EPiC Series in Computing
Volume 101, 2024, Pages 49–57
Proceedings of the 16th International Conference on Bioinformatics and Computational Biology (BICOB-2024)

Machine Learning-Based Automated Detection of ADHD Using Heart Rate Variability Data

Yanqing Ji[1]*, Janet Zhang-Lea[2] and John Tran[3]
[1] Department of Electrical and Computer Engineering, Gonzaga University, Spokane, WA
[2] Department of Human Physiology, Gonzaga University, Spokane, WA
[3] Department of Psychiatry and Behavioral Science, University of California San Francisco Fresno, Fresno, CA.
ji@gonzaga.edu, lea@gonzaga.edu, jtran@fresnocountyca.gov

Abstract

This study addresses the pressing need for effective methods in detecting Attention-Deficit/Hyperactivity Disorder (ADHD), a neurodevelopmental condition significantly impacting individuals' attention, impulse control, and activity regulation. Leveraging advancements in machine learning and wearable technology, the research explores the potential of Heart Rate Variability (HRV) data as a novel source for ADHD detection. Six machine learning algorithms, including Logistic Regression, Random Forest, XGBoost, LightGBM, Neural Network, and Support Vector Machine, were rigorously investigated using an HRV dataset, marking a pioneering effort in utilizing HRV data for ADHD identification. The results demonstrate promising performance, with Logistic Regression exhibiting the highest F1 score (0.71), and Support Vector Machine achieving the highest Matthews Correlation Coefficient (0.44). This study showcases the capacity of machine learning utilizing HRV data for identifying ADHD, contributing to the evolving landscape of machine learning applications in mental health diagnostics.

1 Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder that affects individuals' ability to pay attention, control their impulses, and regulate their level of activity. The disorder has a significant impact on various aspects of a person's life, including academic and occupational performance, relationships, mental health, and overall well-being [1]. ADHD is a diverse, enduring condition that impacts 5.9 to 7.1% of school-aged children and almost 5% of the adult population [2]. ADHD in adults often manifests with a varied clinical presentation that goes beyond the fundamental motor symptoms observed in children. It encompasses a wider spectrum of emotional

* Corresponding Author

dysregulations and functional impairments [3] which necessitate effective ADHD detection methods to minimize the risk of mental health disorders, self-injury, and suicidal tendencies.

Presently, there are no definitive biomarkers or valid objective tests to identify ADHD [4], and to differentiate ADHD diagnosis from other mental disorders, including bipolar disorder [5] or anxiety [6]. ADHD diagnosis primarily relies on subjective clinical evaluation, observation of behavior, and reported symptoms. Due to the subjective nature of this assessment process, there exists a risk of both under and over-diagnosis. Consequently, there is a demand for more objective assessment approaches [7] [8].

Wearable sensor technology combined with machine learning methods has garnered considerable interest as a means to complement established subjective diagnostic procedures in the realm of mental health [9]. To identify or classify ADHD, researchers have utilized data collected via different types of wearable sensors and experimented with various machine learning algorithms. For example, deep convolutional neural networks have been utilized to classify the electroencephalography signal of healthy children from ADHD children with two subtypes of Combined ADHD (ADHD-C) and inattentive AHDH (ADHD-I) [10]. Kol et al. employed a bagged tree classifier to categorize ADHD, ADHD+CD (conduct disorder), and CD automatically on electrocardiography (ECG) signals [11].

Heart rate data emerges as a potential data source for investigating psychotic disorders. It can be utilized to compute heart rate variability (HRV), indicating the extent of variation in the time intervals between successive heartbeats. HRV can provide valuable insights into the autonomic nervous system and overall health. Researchers have found evidence of reduced HRV in neuropsychiatric conditions like depression and psychotic disorders [12] as well as ADHD [13]. However, there are very limited studies employing HRV for ADHD diagnosis in the literature. In a pilot study [14] on this subject, short-term HRV data were gathered from 20 children (10 with ADHD and 10 controls), aged 7-12 years. Statistical analyses, including Mann-Whitney and Wilcoxon tests, were employed to discern parameters distinguishing the ADHD and non-ADHD groups. Another study reported a correlation between lower HRV and poor emotional regulation in adolescents diagnosed with ADHD [15]. Notably, both studies did not incorporate machine learning algorithms, and no accuracy data were reported.

In this study, six different machine learning algorithms were examined for automatically distinguishing individuals with ADHD from other clinical controls through the analysis of heart rate data. Statistical features were derived from patients' HRV data, and a subset of these features was chosen using their associated p-values. The selected subset was then utilized by the six machine learning algorithms: Logistic Regression, Random Forest, XGBoost, LightGBM, Neural Network, and Support Vector Machine. The comparative analysis of these six classifiers indicated that Logistic Regression achieved the highest F1 score. To the best of our knowledge, this is the first study that utilized machine learning-based methodologies for identifying ADHD patients among other patients with different conditions.

2 Methodology

In this section, we present the proposed framework for detecting ADHD among clinical controls, introduce the dataset employed, and describe the chosen feature selection method.

2.1 ADHD Detection Framework

This section presents an outline of the proposed framework designed for identifying ADHD among clinical controls. Figure 1 provides a big picture of the proposed framework, demonstrating its utility in classifying patients into ADHD or non-ADHD categories based on the dataset. In the initial stage, data is collected from patients, including heart rate data recorded through wearable devices, and patient details encompassing background and medical history. Subsequently, statistical features are extracted using a Python library tsfresh [16] and then relevant features are chosen if their p-value is less than 0.05. The next step involves implementing, tuning, and testing various machine learning algorithms with 10-fold cross-validation. This facilitates the identification of an optimal algorithm based on diverse performance metrics, ensuring the accurate identification of ADHD patients.

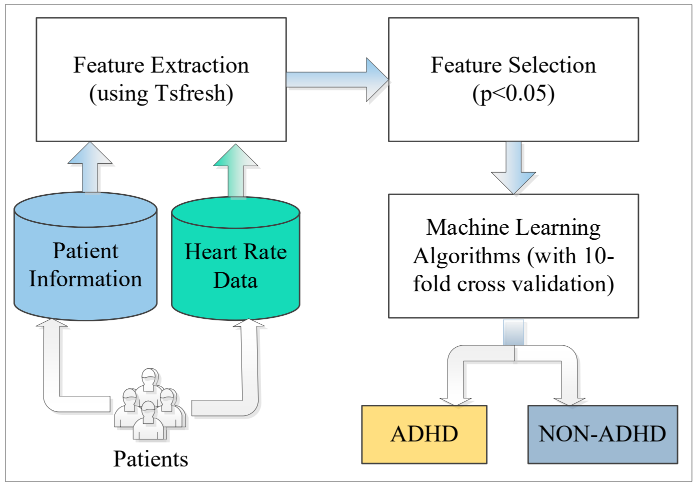

Figure 1: ADHD Detection Framework

2.2 Dataset

This current study utilizes a publicly available 'Hyperaktiv' dataset (accessible at https://osf.io/3agwr) [17]. This dataset includes patient information, activity, and heart rate data retrieved from 103 enrolled patients, with 51 diagnosed with ADHD and 52 with other clinical disorders. The heart rate data is ECG-based and recorded through a compact, chest-worn device powered by a battery. This setup enables unrestricted movement and facilitates extended recording periods. Figure 2 illustrates the HRV log of an ADHD-diagnosed patient over 24 hours. The figure indicates that this patient exhibits reduced HRV during the early morning and the later afternoon. The fluctuations in HRV over time could provide insights into diagnosing ADHD, given the time-dependent characteristics of ADHD hyperactivity. In the current dataset, 80 patients contributed HRV recordings. Among them, 38 individuals with ADHD recorded their HRV for an average duration of 20.5 ± 3.9 hours, while 42 clinical controls logged their HRV for an average of 21 ± 4 hours [17]. The recorded data is organized into one file per participant.

2.3 Feature Extraction and Selection

Utilizing HRV data, 788 statistical features were derived using the open-source Python package tsfresh [16]. This library is specifically crafted for the systematic extraction of pertinent features from time series data to facilitate machine learning tasks. The extracted features encompass a range of attributes, encompassing basic statistics like mean, median, standard deviation, skewness, kurtosis, as well as entropy measures, FFT (Fast Fourier Transform) features, temporal characteristics, and more.

The resultant feature data is stored in a features file, with each row encapsulating the feature data for an individual participant.

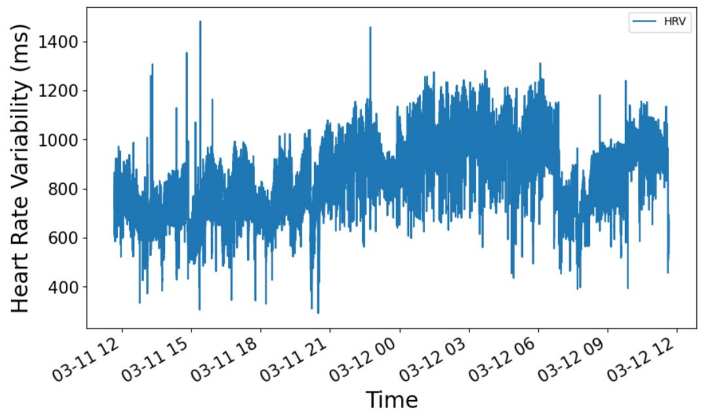

Figure 2: Example of 24 hours (from noon to noon) of HRV recordings from a patient diagnosed with ADHD.

As the number of extracted features is relatively large in comparison to the size of the HRV dataset (consisting of 80 patients), the training phase of a machine learning model may not be able to generalize effectively, potentially diminishing its predictive power unnecessarily. Therefore, a subset of the 788 extracted features was chosen based on their p-values. The threshold for selecting features based on their statistical significance is set at $p < 0.05$. This threshold is a conventional choice and stems from the field of statistics. As a result, 28 features have been selected for utilization in the machine learning algorithms.

2.4 Machine Learning Models and HyperParameters

To classify patients with ADHD and other clinical controls, six distinct machine learning algorithms—Logistic Regression (LG), Random Forest (RF), Extreme Gradient Boosting (XGB), LightGBM (LGBM), Neural Network (NN), and Support Vector Machine (SVM)—have been implemented and tested. The ultimate objective is to develop algorithms that achieve the highest classification accuracy. To enhance the performance of these models, it is crucial to identify and fine-tune essential hyperparameters to precisely tailor them to the provided data, ensuring the creation of accurate models. The chosen algorithms and their specific hyperparameters are described below.

1) *Logistic Regression (LG):* LG is a fundamental machine learning algorithm primarily utilized for binary classification tasks. It focuses on estimating the probability of an instance belonging to a specific class. Employing the logistic function, also known as the sigmoid function, it transforms the linear combination of input features into a range between 0 and 1. The algorithm establishes a decision boundary based on a threshold probability, facilitating the classification of instances. In this study, the model incorporates the L2 penalty, also known as Ridge regularization, to prevent overfitting and enhance generalization performance. The stopping criteria tolerance is set at 0.0001. The model employs a linear solver, and the class weights are set to "balanced" mode.

2) *Random Forest (RF):* RF is a powerful ensemble learning algorithm. This algorithm constructs multiple decision trees during training and predicts the output based on the majority votes of the trees. Each tree in the forest is built on a random subset of a dataset, introducing diversity and mitigating overfitting. RF excels in handling high-dimensional data, capturing complex relationships, and providing robust predictions. In this model, the number of trees is set to 1000, and the entropy function

is utilized to measure the quality of a split. The maximum depth and leaf nodes are both set to none. When selecting features for splitting, the model is set to use the square root of the total available features as the maximum number of features to choose from.

3) *Extreme Gradient Boosting (XGB):* XGB also utilizes an ensemble of decision trees. It sequentially builds trees to correct errors of the preceding ones, resulting in a robust and accurate predictive model. XGB is known for its speed, scalability, and ability to handle diverse datasets. Its incorporation of regularization techniques and advanced features such as tree pruning contributes to its success in achieving state-of-the-art results across various domains. In this work, the number of boosting rounds is set to 1000, and the "verbosity" is set to 0 for silent mode during training. The algorithm is tailored for binary classification using the logistic objective function, while the learning rate is fine-tuned to 0.01, controlling the step size during optimization.

4) *LightGBM (LGBM):* As a cutting-edge gradient-boosting framework, LGBM excels in handling large datasets and complex tasks. Its key innovation lies in the implementation of a histogram-based learning approach, where feature values are binned into discrete intervals, significantly accelerating the training process. LGBM supports classification tasks and offers flexibility in model customization through various hyperparameters. Known for its ability to handle imbalanced datasets, categorical features, and intricate relationships, LGBM has become a popular choice in real-world applications. Being a boosting algorithm, LGBM shares similarities in parameter configuration with XGB. The number of boosting rounds is specified as 1000, and a silent mode is activated for training. The algorithm is tailored for binary classification, and the learning rate is adjusted to 0.01, dictating the step size in the optimization process.

5) *Neural Network (NN):* Inspired by the structure and functioning of the human brain, NNs are characterized by interconnected nodes organized into layers. They can effectively model complex relationships in data for classification. The input layer receives features, and subsequent hidden layers process and learn hierarchical representations, culminating in the output layer's predictions. Training involves adjusting the network's weights through backpropagation, optimizing its ability to generalize from training data to unseen examples. NNs are capable of learning intricate patterns and non-linear relationships, making them instrumental in diverse applications. In this investigation, a compact NN was employed with two hidden layers, a design choice aligned with the limited size of the HRV dataset. The hidden layers consist of 25 and 15 neurons, respectively, utilizing the rectified linear unit function as the chosen activation function. For weight optimization, the 'adam' optimizer is employed, with the maximum training iterations capped at 1000. The initial learning rate is established at 0.001. To maintain result reproducibility, the random seed is set to 42.

6) *Support Vector Machine (SVM):* SVM is a powerful machine learning algorithm utilized for classification tasks. Its main objective is to find a hyperplane that maximizes the margin between different class data points. SVM accommodates both linear and non-linear decision boundaries using kernel functions. By mapping instances into a high-dimensional feature space, SVM determines the optimal hyperplane. Known for its effectiveness in high-dimensional spaces and handling complex datasets, SVM offers flexibility with various kernel functions, making it widely employed in various domains. This study utilizes the radial basis function as the kernel for this model, with a kernel degree set to three. The termination criteria tolerance is established at 0.001, and the shrinking heuristic is configured to true.

2.5 Performance Metrics

Five performance metrics, namely accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC), are employed to evaluate the performance of the studied machine learning algorithms. The definitions of these metrics rely on the confusion matrix, a tabular representation with four quadrants, each illustrating the outcomes of the classifier. These quadrants consist of True Positive

(TP), True Negative (TN), False Positive (FP), and False Negative (FN). The following are the definitions of these five metrics.

Accuracy is a performance metric that measures the overall correctness of a classification model. It is defined as the ratio of correctly predicted instances (both positive and negative) to the total number of instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision evaluates the accuracy of the positive predictions made by a classification model. It is defined as the ratio of correctly predicted positive instances (TP) to the total number of instances predicted as positive (sum of TP and FP).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall, also known as sensitivity or true positive rate, is a performance metric that measures the ability of a classification model to correctly identify all relevant instances of the positive class. It is the ratio of correctly predicted positive instances (TP) to the total number of actual positive instances (sum of TP and FN).

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1-score is the harmonic mean of precision and recall.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

MCC provides a balanced measure of the quality of a binary classification model, especially in cases of imbalanced class distribution. Its formula is as follows:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

3 Results

The HRV dataset was employed to train and evaluate the predictive capabilities of the six machine learning algorithms for ADHD. Following z-score normalization, the transformed data was split into training (70%) and testing (30%) sets, with the latter reserved for the final evaluation phase. Each algorithm underwent training on the stratified 10-fold cross-validated training dataset, resulting in ten distinct models per algorithm (one for each fold). Subsequently, these models were applied to the testing dataset, and the reported results represent the averages across the ten models for each algorithm. The evaluation encompasses five metrics—accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC). A comprehensive overview of the performance of the six classification algorithms is presented in Table I.

Model	Accuracy	Precision	Recall	F1-Score	MCC
LR	0.70	0.64	**0.79**	**0.71**	0.42
RF	0.68	0.65	0.63	0.64	0.35
XG	0.65	0.62	0.66	0.63	0.31
LGBM	0.66	0.62	0.66	0.63	0.32
NN	0.68	0.64	0.71	0.67	0.37
SVM	**0.72**	**0.70**	0.69	0.69	**0.44**

Table 1: Experiment Results for Predicting ADHD on the Test Dataset Averaged across 10 Folds

The experiment results in Table I demonstrate comparable performance between the LR and SVM methods in classifying ADHD and non-ADHD patients using the HRV dataset. Among the six models, the LR model achieves superior recall (0.79) and F1-score (0.71), while the SVM model excels in accuracy (0.72), precision (0.70), and MCC (0.44). Both LR and SVM outperform the remaining four models. Figure 3 displays the AUCROC and AUPRC curves for the SVM model. Overall, these experiments underscore the significant potential of utilizing machine learning with HRV data for ADHD detection.

4 Discussions

This study aimed to employ HRV data for the classification of ADHD and non-ADHD patients through machine learning algorithms. In related work, Hicks et al. utilized activity data from actigraphs for ADHD detection, reporting results for LR, RF, XGB, and LGBM models [17]. A comparison between the two sets of results reveals that HRV data yields comparable outcomes to activity data. For instance, the LR model achieves an F1-Score of 0.71 and an MCC of 0.42 with the HRV dataset, while

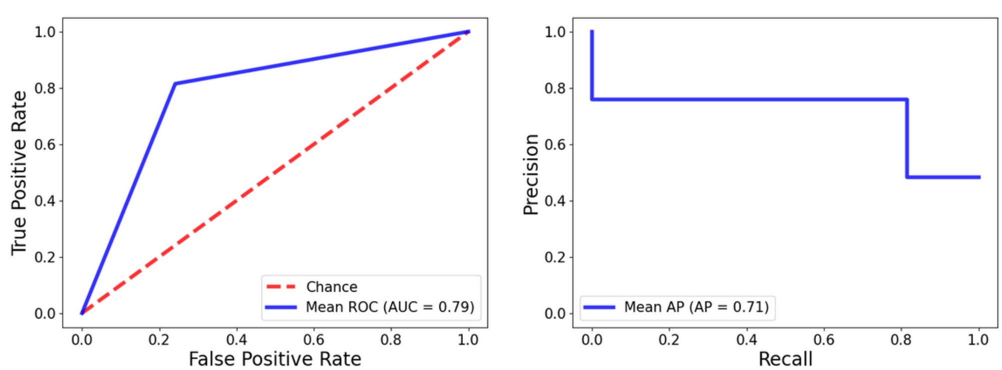

Figure 3: AUCROC AND AUPRC PLOTS FOR THE SVM MODEL.

it achieves 0.68 for F1-Score and 0.46 for MCC with the activity data. This study emphasizes HRV data as a valuable source for detecting ADHD using machine learning algorithms.

A limitation of this study lies in the relatively small size of the HRV dataset, comprising only 38 ADHD patients and 42 clinical controls. The efficacy of machine learning algorithms in automatic

ADHD detection could be substantially enhanced with a larger and more diverse dataset. Encouragingly, the research community's collaboration and willingness to share additional HRV data would significantly contribute to advancing the study and improving the robustness and generalizability of ADHD detection models based on machine learning algorithms.

In the realm of machine learning, the conventional threshold for feature selection based on statistical significance is commonly set at $p < 0.05$. This means that there is less than a 5% chance that the observed feature occurred solely by random chance. While this threshold is rooted in statistical practices, it is not an absolute rule. In our experimentation with the HRV dataset, we explored a more stringent threshold of $p < 0.01$, but it resulted in the selection of only 2 features, which proved insufficient for effective machine learning algorithms. On the other hand, a looser threshold of $p < 0.1$ selected 53 features. Although the LR model exhibited slightly improved performance, the SVM's performance declined with these 53 features. Therefore, setting the threshold at $p < 0.05$ seems to be appropriate for this study.

5 Conclusions

This study explores machine learning algorithms utilizing HRV data, an underexplored domain for ADHD diagnosis. Six algorithms—LR, RF, XGB, LGBM, NN, and SVM—are investigated, with LR and SVM emerging as the top performers. Despite the promising results, the study acknowledges limitations in dataset size and advocates for collaborative efforts to enhance model robustness. This pioneering work demonstrates the potential of HRV-based machine learning for ADHD identification and opens avenues for future research collaborations and improvements in diagnostic models.

References

[1] N. D. Volkow and J. M. Swanson, "Adult attention deficit–hyperactivity disorder," *New England Journal of Medicine,* vol. 369, no. 20, p. 1935–1944, 2013.

[2] E. G. Willcutt, "The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review," *Neurotherapeutics,* vol. 9, no. 3, pp. 490-499, 2012.

[3] P. Shaw, A. Stringaris, J. Nigg and E. Leibenluft, "Emotion dysregulation in attention deficit hyperactivity disorder," *American Journal of Psychiatry,* vol. 171, no. 3, pp. 276-293, 2014.

[4] W. Das and S. Khanna, "A Robust Machine Learning Based Framework for the Automated Detection of ADHD Using Pupillometric Biomarkers and Time Series Analysis," *Scientific Reports,* vol. 11, no. 1, p. 16370, 2021.

[5] M. J. Brus, M. V. Solanto and J. F. Goldberg, "Adult ADHD vs. Bipolar Disorder in the DSM-5 Era," *Journal of Psychiatric Practice,* vol. 20, no. 6, pp. 428-437, 2014.

[6] A. Koyuncu, T. Ayan, E. I. Guliyev, S. Erbilgin and E. Deveci, "ADHD and Anxiety Disorder Comorbidity in Children and Adults: Diagnostic and Therapeutic Challenges," *Current Psychiatry Reports,* vol. 24, pp. 129-140, 2022.

[7] L. G. Murillo, S. Cortese, D. Anderson, A. D. Martino and F. X. Castellanos, "Locomotor activity measures in the diagnosis of attention deficit hyperactivity disorder: Meta-analyses and new findings," *Journal of Neuroscience Methods,* vol. 252, pp. 14-26, 2015.

[8] N. Takahashi, K. Ishizuka and T. Inada, "Peripheral biomarkers of attention-deficit hyperactivity disorder: Current status and future perspective," *Journal of Psychiatric Research,* vol. 137, pp. 465-470, 2021.

[9] E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. J. Oedegaard and J. Tørresen, "Mental health monitoring with multimodal sensing and machine learning: A survey," *Pervasive and Mobile Computing,* vol. 51, pp. 1-26, 2018.

[10] A. Ahmadi, M. Kashefi, H. Shahrokhi and M. A. Nazari, "Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes," *Biomedical Signal Processing and Control,* vol. 63, 2021.

[11] J. E. Koh, C. P. Ooi, N. S. Lim-Ashworth, J. Vicnesh, H. T. Tor, O. S. Lih, R.-S. T. MBBS, U. Acharya and D. S. S. Fung, "Automated classification of attention deficit hyperactivity disorder and conduct disorder using entropy features with ECG signals," *Computers in Biology and Medicine,* vol. 140, p. 105–120, 2022.

[12] G. A. Alvares, D. S. Quintana, I. B. Hickie and A. J. Guastella, "Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: a systematic review and meta-analysis," *Journal of Psychiatry and Neuroscience,* vol. 41, no. 2, pp. 89-104, 2016.

[13] A. Robe, A. Dobrean, I. A. Cristea and C. R. Păsărelu, "Attention-deficit/hyperactivity disorder and task-related heart rate variability: A systematic review and meta-analysis," *Neuroscience & Biobehavioral Reviews,* vol. 99, pp. 11-22, 2019.

[14] M. R. Rukmani, S. P. Seshadri, K. Thennarasu, T. R. Raju and T. N. Sathyaprabha, "Heart Rate Variability in Children with Attention-Deficit/Hyperactivity Disorder: A Pilot Study," *Annals of Neurosciences,* vol. 23, pp. 81-88, 2016.

[15] E. Kvadsheim, O. B. Fasmer, B. Osnes, J. Koenig, S. Adolfsdottir, H. Eichele, K. J. Plessen and L. Sørensen, "Lower Cardiac Vagal Activity Predicts Self-Reported Difficulties With Emotion Regulation in Adolescents With ADHD," *Frontiers in Psychiatry,* vol. 17, no. 11, p. 244, 2020.

[16] M. Christ, N. Braun, J. Neuffer and A. W. Kempa-Liehr, "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," *Neurocomputing,* vol. 307, pp. 72-77, 2018.

[17] S. A. Hicks, A. Stautland, O. B. Fasmer, W. Førland, H. L. Hammer, P. Halvorsen, K. Mjeldheim, K. J. Oedegaard, B. Osnes, V. E. G. Syrstad, M. A. Riegler and P. Jakobsen, "HYPERAKTIV: An Activity Dataset from Patients with Attention-Deficit/Hyperactivity Disorder (ADHD)," in *Proceedings of the 12th ACM Multimedia Systems Conference*, Istanbul, Turkey, 2021.