HYPERAKTIV: An Activity Dataset from Adult Patients with Attention-Deficit/Hyperactivity Disorder (ADHD)

Steven Alexander Hicks*
SimulaMet, Norway

Andrea Stautland
University of Bergen, Norway

Ole Bernt Fasmer†
NORMENT, Haukeland University Hospital, Norway

Wenche Førland
Independent Researcher, Norway

Hugo Lewi Hammer‡
Oslo Metropolitan University, Norway

Pål Halvorsen*
SimulaMet, Norway

Kristin Mjeldheim
Independent Researcher, Norway

Ketil Joachim Oedegaard†
NORMENT, Haukeland University Hospital, Norway

Berge Osnes†
Haukeland University Hospital, Norway

Vigdis Elin Giæver Syrstad†
NORMENT, Haukeland University Hospital, Norway

Michael Alexander Riegler
SimulaMet, Norway

Petter Jakobsen†
NORMENT, Haukeland University Hospital, Norway

Abstract

Machine learning research within healthcare frequently lacks the public data needed to be fully reproducible and comparable. Datasets are often restricted due to privacy concerns and legal requirements that come with patient-related data. Consequentially, many algorithms and models get published on the same topic without a standard benchmark to measure against. Therefore, this paper presents HYPERAKTIV, a public dataset containing health, activity, and heart rate data from adult patients diagnosed with attention deficit hyperactivity disorder, better known as ADHD. The dataset consists of data collected from 51 patients with ADHD and 52 clinical controls. In addition to the activity and heart rate data, we also include a series of patient attributes such as their age, sex, and information about their mental state, as well as output data from a computerized neuropsychological test. Together with the presented dataset, we also provide baseline experiments using traditional machine learning algorithms to predict ADHD based on the included activity data. We hope that this dataset can be used as a starting point for computer scientists who want to contribute to the field of mental health, and as a common benchmark for future work in ADHD analysis.

CCS Concepts

• **Applied computing** → **Health informatics**; • **Computing methodologies** → **Machine learning**; *Cross-validation*; *Supervised learning*.

*Also affiliated with Oslo Metropolitan University, Norway
†Also affiliated with University of Bergen, Norway
‡Also affiliated with SimulaMet, Norway

MMSys 21, September 28-October 1, 2021, Istanbul, Turkey

ACM ISBN 978-1-4503-8434-6/21/09.
https://doi.org/10.1145/3458305.3478454

Keywords

Attention-Deficit Hyperactivity Disorder, ADHD, Actigraphy, Motor Activity, Heart Rate, Machine Learning, Artificial Intelligence, Dataset

ACM Reference Format:
Steven Alexander Hicks, Andrea Stautland, Ole Bernt Fasmer, Wenche Førland, Hugo Lewi Hammer, Pål Halvorsen, Kristin Mjeldheim, Ketil Joachim Oedegaard, Berge Osnes, Vigdis Elin Giæver Syrstad, Michael Alexander Riegler, and Petter Jakobsen. 2021. HYPERAKTIV: An Activity Dataset from Adult Patients with Attention-Deficit/Hyperactivity Disorder (ADHD). In *12th ACM Multimedia Systems Conference (MMSys '21) (MMSys 21), September 28-October 1, 2021, Istanbul, Turkey.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3458305.3478454

1 Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a diverse chronic condition affecting nearly five percent of the adult population. The disorder often severely impacts social and occupational functioning, as well as quality of life for those affected [38]. ADHD diagnostics is currently based on subjective evaluation and clinical observations, and it is therefore a need for more objective methods [23, 35]. Sensory data collected from patients and analyzed by machine learning techniques have gained considerable interest as a tool to support existing subjective diagnostic practices within mental health [14]. Within the field of objective ADHD diagnostics, support vector machines have shown promising abilities of discriminating between children with ADHD and healthy controls in movement data from accelerometers and gyroscopes [26]. Other studies have applied various neural network algorithms to data from the brain, like functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), with promising results for a similar discriminating approach [1].

Analysis of sensor data containing information about the mental health of a person requires reliable and reproducible results. Therefore, in addition to presenting results, it is important to make data, methods and results equally freely available. However, within the medical field, sharing data is often problematic due to ethical and legal restrictions. We have previously shared two datasets containing anonymized actigraph recordings of motor activity collected through clinical studies. First, the DEPRESJON dataset [13] which contains data from bipolar and unipolar depressed patients, as well as healthy controls. Second, the PSYKOSE dataset [19] which contains activation data from patients with schizophrenia. In this paper, we present another openly shared anonymized dataset containing various sensory data collected from patients referred to a private psychiatric out-patient clinic in need of a diagnostic evaluation of either ADHD, mood or anxiety disorders. A total of 103 patients were recruited, 51 of which were diagnosed with ADHD, and 52 with other diagnoses (clinical controls). The Norwegian Regional Medical Research Ethics Committee West approved the original protocol for the data collection, and all processes were in accordance with the Helsinki Declaration of 1975. The data collected were recordings of motor activity and heart rate, the output of a computerized test of attention-related problems, as well as various diagnostic and clinical assessments [10–12]. It is also important to point out that the actigraph equipment used to record motor activity are identical to the equipment used when collecting the files of the DEPRESJON [13] and PSYKOSE [19] datasets. Consequently, the 32 healthy controls included in those two datasets can be used as an additional comparison group when analyzing the new HYPERAKTIV dataset.

Recording of motor activity is a simple method of monitoring human rest and activity cycles, and is ordinarily collected with a wrist-worn actigraphic device that registers acceleration in the three-dimensional space. Data from actigraphs have been applied to studies of sleep [40] and psychiatric diagnosis like bipolar disorder [30], unipolar depression [5] and schizophrenia [39]. Regarding ADHD, studies of actigraph data from both adolescents and adults have recognized a great potential for motor activity in ADHD diagnostics [9, 23]. Activity data from one day is visualized in Figure 1. In addition, a study of circadian rhythmicity in adult ADHD [36] found augmented restlessness in sleep towards the end of the night, and increased activity in the afternoon, when comparing mature humans with ADHD to healthy controls. In previous studies of the motor activity data from the HYPERAKTIV dataset, the ADHD patients did not display evidence of general hyperactivity, but presented different activity patterns to controls for Fourier analyses, intra-daily variability and autocorrelation [11]. This dataset has never been analyzed by machine learning techniques. However, previous analyses of the DEPRESJON dataset using various machine learning approaches provided favorable discriminating abilities, especially for Deep Neural Networks [18].

Conner's Continuous Performance Test II (CPT-II) [8] is a computerized neuropsychological response test frequently used in the assessments of ADHD. Still, the specificity and discriminating abilities of CPT-II are somewhat uncertain and may vary between subtypes of ADHD and cognitive difficulties due to other conditions [25, 35]. Previously, our research group has found the CPT-II to be a useful supplement in the diagnosis of adult ADHD when applying linear and non-linear analytic tools to the current dataset [12].

Heart rate can be recorded using electrocardiogram (ECG) or photoplethysmography (PPG) technology. ECG is regarded as the benchmark and most reliable method [31], although it is not as mobile as wrist-worn PPG devices. The heart rate data included in

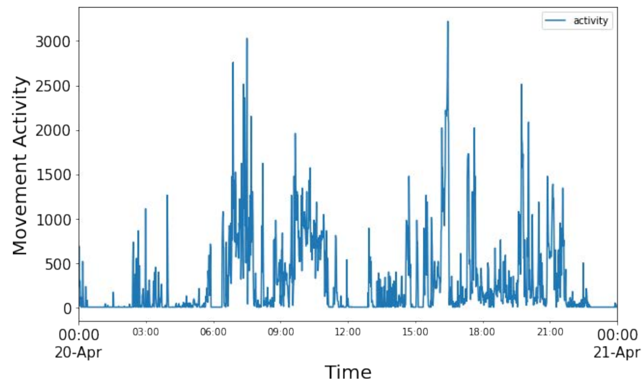


Figure 1: Example of 24 hours (from midnight to midnight) of actigraphy recordings from participant no. 57. Female, age group 17-29, diagnosed with ADHD, unipolar depression, anxiety disorder and cyclothymic temperament.

this dataset is ECG-based and recorded using a small chest-worn battery driven device, allowing free movement and long recordings. Another key point about the present heart rate data is that it has not been analyzed for any publication at the time of writing this paper. Heart rate data such as this can be used to calculate heart rate variability (HRV), which is a measure of variation in time between consecutive heartbeats. HRV is regulated by the autonomous nervous system and is viewed as a marker of autonomic activity. There is substantial evidence of reduced HRV in neuropsychiatric disorders such as depression and psychotic disorders [2], and efforts are being made to utilize this to aid diagnostics and disease management [24, 37]. HRV in relation to ADHD is sparsely studied, and results are somewhat conflicting [21]. Furthermore, previous studies on the subject have mostly focused on pediatric patient groups. Still, one systematic review found evidence of an association between ADHD and autonomic dysregulation [28], and we recently reported an association between lower HRV and poor emotional regulation in adolescents with ADHD [22].

The main contributions of this paper are:

(1) We compile and publish a fully open dataset containing sensory data collected from patients with ADHD and clinical controls, patients with mood or anxiety disorders.
(2) We provide a set of baseline machine learning experiments to benchmark the released dataset and evaluate its technical validity.
(3) We discuss and suggest possible future research directions and application scenarios using the dataset.

The remainder of this paper is organized as follows. First, we give a brief introduction to the medical background related to ADHD and the difficulties associated with this common disorder. Then, we give an in-depth presentation of the dataset, including details on the included files and directories. This is followed by a discussion on potential use cases that can be a starting point for future research. We then use one of the presented use cases to perform two experiments meant as a benchmark for the dataset. Lastly, we conclude this paper with a discussion on the experimental results and future hopes for the dataset.

2 Medical Background

Adult ADHD is a neuropsychiatric disorder characterized by the core symptoms hyperactivity, impulsivity and inattention. Symptoms must be present over time and negatively affect social, academic, or occupational functioning to qualify for the ADHD diagnosis. Inattention seems to be the primary feature of adult ADHD, characterized by disorganization and distraction, problems with staying on task and focused, as well as a propensity for persistent daydreaming. Many adults also have symptoms of impulsivity and hyperactivity, such as talkativeness, constant restlessness, and an inclination towards taking spontaneous decisions without evaluating potential consequences. Among children with ADHD, the latter condition generally has a more substantial manifestation [38]. Another difference between adult and pediatric ADHD is that ADHD seems to be a predominantly male disorder among children [29]. This gender difference becomes more equalized among adults [38]. Stimulants are considered the most effective medications for ADHD treatment. However, such medications have a pronounced potential for misuse and dependency. In general, adult ADHD is associated with an increased risk of developing substance-use disorders [38]. Furthermore, adult ADHD is also often concurrent with other psychiatric conditions such as mood and anxiety disorders, as well as sleeplessness [29, 34, 38].

3 Dataset Structure

The dataset is organized into four different items that can be found in the root directory. The directory *activity_data* contains the activity data collected from all participants, organized into separate files. Each file starts with few lines of metadata before the activity measurements start. The directory *hrv_data* contains the heart rate data collected from all participants, and like the activity data, it is separated into one file per participant. Each file starts with two lines of metadata before the IBI values start. The file *CPT_II_ConnersContinuousPerformanceTest.csv* contains the individual responses of the 360 CPT-II test trials, the omission and commission errors, and the ADHD Confidence Index. The file names *features.csv* contains the pre-extracted features used to perform the experiments presented in Section 7. Each line in the file corresponds to features for a single participant. Lastly, the file *patient_info.csv* contains all the information featured in Section 4. The file consists of 32 different columns, where each line corresponds to features for a single participant.

4 Dataset Details

As previously described, the dataset HYPERAKTIV contains time series of motor activity and heart rate, output from a neuropsychological computer test, the conclusions and sum scores of various diagnostic assessment tools, as well as the participant's sex, age, and prescribed medications. Sex is given as zero (female) and one (male). Participant ages are presented in four groups, where 1 = 17-29 years, 2 = 30-39 years, 3 = 40-49 years and 4 = 50-67 years. Of the 85 patients that recorded motor activity, 23 patients belonged to age group 1, 26 patients to age group 2, 24 to age group 3, and 12 to age group 4. The majority of the participants were not on medications. Among the participants diagnosed with ADHD who provided motor activity recordings, 73 percent were not medicated and only one individual was prescribed stimulants. Some more characteristics of the dataset is shown in Table 1.

Table 1: Characteristics and demographics of the 85 patients having recorded motor activity. Data from clinical assessments are given as mean (standard derivation). Differences tested with the Independent Samples t-test with Levene's test for Equality of Variance, at a significance level of $p < 0.05$. (NS equals $p > 0.05$)

	ADHD	**Controls**	p
N	45	40	
Sex (m/f)	24 / 21	20 / 20	
Bipolar (n)	16	20	
Unipolar (n)	15	10	
Anxiety (n)	18	26	
Substance (n)	12	7	
Other (n)	11	16	
CT (n)	26	22	
MDQ (n)	18	11	
WURS	51.5 (19.1)	29.5 (15.7)	**<0.001**
ASRS	47.6 (11.3)	34.0 (12.3)	**<0.001**
MADRS	13.3 (7.6)	14.0 (8.4)	NS
HASD-A	9.5 (4.6)	9.5 (4.8)	NS
HASD-D	4.4 (3.8)	5.7 (4.2)	NS
Medicated (n)	12	15	

Two experienced and certified psychiatrists performed diagnostic assessments of all patients, using the Mini-International Neuropsychiatric Interview (MINI Plus, version 5.0.0) [32]. The conclusions of the diagnostic interviews are presented as seven diagnostic variables that are scored as not present (0), present (1), or unknown (9). The *ADHD* variable states the general presence of ADHD, and the *ADD* variable identifies the presence of the inattentive subtype of ADHD. The *BIPOLAR*, *UNIPOLAR*, and *ANXIETY* variables specify the presence of Bipolar Disorder, Unipolar Depression and/or Anxiety disorder. Potential drug or alcohol addictions are recognized by the *SUBSTANCE* variable, and the presence of additional psychiatric disorders are stated by the *OTHER* variable. Of the 103 patients assessed with MINI Plus, 51 patients received an ADHD diagnosis, and 23 had the inattentive sub-type.

The Adult ADHD Self-Report Scale (ASRS) is a screening tool for evaluating current symptoms of ADHD. This 18-item scale assesses symptoms of impulsivity, hyperactivity, and inattention. The outputs are sum scores between 0 and 72, and a higher score means more severe symptoms [4].

The Wender Utah Rating Scale for Attention Deficit Hyperactivity Disorder (WURS) is a 25-item questionnaire retrospectively assessing the presence and severity of childhood ADHD symptoms. The scale outputs a sum score between 0 and 100, and higher scores indicate increased manifestation and severity of symptoms [4].

The Mood Disorder Questionnaire (MDQ) is a self-reported screening instrument for bipolar spectrum disorder, containing 13 yes/no questions regarding commonly observed hypomanic/manic symptoms in bipolar disorder, as well as two additional questions. The diagnostic criteria are that at least 7 of the 13 questions are answered yes, several of the symptoms have occurred at once and have caused personal problems. The bipolar spectrum diagnosis is more widely defined than bipolar I and II diagnoses, as it also includes subthreshold cases [17]. The MDQ-POS variable is reported similarly to the MINI Plus variables.

The Cyclothymic temperament scale (CT) is a 21-item self-rated scale, assessing the emotional instability and hypersensitivity to external stimuli typically associated with this bipolar spectrum disorder. CT is originally a part of the comprehensive TEMP-A questionnaire [34]. The variable is reported like the MINI Plus variables.

Montgomery and Asberg Depression Rating Scale (MADRS) is a 10-item clinician-rated instrument, which evaluates the severity of ongoing depression. The sum score (0-60) states the severity of depression, and scores below 10 are classified as the absence of depressive symptoms [15].

Hospital Anxiety and Depression Scale (HADS) is a patient-rated assessment tool for evaluating the severity of the current state of anxiety and depression. The scale consists of 14-items, where seven items rate the anxiety level (HADS-A), and seven rate the depressive state (HADS-D). Each dimension gives a sum score between 0 and 21, and scores below 8 indicate the absence of symptoms regarding the condition in question [33].

Motor activity was collected with a wrist-worn actigraph device (Actiwatch, Cambridge Neurotechnology Ltd, England, model AW4), containing a piezoelectric accelerometer programmed to record the integration of intensity, amount, and duration of movement in the x, y, and z-axes. The sampling frequency was 32 Hz and movements over 0.05 g were recorded. The output is an integer value proportional to the movement intensity for 1-minute epochs [10]. The dichotomous variable ACC identifies the participants who have recorded motor activity (N/Y), the ACC-TIME variable tells when the recordings were started (HH:MM), and ACC-DAYS gives the number of 24h cycles recorded for each participant. The 45 patients with ADHD recorded motor activity for 6.6 ± 1.3 days (mean $\pm$ standard derivation), and the 40 clinical controls recorded motor activity for 7.2 ± 0.9 days.

Heart rate was recorded naturalistically with the chest-worn ECG-based monitoring device Actiheart (Cambridge Neurotechnology Ltd, England) [3]. The time between beats (inter-beat interval (IBI)) is measured in milliseconds. This data is completely raw, without any correction or imputation. When faced with missing data, the Actiheart creates its own artifacts, expressed as several identical IBIs after each other. The time series should be evaluated for these artifacts and for the potential infrequent appearance of extrasystoles. After adequate quality control, the IBI data can be analyzed into HRV measures. The dichotomous variable HRV identifies the participants who have recorded heart rate (N/Y), the HRV-TIME variable tells when the recordings were started (HH:MM), and HRV-HOURS gives the approximate number of hours heart rate was recorded for each participant. A total of 80 participants provided heart rate recordings: 38 ADHD patients for an average of 20.5 ± 3.9 hours, and 42 clinical controls for 21 ± 4.3 hours.

Conner's Continuous Performance Test II (CPT-II) is a computerized neuropsychological test which evaluates impulsivity and sustained attention by response or nonresponse to various letters presented for 250 milliseconds on the PC screen. CPT-II contains 360 trials divided into six blocks, defined by the time interval between the presented letters [25]. CPT-II calculates and outputs various estimates, like omission and commission errors, as well as a clinical ADHD Confidence Index score (range 0–100). However, analyzing the raw data (the responses of the 360 trials) has previously been fruitful for distinguishing between patient groups in the current dataset [12]. In total, 49 clinical controls and 50 participants with ADHD completed the CPT-II test.

HYPERAKTIV is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), and is available for download at https://osf.io/3agwr.

5 Applications and Usage Scenarios

The purpose of publishing this dataset is two-fold. First, we want to make the field of mental health research more accessible for computer scientists that do not have access to private medical data. Second, there is a lot of multimedia research on medical applications where datasets are private, making the work neither directly applicable nor reproducible. We hope that by releasing this dataset, we open up a more transparent and collaborative community for mental health research. As a starting point, we foresee this dataset to have several applications and usage scenarios. A few examples are presented in the following.

- Predict whether a patient has ADHD or not by using the included activity data, heart rate variability, or a combination of the two.
- Use patient-related attributes to analyze associations between ADHD and other illnesses such as bipolar disorder.
- Use patient-related attributes and unsupervised techniques to gain new insight to potentially advance diagnosis and treatment of ADHD and related mental disorders.
- Analyze the heart rate data in context to ADHD.

In Section 7, we perform experiments using some of the aforementioned application scenarios.

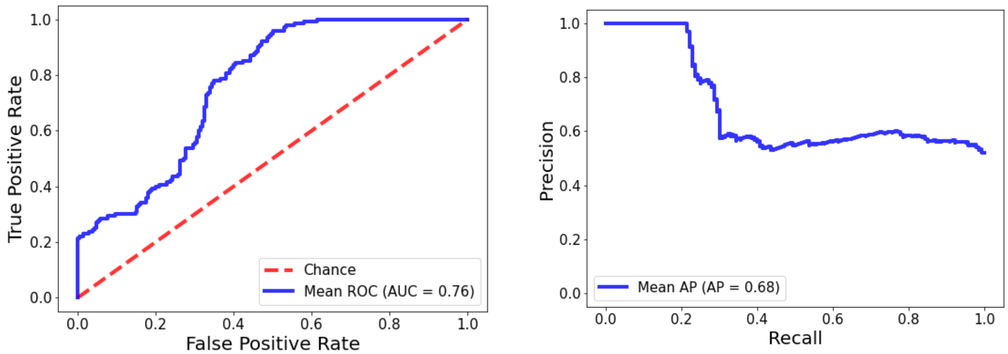
6 Suggested Metrics

As described in Section 5, HYPERAKTIV contains ground truth for several applications, which have different appropriate metrics depending on the task. For classification tasks, we recommend using standard classification metrics for either a binary or multi-class use case [16]. This includes metrics such as precision, recall, F1-score, and Matthews correlation coefficient (MCC). For a regression task, metrics such as mean absolute error or root mean squared error are more appropriate. Regardless of the application, multiple metrics should always be reported for a full evaluation of an algorithm.

7 Experiments

One of the applications mentioned in Section 5 was to predict ADHD based on the provided activity data. This section provides a set of baseline experiments to provide a starting point for researchers who want to get started with HYPERAKTIV. We split this up into two separate experiments. The first experiment predicts if a given patient has ADHD or not based solely on the activity data.

Features used for the experiments were extracted from the activity data using the Python library *tsfresh* [7] and included in the dataset (see Section 4). These features were reduced to only the relevant features using the function *select_features*, which is included in *tsfresh*. The data was split between a training and testing dataset



(a) Area under the ROC curve.

(b) Area under the PRC curve.

Figure 2: AUCROC and AUPRC plots for the best performing model.

Table 2: Baseline results for predicting ADHD on the on the test dataset averaged across 10 folds. The best performing model is highlighted in bold. The models above the midline are the simple baselines.

Model	Accuracy	Precision	Recall	F1-Score	MCC
RB	0.58	0.00	0.00	0.00	0.00
MIN	0.61	0.39	**1.00**	0.56	0.00
MAJ	0.58	0.00	0.00	0.00	0.00
LR	0.71	0.58	0.82	**0.68**	**0.46**
RF	**0.72**	**0.60**	0.76	0.67	0.44
XGB	0.71	0.61	0.67	0.63	0.40
LGBM	0.70	0.60	0.67	0.63	0.39

Table 3: The averaged confusion matrix calculated from the predictions of the 10-fold models that performed best on the testing dataset.

		True ADHD		
		Positive	Negative	Total
Predicted	Positive	14	7	14 + 7
	Negative	4	10	4 + 10
	Total	14 + 4	7 + 10	35

(80% for training and 20% for testing). We combine the clinical controls contained in our previous dataset Psykose [19] to compare against healthy controls. The test data was kept separate from the training process and only used for the final test. Each algorithm was trained on the training dataset using stratified 10-fold cross-validation leading to ten different models per algorithm (one per fold). After the training step, each model was then run on the testing dataset. The reported results are the average of the ten models per algorithm.

We tested four machine learning algorithms in addition to a series of simple prediction rules. The four machine learning methods include logistic regression, random forest, XGBoost [6], and LightGBM [20]. The logistic regression and random forest were implemented in scikit-learn [27], while XBoost and LightGBM were implemented in their respective official libraries. In addition to the aforementioned machine learning methods, we also ran a series of simple prediction rules on the test dataset. This is done to evaluate the effectiveness of the trained models. The simple rules include always predicting the majority class, always predicting the minority class, and random uniform prediction. Each rule was implemented using the dummy classifier included in scikit-learn. The code, dataset (with the used splits), and configurations used for all models to perform the experiments are made open-source and published on GitHub[1].

Looking at the results presented in Table 2, we see that all tested machine learning models beat the four baselines set for predicting ADHD using the activity data. Of the four models, the logistic regression method achieves the best MCC (0.46), which is approximately 0.02 more than the runner-up (random forests). Figure 2 and Table 3 show the AUCROC and AUPRC curves and confusion matrix for the logistic regression model. Overall, the baseline experiments reveal that there is potential in using machine learning with this data but that there is still space for improvement.

8 Conclusion

Open medical datasets are essential for reproducibility and comparability in a field that is well-known for its strict data access. In this paper, we presented HYPERAKTIV, an open dataset meant for ADHD research. The dataset contains activity and heart rate data collected from 103 participants (ADHD patients and clinical controls) together with a set of patient attributes. We discussed several applications of the dataset and suggested a series of metrics that should be used when evaluating machine learning methods trained on HYPERAKTIV. A series of baseline experiments were presented, showing that simple machine learning methods are able to predict ADHD, but with much room for improvement. We hope that this dataset will contribute to more open and collaborative work in the computer science community regarding mental health research.

Acknowledgements

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

[1]https://github.com/simula/hyperaktiv

References

[1] Amirmasoud Ahmadi, Mehrdad Kashefi, Hassan Shahrokhi, and Mohammad Ali Nazari. 2021. Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes. *Biomedical Signal Processing and Control* 63 (2021), 102227.

[2] Gail A Alvares, Daniel S Quintana, Ian B Hickie, and Adam J Guastella. 2016. Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: a systematic review and meta-analysis. *Journal of Psychiatry & Neuroscience* (2016).

[3] Søren Brage, Niels Brage, Paul W Franks, Ulf Ekelund, and Nicholas J Wareham. 2005. Reliability and validity of the combined heart rate and movement sensor Actiheart. *European journal of clinical nutrition* 59, 4 (2005), 561–570.

[4] Erlend Joramo Brevik, Astri J Lundervold, Jan Haavik, and Maj-Britt Posserud. 2020. Validity and accuracy of the Adult Attention-Deficit/Hyperactivity Disorder (ADHD) Self-Report Scale (ASRS) and the Wender Utah Rating Scale (WURS) symptom checklists in discriminating between adults with and without ADHD. *Brain and behavior* 10, 6 (2020), e01605.

[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), 21–28.

[6] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 10. https://doi.org/10.1145/2939672.2939785

[7] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W. Kempa-Liehr. 2018. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing* 307 (2018). https://doi.org/10.1016/j.neucom.2018.03.067

[8] C. Keith Conners and Gill Sitarenios. 2011. *Conners' Continuous Performance Test (CPT)*. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-79948-3_1535

[9] Gianni L Faedda, Kyoko Ohashi, Mariely Hernandez, Cynthia E McGreenery, Marie C Grant, Argelinda Baroni, Ann Polcari, and Martin H Teicher. 2016. Actigraph measures discriminate pediatric bipolar disorder from attention-deficit/hyperactivity disorder and typically developing controls. *Journal of Child Psychology and Psychiatry* 57, 6 (2016).

[10] Ole Bernt Fasmer, Erlend Eindride Fasmer, Kristin Mjeldheim, Wenche Førland, Vigdis Elin Giæver Syrstad, Petter Jakobsen, Jan Øystein Berle, Tone EG Henriksen, Zahra Sepasdar, Erik R Hauge, et al. 2020. Diurnal variation of motor activity in adult ADHD patients analyzed with methods from graph theory. *PloS one* 15, 11 (2020).

[11] Ole Bernt Fasmer, Kristin Mjeldheim, Wenche Førland, Anita L Hansen, Steven Dilsaver, Ketil J Oedegaard, and Jan Øystein Berle. 2015. Motor activity in adult patients with attention deficit hyperactivity disorder. *Psychiatry investigation* 12, 4 (2015), 474.

[12] Ole Bernt Fasmer, Kristin Mjeldheim, Wenche Førland, Anita L Hansen, Vigdis Elin Giæver Syrstad, Ketil J Oedegaard, and Jan Øystein Berle. 2016. Linear and non-linear analyses of Conner's Continuous Performance Test-II discriminate adult patients with attention deficit hyperactivity disorder from patients with mood and anxiety disorders. *BMC psychiatry* 16, 1 (2016).

[13] 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*.

[14] Enrique Garcia-Ceja, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J Oedegaard, and Jim Tørresen. 2018. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing* 51 (2018).

[15] CJ Hawley, TM Gale, T Sivakumaran, Hertfordshire Neuroscience Research group, et al. 2002. Defining remission by cut off score on the MADRS: selecting the optimal value. *Journal of affective disorders* 72, 2 (2002), 177–184.

[16] Steven A. Hicks, Inga Strümke, Vajira Thambawita, Malek Hammou, Michael A. Riegler, Pål Halvorsen, and Sravanthi Parasa. 2021. On evaluation metrics for medical applications of artificial intelligence. *medRxiv* (2021). https://doi.org/10.1101/2021.04.07.21254975

[17] Robert MA Hirschfeld, Janet BW Williams, Robert L Spitzer, Joseph R Calabrese, Laurie Flynn, Paul E Keck Jr, Lydia Lewis, Susan L McElroy, Robert M Post, Daniel J Rapport, et al. 2000. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *American journal of psychiatry* 157, 11 (2000), 1873–1875.

[18] Petter Jakobsen, Enrique Garcia-Ceja, Michael Riegler, Lena Antonsen Stabell, Tine Nordgreen, Jim Torresen, Ole Bernt Fasmer, and Ketil Joachim Oedegaard. 2020. Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. *Plos one* 15, 8 (2020), e0231995.

[19] Petter Jakobsen, Enrique Garcia-Ceja, Lena Antonsen Stabell, Ketil Joachim Oedegaard, Jan Oystein Berle, Vajira Thambawita, Steven Alexander Hicks, Pål Halvorsen, Ole Bernt Fasmer, and Michael Alexander Riegler. 2020. PSYKOSE: A Motor Activity Database of Patients with Schizophrenia. In *Proceedings of the IEEE International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE.

[20] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc.

[21] Julian Koenig, Joshua A Rash, Andrew H Kemp, Reiner Buchhorn, Julian F Thayer, and Michael Kaess. 2017. Resting state vagal tone in attention deficit (hyperactivity) disorder: A meta-analysis. *The World Journal of Biological Psychiatry* 18, 4 (2017), 256–267.

[22] Elisabet Kvadsheim, Ole Bernt Fasmer, Berge Osnes, Julian Koenig, Steinunn Adolfsdottir, Heike Eichele, Kerstin Jessica Plessen, and Lin Sørensen. 2020. Lower Cardiac Vagal Activity Predicts Self-Reported Difficulties With Emotion Regulation in Adolescents With ADHD. *Frontiers in psychiatry* 11 (2020), 244.

[23] Lourdes García Murillo, Samuele Cortese, David Anderson, Adriana Di Martino, and Francisco Xavier Castellanos. 2015. Locomotor activity measures in the diagnosis of attention deficit hyperactivity disorder: Meta-analyses and new findings. *Journal of neuroscience methods* 252 (2015).

[24] Kyoung-Sae Na, Seo-Eun Cho, and Seong-Jin Cho. 2021. Machine learning-based discrimination of panic disorder from other anxiety disorders. *Journal of Affective Disorders* 278 (2021), 1–4.

[25] Iban Onandia-Hinchado, Natividad Pardo-Palenzuela, and Unai Diaz-Orueta. 2021. Cognitive characterization of adult attention deficit hyperactivity disorder by domains: a systematic review. *Journal of Neural Transmission* (2021), 1–45.

[26] Niamh O'Mahony, Blanca Florentino-Liano, Juan J Carballo, Enrique Baca-García, and Antonio Artés Rodríguez. 2014. Objective diagnosis of ADHD using IMUs. *Medical engineering & physics* 36, 7 (2014), 922–926.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011).

[28] Andreea Robe, Anca Dobrean, Ioana A Cristea, Costina R Păsărelu, and Elena Predescu. 2019. Attention-deficit/hyperactivity disorder and task-related heart rate variability: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews* 99 (2019), 11–22.

[29] Kapil Sayal, Vibhore Prasad, David Daley, Tamsin Ford, and David Coghill. 2018. ADHD in children and young people: prevalence, care pathways, and service provision. *The Lancet Psychiatry* 5, 2 (2018), 175–186.

[30] 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.

[31] Fred Shaffer and JP Ginsberg. 2017. An overview of heart rate variability metrics and norms. *Frontiers in public health* 5 (2017), 258.

[32] David V Sheehan, Yves Lecrubier, K Harnett Sheehan, Patricia Amorim, Juris Janavs, Emmanuelle Weiller, Thierry Hergueta, Roxy Baker, Geoffrey C Dunbar, et al. 1998. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of clinical psychiatry* 59, 20 (1998), 22–33.

[33] R Philip Snaith. 2003. The hospital anxiety and depression scale. *Health and quality of life outcomes* 1, 1 (2003), 1–4.

[34] Vigdis Elin Giaever Syrstad, Ketil Joachim Oedegaard, Ole Bernt Fasmer, Anne Halmoy, Jan Haavik, Steven Dilsaver, and Rolf Gjestad. 2020. Cyclothymic temperament: Associations with ADHD, other psychopathology, and medical morbidity in the general population. *Journal of affective disorders* 260 (2020), 440–447.

[35] Nagahide Takahashi, Kanako Ishizuka, and Toshiya Inada. 2021. Peripheral Biomarkers of Attention-deficit Hyperactivity Disorder: Current Status and Future Perspective. *Journal of Psychiatric Research* (2021).

[36] Lorenzo Tonetti, Andreas Conca, Giancarlo Giupponi, Marco Filardi, and Vincenzo Natale. 2018. Circadian activity rhythm in adult attention-deficit hyperactivity disorder. *Journal of psychiatric research* 103 (2018), 1–4.

[37] Gaetano Valenza, Mimma Nardelli, Antonio Lanata, Claudio Gentili, Gilles Bertschy, Markus Kosel, and Enzo Pasquale Scilingo. 2016. Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics. *IEEE journal of biomedical and health informatics* 20, 4 (2016), 1034–1043.

[38] Nora D Volkow and James M Swanson. 2013. Adult attention deficit–hyperactivity disorder. *New England Journal of Medicine* 369, 20 (2013), 1935–1944.

[39] Zi Ying Wee, Samantha Wei Lee Yong, Qian Hui Chew, Cuntai Guan, Tih Shih Lee, and Kang Sim. 2019. Actigraphy studies and clinical and biobehavioural correlates in schizophrenia: a systematic review. *Journal of Neural Transmission* 126, 5 (2019), 531–558.

[40] Eva Charlotte Winnebeck, Dorothee Fischer, Tanya Leise, and Till Roenneberg. 2018. Dynamics and ultradian structure of human sleep in real life. *Current Biology* 28, 1 (2018), 49–59.