules Angst, Kathleen R Merikangas, and Ian B Hickie. 2017. Activation in bipolar disorders: a systematic review. *JAMA psychiatry* 74, 2 (2017), 189–196.
[30] 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.
[31] Gaetano Valenza, Mimma Nardelli, Antonio Lanata, Claudio Gentili, Gilles Bertschy, Rita Paradiso, and Enzo Pasquale Scilingo. 2014. 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* 18, 5 (2014), 1625–1635.
[32] Marianne Agergaard Vammen, Sigurd Mikkelsen, Åse Marie Hansen, Jens Peter Bonde, Matias B Grynderup, Henrik Kolstad, Linda Kærlev, Ole Mors, Reiner Rugulies, and Jane Frlund Thomsen. 2016. Emotional demands at work and the risk of clinical depression: A longitudinal study in the Danish public sector. *Journal of occupational and environmental medicine* 58, 10 (2016), 994–1001.
[33] Nicola Vanello, Andrea Guidi, Claudio Gentili, Sandra Werner, Gilles Bertschy, Gaetano Valenza, Antonio Lanata, and Enzo Pasquale Scilingo. 2012. Speech analysis for mood state characterization in bipolar patients. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2104–2107.