A Cross-Validation Approach for Classifying Physical Activity Intensity : A Case Study in Children with Attention Deficit/Hyperactivity Disorder

Ibrahim Traore
University of Champollion - ISIS Castres
Toulouse INP, IRIT laboratory, University of Toulouse, France
0009-0002-8587-6054

Imen Megdiche
University of Champollion - ISIS Castres
IRIT laboratory, University of Toulouse, France
0000-0002-1331-8662

Jérôme Marquet-Doléac
Health Faculty
University Toulouse III - Paul Sabatier, France
0009-0002-6578-2309

Lotfi Chaari
Toulouse INP, IRIT laboratory, University of Toulouse, France
0000-0002-3590-0370

***Abstract*—To assess physical activity intensity using raw acceleration data, various thresholds have been proposed based on different metrics, making it challenging to select the appropriate threshold, particularly when trying to find a threshold adapted to a similar study in terms of device type, population, and device placement. In this study, a cross-validation method is proposed that does not take into account the specific details of the data recording device, the placement of the device, and the characteristics of the population being recorded. We collected acceleration data from 18 healthy children and 18 children with attention deficit/hyperactivity disorder (ADHD) in normal living conditions using the Actigraph GT9X. After collection, the raw data underwent processing steps such as preprocessing, segmentation, and metric extraction. Subsequently, five intensity thresholds were applied to this data, and a voting method using an aggregation approach was used to combine the classifications from each threshold to obtain a final classification. The findings indicated that 97.2% of the voting classifications are reliable (total and approximate decisions) for children with ADHD (97.4% for healthy children), while 2.8% (2.6% for healthy children) should be considered with caution. In conclusion, our approach is flexible and adaptable to different devices and population groups, making it a valuable tool for assessing physical activity in various research contexts.**

***Index Terms*—Accelerometry, Activity Metrics, Cut-points, Physical activity, Hyperactivity, Attention Deficit Hyperactivity Disorder**

I. Introduction

In recent years, technological advancements in the field of sensors have enabled the development of wearable devices, capable of recording movements over extended periods. These devices have been used to study various conditions, including Attention Deficit Hyperactivity Disorder (ADHD).

ADHD is a neurodevelopmental disorder that affects 3 to 5% of children worldwide [1]. This disorder is characterized by inattention, impulsivity, and hyperactivity, which significantly interfere with daily life activities.

To study the motor overflow of children with ADHD, it is necessary to quantify their movements in everyday life conditions. However, classifying the intensity of movements from data collected by wearable devices can be challenging. This classification involves translating accelerometer data into categories of physical activity intensity: sedentary, light, moderate, and vigorous.

In a previous study [2], a smartwatch was employed to objectively detect detailed movements in children with ADHD using gyroscope and accelerometer measurements in a school environment.

The work of [3] suggests using thresholds on accelerometer data calculated as Counts Per Minute (CPM). The challenge lies in generalizing this approach to various accelerometer devices [3], [4].

This article proposes a cross-validation approach to be applied to raw triaxial accelerometer data recorded in everyday life conditions. We apply this approach in the case of children with ADHD. The approach is based on the concept of "wisdom of the crowd," which is the idea that large groups of people are collectively smarter than individual experts when it comes to problem-solving, decision-making, innovating, and predicting. This concept is widely used in various fields, including computer science, to enhance result accuracy.

Our approach stands out for its high flexibility and adaptability to various devices and population groups. It can be used with different types of data collection devices, such as actigraphs placed on various parts of the body or modern wearable devices like smartwatches. Furthermore, it can be tailored to different populations, whether they are adults, adolescents, children, or specific groups such as individuals

TABLE I
ACCELEROMETER RAW DATA CUT POINTS IN CHILDREN (MG).

Source	Device	Age	Metric	Cut-Points	Accuracy / Sensitivity (%)
Hildebrand 2014, 2016 (ENMO_ACT)	ActiGraph GT3X+ Non-dominant wrist	7-11 yr	ENMO	Light: 35.6 Moderate: 201.4 Vigorous: 707.0	Sedentary : 96 Light: 96 Moderate: 50 Vigorous: 79
Hildebrand 2014, 2016 (ENMO_GENE)	GENEActiv Non-dominant wrist	7-11 yr	ENMO	Light: 56.3 Moderate: 191.6 Vigorous: 695.8	Sedentary : 96 Light: 96 Moderate: 64 Vigorous: 79
Phillips 2013 (ENMOa_LW)	GENEA Left wrist	8-14 yr	ENMOa	Light: 87.5 Moderate: 250 Vigorous: 750	Sedentary : 94.7 Light: N/A Moderate: 88.1 Vigorous: 91.3
Phillips 2013 (ENMOa_RW)	GENEA Right wrist	8-14 yr	ENMOa	Light: 75 Moderate: 275 Vigorous: 700	Sedentary : 94.8 Light: N/A Moderate: 82.4 Vigorous: 89.3
Schaefer 2014 (BFEN)	GENEActiv Non-dominant wrist	6-11 yr	BFEN	Light: 190 Moderate: 314 Vigorous: 998	Sedentary : 83.3 Light: 27.6 Moderate: 41 Vigorous: 88.7

with certain medical conditions. This versatility and broad applicability make our approach a valuable tool for assessing physical activity in various research contexts.

The rest of this article is organized as follows. Section 2 provides a literature review. The proposed method is then elaborated upon in Section 3. The results of our research, our analysis, and an interpretation of these results are presented in Section 4. The conclusion section summarizes the main points discussed and presents some perspectives.

II. LITERATURE REVIEW

Usually, accelerometer manufacturers provided proprietary measurements called ”activity counts” (AC), which were widely used to assess the volume or intensity of physical activity, as well as to predict energy expenditure [5]–[10]. At that time, these measurements were the only outputs available for accelerometers used in research. However, recently, high-resolution raw accelerometer data has become accessible on various devices, including the ActiGraph GT9X accelerometer. Researchers have started to adopt new analytical approaches to directly exploit raw data rather than relying on software provided by the manufacturer [11]–[18] .

In our literature search, we identified two systematic reviews [3], [19] that compiled articles presenting physical activity thresholds based on raw accelerometer data and activity metrics [20]–[23]. Furthermore, we identified an R package (GGIR) [24] designed for processing raw accelerometer data, which provides a list of thresholds along with instructions for their use with GGIR. By cross-referencing information from these sources, we selected articles offering specific thresholds for children as shown in Table I.

Hildebrand et al. [25], [26] conducted a comparative study using the ActiGraph GT3X+ (AG) and the GENEActiv (GA) placed on the hip and wrist of children and adults. They observed a significant difference in acceleration values between placements on the hip and wrist. In their study, they developed intensity thresholds for physical activity in children and adults using the Euclidean Norm Minus One (ENMO) metric. ENMO is calculated by taking the Euclidean norm (vector magnitude) of the three raw acceleration signals minus 1.

Phillips et al. [27] established intensity thresholds using the GENEA and calibrated them against oxygen consumption (VO2). They defined thresholds for classifying activities into sedentary, light, moderate, and vigorous intensities, depending on whether the GENEA was worn on the wrist or hip. These thresholds were defined through values of Euclidean Norm Minus One with Amplitude (ENMOa). ENMOa is a variation of ENMO that takes into account the amplitude of acceleration signals. To calculate ENMOa, the absolute value of acceleration for each axis (x, y, and z) was taken, 1g was subtracted, and the sum of these absolute values was computed.

Schaefer et al. [28] also established intensity thresholds for physical activity using the GENEActiv placed on the wrist in primary school children. For threshold definition, they used the Butterworth-filtered Euclidean Norm Minus One (BFEN) metric, which is a filtered version of ENMO. It uses a Butterworth with a frequency range of 0.2 to 15 Hz to attenuate high frequencies not related to physical activity.

We observed that the thresholds commonly proposed in the literature are not effective in classifying all levels of physical activity intensity. It is important to note that the thresholds do not make the same classification errors. For example, the thresholds proposed by Hildebrand et al. showed better classification performance for sedentary/light and vigorous activities, while their performance was lower for moderate activities. Similarly, the thresholds established by Phillips et al. exhibited lower sensitivity and specificity values for moderate intensity compared to other intensities. We therefore propose an approach based on the principle of wisdom of the crowd as shown in Fig. 1. The approach enhances the accuracy, efficiency, and reliability of classification results by incorporating the contributions of multiple thresholds.

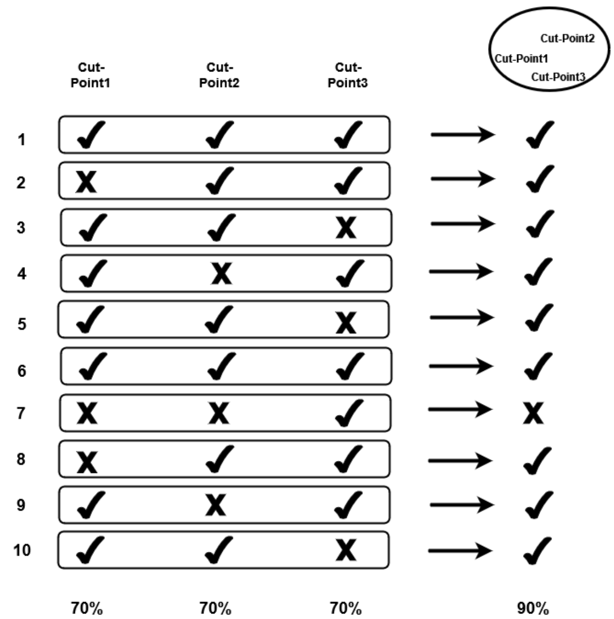

Fig. 1. Wisdom of the crowd concept on a set of cut-points.

III. Our approach

This section outlines the different steps of our approach for the development of a voting system as depicted in Fig. 2. These steps include data acquisition, signal preprocessing, signal segmentation, metrics values extraction/computation, and classification. Various data science techniques are introduced to enhance the classification outcome.

A. Data acquisition

For data acquisition, a sample of 18 children (15 boys and 3 girls) diagnosed with ADHD and 18 healthy children (15 boys and 3 girls) was assembled from 6.5 to 11.5 years (mean=8.92 (±1.42)). Boys are more commonly affected by ADHD than girls [29]. The cohort of selected patients depended on the census collection. The study's objectives were carefully explained to all participants, and consent was obtained from their legal guardians before their participation.

We used a commercial tri-axial accelerometer, the ActiGraph GT9X monitor, which is widely available on the market. This accelerometer has a dynamic range of ±8 g (where g represents the gravitational force of Earth) and offers a sampling frequency range from 30 to 100 Hz. To initialize the ActiGraph GT9X monitor, we connected it to a computer and configured a sampling frequency of 30 Hz with a recording duration of one minute, using ActiLife V6 software.

The data collected by the triaxial accelerometer consists of numerical values representing acceleration along the three axes (x, y, and z). These recordings were conducted in everyday life conditions, where children were encouraged to wear the ActiGraph GT9X on their non-dominant wrist for one week. Each day during that week, the children logged their activities in a journal with the assistance of their parents. Subsequently, raw accelerometer data was exported from the ActiLife software in .gt3x format. During the one-week recording period, we collected between 17 and 18 million lines of data per child.

B. Data Preprocessing

Data collected by accelerometers during daily life can be divided into time intervals when the device is worn and time intervals when it is not. Time intervals when the device is not worn include periods of sleep, showering, and aquatic activities.

It is crucial to make a precise distinction between the time intervals when the device is worn and those when it is not, as the wearing duration serves as the basis for assessing time spent at different levels of physical activity intensity. However, this differentiation can be challenging because constant readings of zero can occur for various reasons, such as removing the accelerometer during some activities (like aquatic activities) or for no specific reason during sleep or while sitting without movement for extended periods.

In our study, we assessed non-wear time of the accelerometer using the standard deviation and the range of values for each accelerometer axis, calculated over consecutive 30-minute segments. For a segment to be considered non-wear time, the standard deviation had to be less than 3.0 mg for at least two of the three axes, or the range of values had to be less than 50 mg for at least two of the three axes [30].

Data corresponding to non-wear periods were excluded from the overall dataset for each participant.

As shown in Fig. 2, in this step our row dataset has the x,y,z accelerations data as features and is indexed by time seconds.

C. Signal Segmentation

In order to extract useful information from the raw data, it is important to divide data into smaller segments using a non-overlapping sliding window approach. This approach involves using a fixed-size window that moves along data. By reducing the window size, activities can be detected more quickly with low computational cost. However, it is possible that the window may not contain the complete cycle of the performed activity.

On the other hand, increasing the window size allows for the detection of more complex activities but comes with a higher computational cost. Therefore, it is not recommended to increase the window size for real-time applications. According to the literature, the most commonly used windows in activity

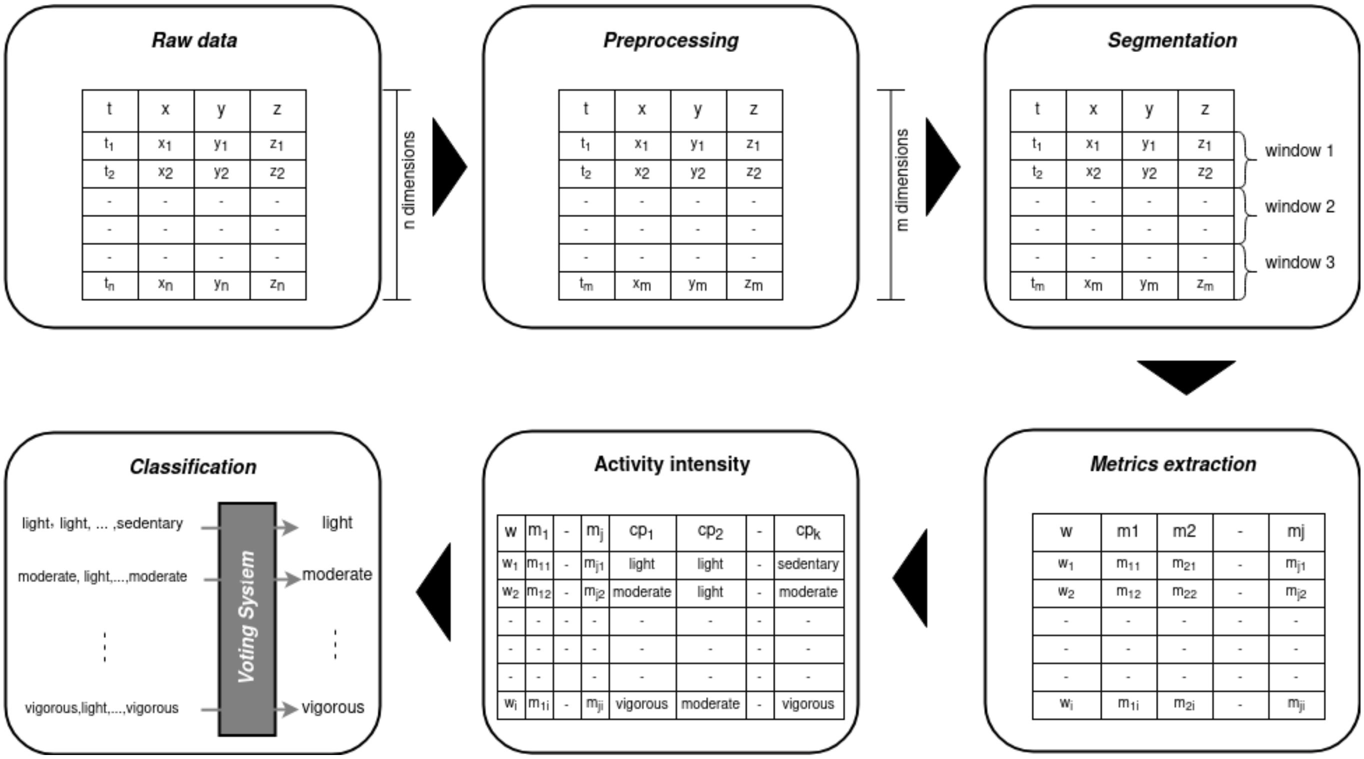

Fig. 2. The proposed pipeline in our approach

recognition systems are of small size. A window of 5 to 15 seconds is considered sufficient for recognizing an activity [31].

In our study, we divided data into 5-second windows. Since the accelerometer was initialized at a frequency of 30 Hz (30 samples per second), the number of samples in a window is 150 (30 Hz x 5 seconds).

D. Metrics extraction

The measured acceleration signal can be converted into an activity value using an activity metric for fixed-length windows. The activity metric defines the method by which activity values are calculated from the preprocessed acceleration signal. There are several metrics that can be applied to a dataset.

In our study, we transformed the extensive raw data into a reduced set of metric values. The purpose of this processing step was to extract metric values for each window, accurately representing the flow of raw data. To achieve this, we extracted the three metrics mentioned in the literature review and summarized them as follow :

$$ENMO = \frac{1}{n}\sum_{i=1}^{n}\max(r_i - 1, 0)$$

, where r_1, r_2,…,r_n are the n magnitude of acceleration values of the given epoch;

$$ENMOa = \sum|\sqrt{x^2+y^2+z^2} - g|$$

, where x, y and z are accelerations in each axis;

$$BFEN = \sum_{i=1}^{f}|\sqrt{x^2+y^2+z^2}|/(f)$$

, where f is sampling frequency; x, y and z are accelerations in each axis.

Let $M = \{m_1, m_2, \ldots, m_j\}$ be a set of metrics, and $W = \{w_1, w_2, \ldots w_i\}$ a set of windows. For each m_i and w_j, we can define a value m_{ij} that represents the value of metric m_i for window w_j. The files from the ActiGraph GT9X (in .gt3x format) were exported into the statistical software R, and the GGIR package was used to perform metric calculations. This allowed us to obtain precise measurements of motion-related acceleration levels.

E. Activity Intensity

Research studies have established specific threshold values for body acceleration measured in milligravity units (mg). For example, physical activity with body acceleration above a threshold in mg can be classified as moderate, while activity with body acceleration below this threshold can be considered light. This approach allows to classify the intensity of activity based on accelerometer data, corresponding to the level of physical effort exerted by an individual through their movements.

Thus, the activity data we calculated from the metrics were categorized into different levels of physical activity intensity, namely sedentary, light, moderate, or vigorous, using cut-points provided by previous studies (see Table I). The goal is to obtain a classification of physical activity intensities based on each cut-point used.

Let $CP = \{cp_1, cp_2, \ldots, cp_k\}$ be a set of cut-points, $M = \{m_1, m_2, \ldots, m_j\}$ a set of metrics, and $W = \{w_1, w_2, \ldots, w_i\}$ a set of windows. For each cp_i and each w_j, we can define a result cp_{ij} ('sedentary', 'light', 'moderate', 'vigorous') that represents the outcome of applying cut-point cp_i to the value of a metric calculated for window w_j.

F. Classification with voting system

Due to the difficulty of thresholds (or cut-points) in efficiently identifying all physical activity intensities, we propose an approach based on a voting system, utilizing the idea of "Wisdom of the Crowd". Our approach involves using multiple independent classification thresholds to generate a voted classification on a dataset. Then, by employing an aggregation method like majority voting, the classifications from each threshold can be combined to obtain a final classification (exp. Fig.1).

The fundamental idea behind this system is that individual errors can offset each other when aggregated. Therefore, the collective response obtained tends to be more accurate, unbiased, and closer to the truth than the opinion of a single classification threshold.

The voting system we propose (see Algorithm 1) involves collecting the most frequent classifications from each previously used classification threshold, while taking into account their classification performance.

IV. Evaluation

A. Evaluation Methodologies

We used three different methodologies to evaluate the reliability of the voting system by analyzing the voting outcomes.

- The first approach involves obtaining information on voting agreements. This entails retrieving the number of studies that provide the same intensity classification as the voting system result for each window. It helps identify challenging cases by examining instances where there is low agreement among the votes. The analysis of voting agreements helps detect these problematic cases that require additional attention.
- The second approach is to study the approximation of intensity (see Algorithm 2) classifications of cut-points that do not match the voting result. Since the value ranges for each intensity differ (e.g., sedentary from 0 to 35.6, light from 35.6 to 201.4), we identified the position of the measured value relative to percentiles. This approach allows us to determine, for non-unanimous votes, whether the classifications (different from the voting result) of the cut-points are approximate or not compared to the voting result. For this purpose, we defined three decision classes: total, approximate, and non-approximate. These decisions provide insights into the consistency of non-unanimous votes, which is crucial for evaluating the quality of the voting system result. The analysis of approximation provides additional information to better understand cases

Algorithm 1: Algorithm of Voting System

```
Input  : A matrix containing the cut points classifications for each window
Output: A matrix containing the voting result for each window
for row in matrix rows do
    frequent_values ← Compute the value counts of each element in the row;
    most_frequent_values ← Find elements with the maximum count in frequent_values;
    if length of most_frequent_values is 1 then
        most_frequent_value ← The only element in most_frequent_values;
    end
    else
        if 'sedentary' is in most_frequent_values and any element in
          ['ENMO_ACT','ENMO_GENE', 'ENMOa_RW','ENMOa_LW'] is 'sedentary' in the row then
            most_frequent_value ← 'sedentary';
        end
        else if 'light' is in most_frequent_values and any element in
          ['ENMO_ACT','ENMO_GENE'] is 'light' in the row then
            most_frequent_value ← 'light';
        end
        else if 'moderate' is in most_frequent_values and any element in
          ['ENMOa_LW','ENMOa_RW'] is 'moderate' in the row then
            most_frequent_value ← 'moderate';
        end
        else if 'vigorous' is in most_frequent_values and any element in
          ['ENMOa_LW','ENMOa_RW','BFEN'] is 'vigorous' in the row then
            most_frequent_value ← 'vigorous';
        end
        else
            most_frequent_value ← The first element in most_frequent_values;
        end
    end
    return most_frequent_value;
end
```

of low agreement among votes identified by the first approach.
- The third approach involves calculating the distance between cut-points classifications and the voting result. By measuring the classification distance, we can quantify the similarity or divergence of decisions. Since we propose using a set of cut-points in this study, the calculation of classification distance helps identify and select the most consistent cut-points for better results.

These approaches are applied separately to the data of the groups of healthy children and those with ADHD to observe the results between these different populations.

B. Results

The analysis of voting agreements (see Fig. 3) indicates that for children with ADHD, 54.4% (53.4% in healthy children) of the voting results were unanimous (5 out of 5), 33.5% (32.6% in healthy children) were obtained by 4 out of 5 votes, 11.2% (13.2% in healthy children) were obtained by 3 out of 5 votes, and only 0.9% (0.8% in healthy children) were obtained by 2 out of 5 votes. Regarding the alignment of decisions with respect to the voting system (see Fig. 4), it is observed that for children with ADHD, 54.4% (53.4% in healthy children) of the decisions are in total agreement, 42.8% (43.9% in healthy children) are approximate, and only 2.8% (2.7% in healthy children) are in disagreement. According to the results of distance calculations (see Table II), the cut-points "ENMO_GENE" (0.93-0.95), "ENMOa_LW" (0.94-0.95), and "ENMOa_RW" (0.96-0.96) show a high proportion of similarity with the voting system result in both cases. As for BFEN and ENMO_ACT, their similarity proportion is lower than the other three but still significant (0.66-0.69 and 0.84-0.86) for both groups of children. Specifically, the cut-point "ENMO_ACT" stands out by having a perfect similarity (100%) with the voting system for vigorous intensities. The cut-points "ENMO_GENE" and "ENMOa_LW" show high similarities (99%) with the voting system for vigorous intensities, following "ENMO_ACT" in both types of populations. The cut-point "ENMOa_RW" distinguishes itself with significant similarity (99%) with the voting system for moderate intensities.

C. Discussion

The diversity of intensity threshold sets for children poses challenges in selecting an appropriate threshold due to various data collection-related parameters. Although many studies have been conducted to calibrate threshold values using raw accelerometer data, there are no widely recognized threshold values for assessing physical activity intensity in children with ADHD. This is evident from the variety of threshold sets used in the literature. This situation creates significant challenges for researchers because the choice of "optimal" threshold values often relies on individual researchers' decisions or the adoption of already established thresholds in use. Furthermore, researchers have also developed their own sets of threshold values for their studies based on their judgment of previously published threshold values. This decision-making process appears more arbitrary than scientific, making it significantly challenging to estimate and synthesize moderate to high-intensity physical activity levels among children across different studies.

Algorithm 2: Algorithm of Vote Decision Reliability

```
Input  : A matrix containing the cut points
         classifications and voting results for each
         window
Output: A matrix containing the decision voting result
         for each window
for row in matrix rows do
    freq_intensity ← Most Frequent intensity value
      in the row;
    other_values ← List of values other than the
      frequent intensity in the row;
    decision ← Initialize to empty;
    if length of other_values is 0 then
        decision ← 'Total';
    end
    else
        if freq_intensity is 'sedentary' and more
          than half of other_values are 'light' with
          values in the first 50 centiles then
            decision ← 'Approximate';
        end
        else if freq_intensity is 'light' and more than
          half of other_values are 'sedentary' with
          values in last 50 centiles or 'moderate' with
          values in first 50 centiles then
            decision ← 'Approximate';
        end
        else if freq_intensity is 'moderate' and more
          than half of other_values are 'light' with
          values in last 50 centiles or 'vigorous' with
          values in first 50 centiles then
            decision ← 'Approximate';
        end
        else if freq_intensity is 'vigorous' and more
          than half of other_values are 'moderate'
          with values in last 50 centiles then
            decision ← 'Approximate';
        end
        else
            decision ← 'Non-Approximate';
        end
    end
    return decision;
end
```

We have developed a voting-based system using a set of threshold values to classify physical activity intensity from raw accelerometer data. This approach does not take into account device-specific, placement-specific, or population-specific fac-

TABLE II
CLASSIFICATION SIMILARITY WITH THE VOTING SYSTEM OUTCOME.

Intensity		ENMO_ACT	ENMO_GENE	ENMOa_LW	ENMOa_RW	BFEN
Non-ADHD	global	0.84	0.93	0.94	0.96	0.69
	sedentary	0.996	0.952	0.937	0.991	0.957
	light	0.65	0.9	0.979	0.907	0.55
	moderate	0.984	0.965	0.923	0.999	0.25
	vigorous	1	0.999	0.999	0.908	0.65
ADHD	global	0.86	0.95	0.95	0.96	0.66
	sedentary	0.998	0.95	0.947	0.994	0.981
	light	0.7	0.96	0.991	0.918	0.5
	moderate	0.99	0.97	0.92	1	0.24
	vigorous	1	0.999	0.999	0.907	0.61

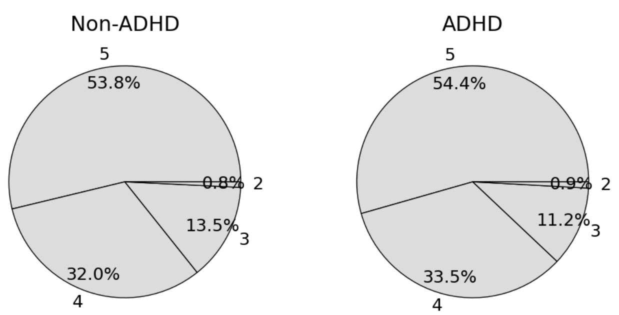

Fig. 3. Vote Distribution.

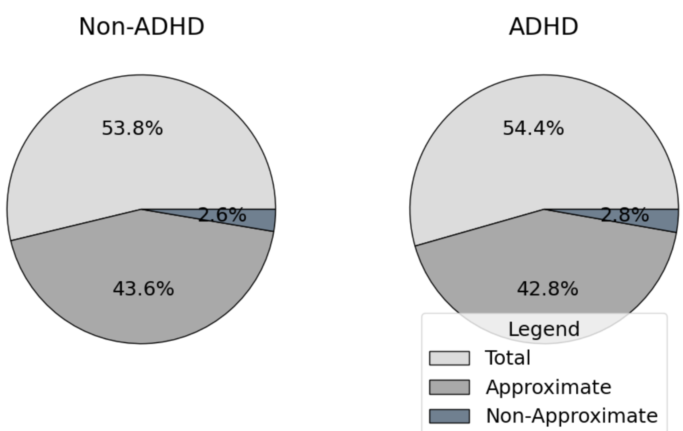

Fig. 4. Vote Decision Reliability.

tors. For example, to apply our approach to data recorded with an actigraph placed on the hip in adults, one can simply use predefined thresholds for adults, such as those reported in the literature for raw data recorded using a hip-worn accelerometer.

It is worth mentioning that some studies use other metrics, such as Mean Amplitude Deviation (MAD), to develop physical activity intensity thresholds. Our approach is not specifically dependent on any particular metric, but we recommend its use with metrics derived from raw acceleration data rather than summary measures like Counts per Minute (CPM). It is also important to note that some activity metrics may require data filtering at the preprocessing stage before their use.

For our study, we selected five widely used sets of threshold values for raw accelerometer data with the aim of contributing to consensus on their specific application to children with ADHD.

Regarding the classification of vigorous physical activity, the ENMO_ACT, ENMO_GENE, and ENMOa_LW threshold sets showed significantly better classification similarity with the voting system result than the ENMOa_RW and BFEN threshold sets. These classifiers appear to accurately identify periods when subjects were engaged in intense physical activities. The ENMOa_RW classifier has a high similarity proportion for moderate activities compared to the voting system result, indicating improved performance in detecting activities requiring moderate physical effort.

In contrast, the BFEN classifier exhibits the greatest discrepancy with the voting system result for all intensities. This divergence can be attributed to the data preprocessing technique used in BFEN, as it is the only classifier among the five to employ a metric that applies a bandpass filter to the data.

V. CONCLUSION

To the best of our knowledge, this study is the first one applying a cross-threshold approach to multiple sets of raw accelerometer data to estimate physical activity intensity in children with ADHD.

Recent versions of accelerometers provide raw acceleration data, theoretically allowing for comparisons across brands. However, no approach has been defined to optimize the selection of thresholds set on raw triaxial acceleration data for children with ADHD.

Overall, the results support the application of the voting system on threshold sets as 97.2% and 97.4% of the voting system results are reliable (total and approximate decisions) for children with and without ADHD, respectively. These findings reveal a similar classification between participants in healthy and ADHD groups. This suggests that the approach proposed in this study works effectively for both groups.

Based on these results, we plan to build an artificial intelligence model [32] to manage motor overflow in children with ADHD in a school environment and assess its applicability to a broader population in our future research.

REFERENCES

[1] G. V. Polanczyk, E. G. Willcutt, G. A. Salum, C. Kieling, and L. A. Rohde, "Adhd prevalence estimates across three decades: an updated systematic review and meta-regression analysis," *International journal of epidemiology*, vol. 43, no. 2, pp. 434–442, 2014.

[2] L.-C. Lin, C.-S. Ouyang, C.-T. Chiang, R.-C. Wu, and R.-C. Yang, "Quantitative analysis of movements in children with attention-deficit hyperactivity disorder using a smart watch at school," *Applied Sciences*, vol. 10, no. 12, p. 4116, 2020.

[3] Y. Kim, M. W. Beets, and G. J. Welk, "Everything you wanted to know about selecting the "right" actigraph accelerometer cut-points for youth, but…: a systematic review," *Journal of Science and Medicine in Sport*, vol. 15, no. 4, pp. 311–321, 2012.

[4] M. de Almeida Mendes, I. C. da Silva, V. V. Ramires, F. F. Reichert, R. C. Martins, and E. Tomasi, "Calibration of raw accelerometer data to measure physical activity: a systematic review," *Gait & posture*, vol. 61, pp. 98–110, 2018.

[5] D. R. Bassett Jr, A. V. Rowlands, and S. G. Trost, "Calibration and validation of wearable monitors," *Medicine and science in sports and exercise*, vol. 44, no. 1 Suppl 1, p. S32, 2012.

[6] S. J. Strath, K. A. Pfeiffer, and M. C. Whitt-Glover, "Accelerometer use with children, older adults, and adults with functional limitations," *Medicine and science in sports and exercise*, vol. 44, no. 1 Suppl 1, p. S77, 2012.

[7] R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Masse, T. Tilert, M. McDowell *et al.*, "Physical activity in the united states measured by accelerometer," *Medicine and science in sports and exercise*, vol. 40, no. 1, p. 181, 2008.

[8] S. G. Trost, R. R. Pate, J. F. Sallis, P. S. Freedson, W. C. Taylor, M. Dowda, and J. Sirard, "Age and gender differences in objectively measured physical activity in youth," *Medicine & science in sports & exercise*, vol. 34, no. 2, pp. 350–355, 2002.

[9] K. R. Evenson, D. J. Catellier, K. Gill, K. S. Ondrak, and R. G. McMurray, "Calibration of two objective measures of physical activity for children," *Journal of sports sciences*, vol. 26, no. 14, pp. 1557–1565, 2008.

[10] P. Freedson, D. Pober, and K. F. Janz, "Calibration of accelerometer output for children," *Medicine & Science in Sports & Exercise*, vol. 37, no. 11, pp. S523–S530, 2005.

[11] J. Bai, C. Di, L. Xiao, K. R. Evenson, A. Z. LaCroix, C. M. Crainiceanu, and D. M. Buchner, "An activity index for raw accelerometry data and its comparison with other activity metrics," *PloS one*, vol. 11, no. 8, p. e0160644, 2016.

[12] M. Aittasalo, H. Vähä-Ypyä, T. Vasankari, P. Husu, A.-M. Jussila, and H. Sievänen, "Mean amplitude deviation calculated from raw acceleration data: a novel method for classifying the intensity of adolescents' physical activity irrespective of accelerometer brand," *BMC sports science, medicine and rehabilitation*, vol. 7, pp. 1–7, 2015.

[13] C. V. Bouten, K. R. Westerterp, M. Verduin, and J. D. Janssen, "Assessment of energy expenditure for physical activity using a triaxial accelerometer." *Medicine and science in sports and exercise*, vol. 26, no. 12, pp. 1516–1523, 1994.

[14] M. P. Rothney, M. Neumann, A. Béziat, and K. Y. Chen, "An artificial neural network model of energy expenditure using nonintegrated acceleration signals," *Journal of applied physiology*, vol. 103, no. 4, pp. 1419–1427, 2007.

[15] D. Eslinger, A. V. Rowlands, T. L. Hurst, M. Catt, P. Murray, and R. G. Eston, "Validation of the genea accelerometer," 2011.

[16] I. C. Gyllensten and A. G. Bonomi, "Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life," *IEEE transactions on biomedical engineering*, vol. 58, no. 9, pp. 2656–2663, 2011.

[17] S. Zhang, A. V. Rowlands, P. Murray, T. L. Hurst *et al.*, "Physical activity classification using the genea wrist-worn accelerometer," 2012.

[18] M. Brandes, V. H. VT, V. Hannoever, and S. Brage, "Estimating energy expenditure from raw accelerometry in three types of locomotion." *Medicine and science in sports and exercise*, vol. 44, no. 11, pp. 2235–2242, 2012.

[19] V. Farrahi, M. Niemelä, M. Kangas, R. Korpelainen, and T. Jämsä, "Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches," *Gait & posture*, vol. 68, pp. 285–299, 2019.

[20] J. H. Migueles, A. V. Rowlands, F. Huber, S. Sabia, and V. T. van Hees, "Ggir: a research community–driven open source r package for generating physical activity and sleep outcomes from multi-day raw accelerometer data," *Journal for the Measurement of Physical Behaviour*, vol. 2, no. 3, pp. 188–196, 2019.

[21] K. Bakrania, T. Yates, A. V. Rowlands, D. W. Esliger, S. Bunnewell, J. Sanders, M. Davies, K. Khunti, and C. L. Edwardson, "Intensity thresholds on raw acceleration data: Euclidean norm minus one (enmo) and mean amplitude deviation (mad) approaches," *PloS one*, vol. 11, no. 10, p. e0164045, 2016.

[22] V. T. Van Hees, L. Gorzelniak, E. C. Dean León, M. Eder, M. Pias, S. Taherian, U. Ekelund, F. Renström, P. W. Franks, A. Horsch *et al.*, "Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity," *PloS one*, vol. 8, no. 4, p. e61691, 2013.

[23] V. T. Van Hees, Z. Fang, J. Langford, F. Assah, A. Mohammad, I. C. Da Silva, M. I. Trenell, T. White, N. J. Wareham, and S. Brage, "Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents," *Journal of applied physiology*, vol. 117, no. 7, pp. 738–744, 2014.

[24] B. Maczák, G. Vadai, A. Dér, I. Szendi, and Z. Gingl, "Detailed analysis and comparison of different activity metrics," *Plos one*, vol. 16, no. 12, p. e0261718, 2021.

[25] M. Hildebrand, V. H. VT, B. H. Hansen, and U. Ekelund, "Age group comparability of raw accelerometer output from wrist-and hip-worn monitors." *Medicine and science in sports and exercise*, vol. 46, no. 9, pp. 1816–1824, 2014.

[26] M. Hildebrand, B. H. Hansen, V. T. van Hees, and U. Ekelund, "Evaluation of raw acceleration sedentary thresholds in children and adults," *Scandinavian journal of medicine & science in sports*, vol. 27, no. 12, pp. 1814–1823, 2017.

[27] L. R. Phillips, G. Parfitt, and A. V. Rowlands, "Calibration of the genea accelerometer for assessment of physical activity intensity in children," *Journal of science and medicine in sport*, vol. 16, no. 2, pp. 124–128, 2013.

[28] C. A. Schaefer, C. R. Nigg, J. O. Hill, L. A. Brink, and R. C. Browning, "Establishing and evaluating wrist cutpoints for the geneactiv accelerometer in youth," *Medicine and science in sports and exercise*, vol. 46, no. 4, p. 826, 2014.

[29] S. Dalsgaard, E. Thorsteinsson, B. B. Trabjerg, J. Schullehner, O. Plana-Ripoll, I. Brikell, T. Wimberley, M. Thygesen, K. B. Madsen, A. Timmerman *et al.*, "Incidence rates and cumulative incidences of the full spectrum of diagnosed mental disorders in childhood and adolescence," *JAMA psychiatry*, vol. 77, no. 2, pp. 155–164, 2020.

[30] V. T. van Hees, F. Renström, A. Wright, A. Gradmark, M. Catt, K. Y. Chen, M. Löf, L. Bluck, J. Pomeroy, N. J. Wareham *et al.*, "Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer," *PloS one*, vol. 6, no. 7, p. e22922, 2011.

[31] V. Farrahi, M. Niemelä, M. Kangas, R. Korpelainen, and T. Jämsä, "Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches," *Gait & posture*, vol. 68, pp. 285–299, 2019.

[32] M. Fakhfakh, B. Bouaziz, F. Gargouri, and L. Chaari, "Prognet: Covid-19 prognosis using recurrent and convolutional neural networks," *medRxiv*, pp. 2020–05, 2020.