A Preliminary Investigation into Quantitative Assessment of ADHD Treatment Efficacy on Hyperactivity Levels via Actigraphy*

Joshua Putris, Aybike Aydın, Mustafa Balkanas, Ayşe Elif Söylemezoğlu, Nihal Serdengeçti, Tayyib Kadak, Mahmut Cem Tarakçıoğlu, and Hakan Töreyin, *Member, IEEE*

***Abstract*— The diagnosis and assessment of attention deficit hyperactivity disorder (ADHD) in clinical practice heavily rely on subjective and biased scales. This study explores the feasibility of using actigraphy measurements to objectively assess and monitor the treatment response of ADHD in children receiving medication. A cohort of ten children underwent evaluation using three scales, one administered by child-adolescent psychiatrists and two completed by the parents, both pre- and post-medication. In addition, two sets of actigraphy recordings were collected, each spanning seven consecutive days, before and after medication administration. The study revealed that changes in the median, mean, and skewness of accelerations in spherical coordinates exhibited stronger correlations with changes in the scale scores in comparison to other features. Additionally, binary classifications using feature sets with top correlations and PCA features defining 95% variability showed better predictions for ADHD treatment response assessed by TURGAY DSM-IV-S (82%) and WFIRS-life skills (81.4%) scores compared to WFIRS-school behavior (63.9%) scores. These findings represent the first reported correlations between ADHD scales and a broad range of features. Additionally, they demonstrate the feasibility of using actigraphy data to predict ADHD treatment response for the first time.**

***Clinical Relevance*— The utilization of actigraphy for objective and reliable measurement of ADHD symptoms and functional impairment can serve as a valuable complement to subjective evaluations conducted by parents and clinicians, and thus aiding in determining an effective treatment plan and identifying priority areas for intervention.**

I. Introduction

Attention deficit hyperactivity disorder (ADHD) is a common neurodevelopmental issue affecting around 2% to 9% of children of school age in the world [1]. Inattention, increased motor activity, and impulsivity are the hallmarks of ADHD, which is characterized by dysregulated cognitive and behavioral processes [2].

ADHD is typically diagnosed through a combination of clinical evaluation, medical history review, symptom rating scales, and cognitive-neuropsychological tests [3]. In assessing therapy efficacy, various methods are employed, including the use of clinical assessment tools. One commonly used tool is the Clinical Global Impressions scale (CGI), which enables physicians to evaluate the severity of symptoms, changes over time, and response to treatment in psychiatric disorders. It plays a significant role in assessing treatment effectiveness [4]. Assessing functional impairment is a crucial factor in determining therapy efficacy, and scales such as the Weiss Functional Impairment Rating scale [5] can be employed to evaluate functioning. While effective, these methods may suffer from subjectivity and bias. In contrast, objective measures can potentially improve the outcome sensitivity in clinical diagnosis and treatment efficacy monitoring of ADHD [6]. On the other hand, there is not a recognized and objective method for assessing ADHD that is utilized in clinics.

This study aims to address the need for an objective ADHD assessment method in children by investigating the potential use of actigraphy-based acceleration measurements. The study focuses on monitoring changes in hyperactivity level and functional impairments in children with ADHD receiving medication.

Unlike previous studies that primarily use actigraphy for ADHD diagnosis [7]–[9], this work takes a longitudinal approach to evaluate medication efficacy using actigraphy. Specifically, the study examines the correlation between actigraphy data and scales measuring ADHD symptom severity and functional impairments.

There are four key differences between this study and previous research on actigraphy-based medication assessment. Firstly, it extends beyond sleep-related activities, as explored in previous studies [10]–[12], to encompass a full week of activities, as the scales used consider activities beyond sleep. Secondly, it collects seven consecutive days of data for each pre- and post-medication session (168 hours/session), providing a larger data volume compared to [13], and controlling for potential variations from routine activities. Thirdly, it explores a wide range of statistical features derived from actigraphy recordings, going beyond the commonly investigated variance [13] and intensity features [14]. Lastly, the study explores data-driven classification methods to predict the ADHD treatment response (i.e., improvement vs. worsening/stability) captured by the clinical scales. Notably, in this study, unlike other studies, in addition to the ADHD symptom severity completed by the parents, both the functional impairment assessed by the parents and the assessment of the general level of functioning by the clinician are included to increase the response to treatment indicators.

*Research supported in part by the San Diego State University Rapid Response Program.

J. Putris and H. Töreyin are with the San Diego State University, San Diego, CA 92182 USA (Corresponding author phone: 619-594-0795; e-mail: htoreyin@sdsu.edu).

A. Aydın, M. Balkanas, A. E. Söylemezoğlu, N. Serdengeçti, T. Kadak, and M. C. Tarakçıoğlu are with the Istanbul University – Cerrahpaşa, Cerrahpaşa Faculty of Medicine, İstanbul, 34147 Turkey).

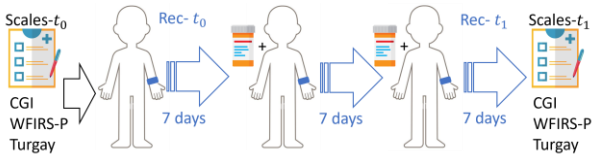

Figure 1. The data collection procedure encompasses initial assessments of subjects using three scales, namely CGI, WFIRS-P, and Turgay (Scales-t_0), followed by a seven-day recording period (Rec-t_0). Subsequently, the subjects commence medication, and during the second week of medication intake, another seven-day recording period (Rec-t_1) is conducted. At the conclusion of Rec-t_1, subsequent assessments using the same scales (Scales-t_1) are administered.

II. Methods

A. Participants, Clinical Assessment, and Data Collection

The compliance of this study with ethical standards was evaluated and approved by the Istanbul University-Cerrahpaşa Clinical Research Ethics Committee (IRB# 514022). A total of ten subjects (S1-to-S10, 3 girls, 7 boys, age: 8.6±2.6) were recruited in the study. Patients who presented to the child psychiatry outpatient clinic with complaints of hyperactivity and impulsivity and who were subsequently diagnosed with ADHD according to the fifth edition of the diagnostic and statistical manual of mental disorders criteria (DSM-5) based on clinical evaluation constituted the study's sample size of ten.

The data collection process is depicted in Figure 1. Pre-medication scale score assessments are referred to as Scales-t_0, while assessments at the end of the second week of medication (methylphenidate hydrochloride, osmotic-release oral system methylphenidate; no dosage increase through the study) are labeled as Scales-t_1. Child-adolescent psychiatrists used the CGI-Severity (CGI-S) score to evaluate the children's clinical functioning. Additionally, parents completed the Weiss functional impairment rating scale-parent report (WFIRS-P) to assess impaired functioning and the Turgay DSM-IV based child and adolescent behavior disorders screening and rating scale (Turgay DSM-IV-S) to assess the severity of hyperactivity and impulsivity. The actigraphy dataset comprises two recordings per subject, each spanning 7 consecutive days of continuous 3-axes acceleration data collected from the dominant wrist. One recording, labeled Rec-t_0, is obtained prior to medication initiation, while the second recording, labeled Rec-t_1, is obtained during the second week of medication. Both recordings automatically commence at 7 am on the day following the clinic visit. Throughout the measurements, subjects were instructed to keep the watches on at all times.

B. Actigraphy Hardware

The study utilizes two different watches, namely GENEActiv (Activinsights, Huntingdon, UK) and AX3 (Axivity Ltd., Newcastle upon Tyne, UK). Both watches can store raw 3-axes accelerometer data continuously for at least seven days, with a sampling rate of 20 Hz for GENEActiv and 25 Hz for AX3. GENEActiv is used for measurements from subjects S1 and S2, while AX3 is used for data from subjects S3 to S10. The equivalence and agreement between the two watches in measuring acceleration is demonstrated in [15], with an intraclass correlation coefficient of 0.95 (95% confidence interval: 0.87-0.98).

B. Data Analysis

All data analyses are conducted using MATLAB (MathWorks, Natick, MA). The overview of the analyses is summarized in Figure 2. The dataset is initially examined to identify any missing data or outliers. This examination includes visually inspecting the data for missing values and checking if any data points exceed the predefined dynamic ranges of ±8g, which serve as outlier thresholds. As a result, it is determined that all recordings in the dataset are free of missing data or outliers. However, the WFIRS-P scores for one subject (S10) were missing and are therefore excluded from analyses involving WFIRS-P.

Then, the accelerometer data captured in three-dimensional Cartesian coordinates (x, y, z) is converted into spherical coordinates (ρ, θ, ϕ) to provide a more natural representation of wrist movements. The data from one of the subjects transformed to the new coordinates is presented in Figure 2. For each spherical coordinate, a set of 15 features is computed, including maximum, minimum, mean (μ), standard deviation, median (η), mode, range, skewness (β), kurtosis, coefficient of variation, variance, interquartile range between upper (Q3) and lower (Q1) quartiles, entropy, standard error, and median absolute deviation; totaling 45 features.

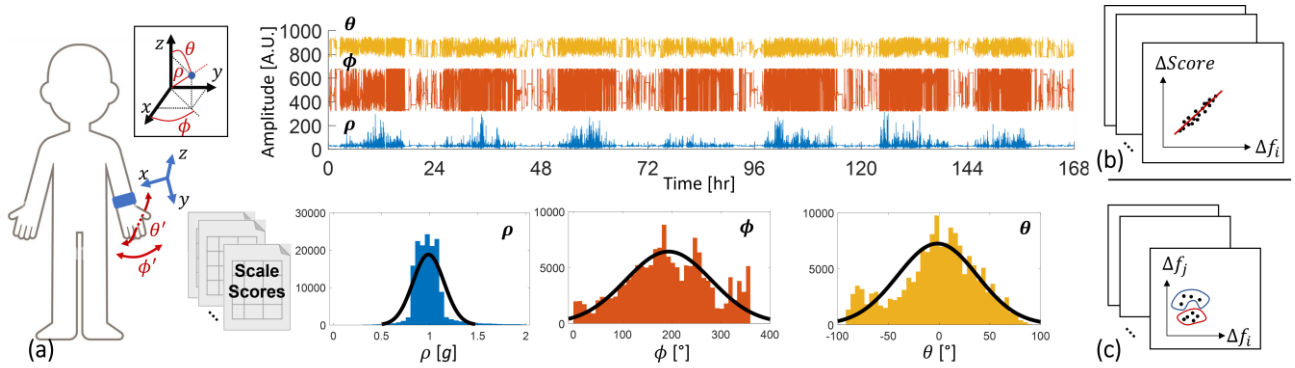

Figure 2. Data analysis steps. (a) The raw 3D acceleration data is converted into spherical coordinates. The ρ direction captures the amplitude of the accelerations, while the ϕ and θ directions capture mostly the mediolateral (ϕ') and anteroposterior (θ') swings of the wrist. From the recorded data and their distributions, a total of 45 features are extracted. (b) Correlations are calculated between the changes in these features, Δf_i, from Rec-t_0 to Rec-t_1 and the corresponding changes in scale scores, $\Delta Score$, from Scales-t_0 to Scales $- t_1$. (c) Subsets of features exhibiting the highest correlations and new features obtained through principal component analysis (PCA) are utilized to predict the ADHD treatment response as detected by the scales.

TABLE I. SCALE SCORE CHANGES FOLLOWING MEDICATION

	ΔC	*Cb*	*ΔW1*	*W1b*	*ΔW2*	*W2b*	*ΔT*	*Tb*
S1	-2	1	0	0	-0.2	1	-4	1
S2	-2	1	0	0	0.4	0	0	0
S3	-2	1	-0.99	1	-0.8	1	-9	1
S4	-2	1	-1.33	1	-0.8	1	-15	1
S5	-2	1	0	0	-0.1	1	-5	1
S6	-3	1	0.17	0	-0.1	1	10	0
S7	-3	1	-0.67	1	-0.9	1	-19	1
S8	-3	1	-0.33	1	-0.5	1	-9	1
S9	-2	1	0	0	0.2	0	-6	1
S10	-2	1	N/A	N/A	N/A	N/A	-12	1

Correlations between the *changes* in actigraphy feature values from Rec-t_0 to Rec-t_1 and the Scale-t_0 to Scale-t_1 *changes* in the CGI scale (*ΔC*), WFIRS-school subscale (*ΔW1*) assessing functional impairment in school-behavior domain, WFIRS-life subscale (*ΔW2*) assessing functional impairment in life-skills domain, and Turgay DSM-IV-S hyperactivity-impulsivity subscale (*ΔT*) are analyzed (Figure 2(b)). Pearson correlation coefficients (r) are computed for each feature-score pair. The features are ranked based on r for each scale. Subsequently, the feasibility of predicting the ADHD treatment response in patients, as clinically determined by *ΔC*, *ΔW1, ΔW2*, and *ΔT*, is explored using two new feature sets (Figure 2(c)). One set consists of the highest-ranked features, while the other set consists of principal components derived from the complete feature set, serving as predictors in the analysis.

III. RESULTS AND DISCUSSION

The scale score changes of ΔC, ΔW1, ΔW2, and ΔT are presented in Table I. Negative values indicate improvement, denoted by the binary value '1' in the corresponding columns Cb, W1b, W2b, and Tb. Positive values indicate a worsening, while zero represents stability, and both are indicated by the binary value '0'. All subjects show improvement based on ΔC scores. However, ΔW1, assessing school behavior changes, reflect worsening in five out of nine subjects. These score variations are expected because there may be assessment differences between the clinician and the parents.

A. Correlations Between Changes in Features and Scores

Table II summarizes the correlation coefficient analysis, which is conducted on two versions of the data: one considering the entire 168-hour recordings and another focusing on workday hours between 8 am and 5 pm over five workdays, totaling 45 hours. The resulting correlation coefficients from the weekday hours are denoted as *ΔC-wd, ΔT-wd, ΔW1-wd,* and *ΔW2-wd* in the table. The individually obtained correlation scores (r) between the features and the scale score changes range from -0.88 to 0.86. To rank the features across all four scales, a feature score (S) is calculated by taking the square root of the mean of the squared correlation coefficients of the feature. Table II presents the seven features that achieve S>0.5, identified as the best features, along with their individual r values for each scale and S values. Remarkably, median (η) emerges as the statistical feature that consistently ranks among the best features in all three spherical coordinate directions. The highest S score is obtained for the η_ρ. The other prominent features are μ_ϕ, ϕ_ρ, β_ϕ, and β_θ.

A further ranking is conducted to determine which scale exhibits the highest correlation with respect to these features.

TABLE II. PEARSON CORRELATION COEFFICIENTS (r) BETWEEN STATISTICAL FEATURES FROM ρ, θ, ϕ AND *ΔC, ΔW1, ΔW2, AND ΔT*

	ΔC-wd	*ΔC*	*ΔT-wd*	*ΔT*	*ΔW1-wd*	*ΔW1*	*ΔW2-wd*	*ΔW2*	S
η_ρ	-0.16	-0.02	-0.49	-0.78	-0.82	-0.83	-0.78	-0.75	**0.65**
μ_ϕ	0.57	0.12	0.20	0.62	0.48	0.86	0.61	0.80	**0.59**
μ_ρ	-0.17	0.05	-0.57	-0.70	-0.65	-0.65	-0.76	-0.60	**0.57**
η_ϕ	0.38	-0.02	0.17	0.60	0.57	0.85	0.54	0.77	**0.56**
η_θ	0.69	0.63	-0.45	-0.68	-0.55	-0.65	-0.17	-0.31	**0.55**
β_ϕ	-0.41	-0.10	-0.14	-0.63	-0.41	-0.88	-0.45	-0.78	**0.54**
β_θ	-0.55	-0.45	0.33	0.61	0.60	0.77	0.23	0.46	**0.52**
S	**0.46**	**0.3**	**0.37**	**0.66**	**0.6**	**0.79**	**0.55**	**0.66**	

η, μ, and β correspond to median, mean, and skewness, respectively. $|r|$ values greater than 0.65 are highlighted in red. S correspond to feature score.

TABLE III. ADHD TREATMENT RESPONSE PREDICTION

Scale	Length/Feature	Classifier	TP	TN	FP	FN	Accuracy (%)
Tb	*All/Top*	RUSBoost	8	2	0	0	100
		SVM	8	1	1	0	90
	All/PCA	RUSBoost	8	0	2	0	80
		LR	8	0	2	0	80
	Wd/Top	LR	8	1	1	0	90
		SVM	8	1	1	0	90
	Wd/PCA	LR	7	0	2	1	70
		SVM	7	0	2	1	70
W1b	*All/Top*	SVM	2	5	0	2	77.8
		LR	3	4	1	1	55.6
	All/PCA	RUSBoost	2	4	1	2	66.7
		SVM	1	4	1	3	55.6
	Wd/Top	SVM	3	3	2	1	66.7
		RUSBoost	1	5	0	3	66.7
	Wd/PCA	RUSBoost	3	3	2	1	66.7
		SVM	1	3	2	3	44.4
W2b	*All/Top*	SVM	7	0	2	0	77.8
		RUSBoost	5	1	1	2	66.7
	All/PCA	LR	7	1	1	0	88.9
		SVM	7	0	2	0	77.8
	Wd/Top	SVM	7	1	1	0	88.9
		RUSBoost	5	2	0	2	77.8
	Wd/PCA	LR	7	2	0	0	100
		SVM	7	1	1	0	88.9

The corresponding S values are presented in the bottom row of Table II. *ΔW1* demonstrates the strongest correlation, followed by *ΔW2* and *ΔT*. Notably, with respect to *ΔT*, *ΔW1*, and *ΔW2*, which are obtained based on subscales completed by parents, the correlations derived from the complete recordings exhibit higher scores in comparison to the 45-hour segments. Given the restricted evaluation timeframe of working parents, limited to non-working hours, this outcome aligns with expectations.

B. ADHD Treatment Response Prediction

The recordings are assessed to determine their predictability of ADHD treatment response with medication. The binary labels used in Table I to represent the scale score assessments are utilized as the two classes for prediction. It is important to note that the dataset comprises a limited number of subjects with 45 features, necessitating feature reduction to prevent overfitting. Two feature sets are employed for prediction purposes. The first set comprises the seven highest-ranked features as indicated in Table II. The second feature set is generated by applying principal component analysis (PCA) to the complete set of 45 features, selecting the principal components that account for 95% of the variance. Each feature set is obtained using two versions of the recordings: one utilizing the complete recordings and the other considering the 45-hour segments of work hours. This results in a total of four feature vectors per scale change prediction. It is worth noting

that the prediction is performed only for *W1b, W2b*, and *Tb*. The *Cb* is excluded due to the absence of the '0' class representing worsening or stability. The performance of three classifiers is compared using the MATLAB Classification Learner toolbox. In this feasibility study, three classifiers are employed. Logistic regression (LR) and support vector machine with a linear kernel (SVM) are known to be suitable for handling small datasets as in this study, while random under-sampling boost (RUSBoost) is particularly advantageous in the context of imbalanced class distributions such as in *W2b* and *Tb*.

The results of ADHD treatment response prediction are presented in Table III, including the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The accuracy percentage, calculated as $Accuracy = 100 * (TP + TN)/(N)$, is also provided, where N represents the number of subjects with N=10 for *Tb* and N=9 for *W1b* and *W2b*. The results of the top two classifiers for each prediction task are reported. The SVM classifier achieves the highest average accuracy across all predictions (74.8%), followed by RUSBoost (71.2%), and LR (67.4%). However, one-way analysis of variance (ANOVA) indicates that there is no significant difference among the prediction performances of the three models ($p = 0.52$). While the average accuracy obtained using the top seven features (83.5%) is slightly higher than that of the PCA feature set (78.7%), the difference is not significant ($p = 0.35$). On the other hand, when considering only *Tb* and *W1b* predictions, the difference between the average accuracies obtained using the top seven features is larger than PCA (83.6% vs. 70.9%, $p < 0.05$).

The classifiers demonstrated similar average accuracies for *Tb* (LR: 80%, SVM: 82%, RUSBoost: 77.5%) and *W2b* (LR: 75%, SVM: 81.4%, RUSBoost: 72.3%), which were higher than those obtained for *W1b* (LR: 47.2%, SVM: 61.1%, RUSBoost: 63.9%). For LR and SVM classifiers, these differences were found to be statistically significant ($p < 0.05$). These results are counterintuitive considering the higher S score corresponding to the *ΔW1* column in Table II. However, it is important to note that the correlation analysis in Table II used continuous-valued vectors for feature and score changes. Binary versions of score changes may not exhibit strong correlations with the same features. Future work will include a biserial correlation coefficient analysis to further investigate this.

IV. Conclusion

This study presents an analysis of the correlations between changes in acceleration measurements obtained from a wrist-worn watch and three scales, one evaluated by a clinician and two by parents, to assess the medication response in children with ADHD. Notably, this study is the first to explore the feasibility of predicting the treatment response from acceleration measurements. The findings, based on an initial study of ten children, reveal that mean, median, and skewness features derived from 3D acceleration measurements on spherical coordinates exhibit stronger correlations compared to other features. Furthermore, the results obtained from three binary classifiers demonstrate that the acceleration data can capture the ADHD treatment response in patients as assessed by the TURGAY DSM-IV-S and the WFIRS-life skills subscale scores with accuracies ranging from 72.3% to 82%. Future studies could involve a larger subject pool and longer observation periods to enhance generalizability and investigate predicting the magnitude as well as the direction of ADHD treatment response.

References

[1] A. Singh, C. J. Yeh, N. Verma, and A. K. Das, "Overview of Attention Deficit Hyperactivity Disorder in Young Children," *Health Psychol Res*, vol. 3, no. 2, p. 2115, Apr. 2015, doi: 10.4081/hpr.2015.2115.

[2] Jensen, P.S., "Clinical considerations for the diagnosis and treatment of ADHD in the managed care setting," *Am J Manag Care*, vol. 15, no. 5 Suppl, pp. S129–S140, 2009.

[3] Turgay A., "Turgay's DSM-IV based ADHD and disruptive behaviour disorders screening scale." Integrative Therapy Institute Publication,Toronto-Ontario, Canada, 1997.

[4] J. Busner and S. D. Targum, "The Clinical Global Impressions Scale," *Psychiatry (Edgmont)*, vol. 4, no. 7, pp. 28–37, Jul. 2007.

[5] M. Weiss and C. Murray, "Assessment and management of attention-deficit hyperactivity disorder in adults," *CMAJ*, vol. 168, no. 6, pp. 715–722, Mar. 2003.

[6] C. R. Sumner, V. S. Haynes, M. H. Teicher, and J. H. Newcorn, "Does Placebo Response Differ between Objective and Subjective Measures in Children with Attention-Deficit/Hyperactivity Disorder?," *Postgraduate Medicine*, vol. 122, no. 5, pp. 52–61, Sep. 2010, doi: 10.3810/pgm.2010.09.2201.

[7] M. L. Wolraich *et al.*, "Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents," *Pediatrics*, vol. 144, no. 4, p. e20192528, Oct. 2019, doi: 10.1542/peds.2019-2528.

[8] F. De Crescenzo *et al.*, "The use of actigraphy in the monitoring of methylphenidate versus placebo in ADHD: a meta-analysis," *ADHD Atten Def Hyp Disord*, vol. 6, no. 1, pp. 49–58, Mar. 2014, doi: 10.1007/s12402-013-0122-x.

[9] M. Muñoz-Organero, L. Powell, B. Heller, V. Harpin, and J. Parker, "Automatic Extraction and Detection of Characteristic Movement Patterns in Children with ADHD Based on a Convolutional Neural Network (CNN) and Acceleration Images," *Sensors*, vol. 18, no. 11, Art. no. 11, Nov. 2018, doi: 10.3390/s18113924.

[10] S. Ironside, F. Davidson, and P. Corkum, "Circadian motor activity affected by stimulant medication in children with attention-deficit/hyperactivity disorder," *Journal of Sleep Research*, vol. 19, no. 4, pp. 546–551, 2010, doi: 10.1111/j.1365-2869.2010.00845.x.

[11] M. Sanabra, T. Gómez, and J. A. Alda-Díez, "[The effects of pharmacological treatment with stimulants on circadian activity patterns in children with attention deficit hyperactivity disorder]," *Rev Neurol*, vol. 71, no. 12, pp. 438–446, Dec. 2020, doi: 10.33588/rn.7112.2020351.

[12] M. M. Solleveld *et al.*, "Effects of 16 Weeks of Methylphenidate Treatment on Actigraph-Assessed Sleep Measures in Medication-Naive Children With ADHD," *Frontiers in Psychiatry*, vol. 11, 2020, Accessed: May 20, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2020.00082

[13] C.-S. Ouyang, R.-C. Yang, C.-T. Chiang, R.-C. Wu, and L.-C. Lin, "Objective Evaluation of Therapeutic Effects of ADHD Medication Using a Smart Watch: A Pilot Study," *Applied Sciences*, vol. 10, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/app10175946.

[14] K. Konrad, T. Günther, M. Heinzel-Gutenbrunner, and B. Herpertz-Dahlmann, "Clinical Evaluation of Subjective and Objective Changes in Motor Activity and Attention in Children with Attention-Deficit/Hyperactivity Disorder in a Double-Blind Methylphenidate Trial," *Journal of Child and Adolescent Psychopharmacology*, vol. 15, no. 2, pp. 180–190, Apr. 2005, doi: 10.1089/cap.2005.15.180.

[15] A. V. Rowlands *et al.*, "Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent?," *Medicine & Science in Sports & Exercise*, vol. 50, no. 2, p. 257, Feb. 2018, doi: 10.1249/MSS.0000000000001435.