Data-Driven Techniques for Estimating Energy Expenditure in Wheelchair Users

Roya Doshmanziari, Håkon Strand Aandahl, Håvard Pettersen Reierstad, Marius Lyng Danielsson, Julia Kathrin Baumgart, and Damiano Varagnolo

***Abstract*— Providing feedback on energy expenditure (EE) may be an important tool to support obesity prevention among manual wheelchair users (MWU). This paper presents a data-driven approach for estimating EE based on data collected from 40 participants (20 MWU and 20 controls without disability) across different activities (lying, sitting and wheelchair propulsion at different intensities). We extracted features from heart rate, inertial measurement units (IMU), and individual personal characteristics to develop activity classification and EE estimation algorithms and investigate the influence of personal characteristics on EE estimates. Support Vector Machines were selected as classifiers, while Support Vector Regressors, Gaussian Processes, Random Forest, XGBoost, and Neural Networks were selected as regression models. High classification accuracy was achieved with minor confusion between activities and EE estimation results showed high generalisation capabilities of the trained models on unseen participants. We explored the impact of changing the position of the IMU on the accuracy of EE estimations. We recommend the wrist as the primary location for sensor placement. It provides a good trade-off between accuracy, high wear compliance rates and the possibility of integrating our algorithms in already existing wearable devices. Our findings showed that including data collected from people without disabilities to develop EE estimation algorithms for MWU did not enhance the estimation accuracy. In conclusion, data-driven algorithms based on wearable sensors and personal characteristics are effective for activity classification and EE estimation in MWU, but need to be personalized and further developed for daily life settings to be ecologically valid.**

***Index Terms*— Manual wheelchair users, energy expenditure, inertial measurement units, wearable sensors.**

I. Introduction

COMPARED to people without a disability, manual wheelchair users (MWU) have a less active lifestyle, which has been linked to a higher prevalence of diabetes, cardiovascular disease and obesity [1], [2]. In fact, the prevalence of obesity is 2.5 times higher in MWU [3], due to an energy intake that exceeds the energy expenditure (EE) in this group. Strategies to prevent or counteract obesity in MWU include reducing energy intake and/or increasing EE through increased physical activity (PA).

Providing feedback on the daily life PA behaviour and EE may be one strategy to enable MWU to improve their balance between energy intake and expenditure. However, gold standard methods for measuring EE (e.g., doubly labelled water or (in)direct calorimetry [4], [5], [6]) are impractical to use in daily life settings since these methods require specialized and expensive equipment and cannot easily be integrated into a MWU's daily routine. While commercial, low-cost and non-invasive wearable devices are now commonly used as alternatives to provide feedback on EE, their accuracy is variable and often insufficient even in people without a disability [7], [8], [9], [10].

In line with this, the only two studies that validated two different versions of the commercially available Apple watch with a wheelchair specific setting,[1] found that the EE estimation algorithms underestimated MWUs' EE by almost 30 percent, and the error was highly variable [11], [12]. While research-grade wearable devices seem to be more accurate for estimating EE in MWU on a group level (example of EE estimation error in previous studies: $-5.65 \pm 32.61\%$ [13] and $16.8 \pm 15.8\%$ [14]), the high error variability still remains a challenge. This indicates that neither commercially nor research-grade wearable devices are accurate enough for tracking EE in MWU, and that further development of EE estimation algorithms is necessary.

[1]Apple Watch's "Wheelchair Mode" is a feature designed specifically for wheelchair users to help them track their activity and fitness levels accurately. This mode includes customized activity tracking based on counting wheelchair pushes throughout the day.

Received 14 November 2023; revised 12 November 2024 and 13 January 2025; accepted 28 January 2025. Date of publication 30 January 2025; date of current version 13 February 2025. This work was supported by the Strategic Funding of Faculty of Medicine and Health Science at the Norwegian University of Science and Technology (NTNU), an NTNU Innovation Scholarship and NTNU Discovery. *(Corresponding author: Damiano Varagnolo.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Norwegian Centre for Research Data, Norwegian Agency for Shared Services in Education and Research, under Project No. 216680.

Roya Doshmanziari and Håvard Pettersen Reierstad, are with the Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7034 Trondheim, Norway (e-mail: roya.doshmanziari@ntnu.no).

Håkon Strand Aandahl is with the Department of Physical Education and Sport Science, Faculty of Teacher Education and Arts, Nord University, 7600 Levanger, Norway (e-mail: Hakonsaa@gmail.com).

Marius Lyng Danielsson and Julia Kathrin Baumgart Centre for Elite Sports Research, Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, 7034 Trondheim, Norway (e-mail: Marius.L.Danielsson@ntnu.no; Julia.K.Baumgart@ntnu.no).

Damiano Varagnolo is with the Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7034 Trondheim, Norway, and also with the Department of Information Engineering, University of Padua, 35122 Padua, Italy (e-mail: damiano.varagnolo@ntnu.no).

Digital Object Identifier 10.1109/TNSRE.2025.3537333

So far, EE estimation algorithms for MWU and people without a disability are mostly based on accelerometry data, which is collected in laboratory/controlled settings [15], [16]. Data-driven approaches have been shown to distinguish between standardized types of activities (exercising, biking, ambulating, sitting, lying, wheelchair propulsion, etc.) in people without a disability with high accuracy [17], [18], and in MWU with moderate to high accuracy [19]. However, simply identifying the activity type and providing an activity-type specific EE estimate is not sufficient, as EE varies both within different activities for the same individual (depending on how vigorously they engage in the activity) and between individuals (due to inter-individual differences in the energy metabolism). Previous studies also suggest that heart rate (HR) is a promising biomarker for indicating the intensity of a certain activity. HR, when individually calibrated, highly correlates with EE (although it is not a determinant of this estimand) [20], [21]. Furthemore, personal characteristics such as age, body mass, sex and the fitness level of individuals have been shown to have an impact on EE [22], [23], [24]. In addition, in MWU, the heterogeneity in disability-related differences in EE should be considered. Thus, to improve EE estimations, especially among MWU, there is a need for more sophisticated and individualized data-driven modelling based on combined accelerometry and HR data.

Indeed, in a previous study [25], we showed that it is possible to develop EE estimators during wheelchair propulsion that are accurate at least in a control group without disability. These estimators were developed via data-driven modelling by combining HR (measured with a chest-band) and accelerometery data collected by a wrist mounted inertial measurement unit (IMU) sensor, in addition to personal characteristics (age, bodymass, sex, height, etc.). However, similar approaches still need to be tested in our target population, i.e., MWU. Moreover, the optimal placement of IMU for MWU is still uncertain and needs further investigation. This aspect has been previously explored and shown to impact the accuracy of EE estimations in people without a disability [26], [27], [28], [29], [30]. Lastly, since collecting data from a large number MWU is challenging, as they constitute a minority group, there is a need to investigate whether the data can be augmented by data from participants without a disability. While significant differences exist between individuals with and without disabilities, augmenting MWU data with non-MWU data may help address data scarcity and improve model training.

We therefore investigate whether the developed EE estimators are equally accurate for MWU, and use data-driven techniques to: *A)* extract newly defined, informative movement features from IMU measurements and explore the variability of the extracted features among participants during wheelchair propulsion; *B)* investigate the classification of rest lying, rest sitting, and wheelchair activity using wearable sensors data (HR and IMU measurements), while developing EE estimates and examining the influence of personal characteristics on customizing EE estimation algorithms; *C)* explore the impact of incorporating IMU sensors data mounted on various locations of the body and wheelchair in order to enhance EE estimation accuracy. Lastly, we aim to *D)* answer the following question: considering the complexity of recruiting MWU and collecting data from this population, what are the possibilities of using data collected from people without a disability to improve the performance of the MWU-oriented EE estimation algorithms?

To do so, the paper is structured as follows: Section II presents the experimental protocol and collected data. Section III describes pre-processing of the data and feature extraction steps. Section IV introduces the classification and regression models. Section V details and discusses our quantitative results. Finally, Section VI concludes the paper with some considerations and future research directions.

TABLE I
EXPERIMENTAL SETUP FOR THE THREE SEPARATE TEST DAYS, DETAILING PRE-DETERMINED INCLINES AND TREADMILL SPEEDS (km/h) FOR EACH STAGE

Incline	0.5%		2.5%		5.0%	
	Male	Female	Male	Female	Male	Female
Stage 1	4	3	3	2	2	1
Stage 2	6	5	4	3	3	2
Stage 3	8	7	5	4	4	3

II. EXPERIMENTAL SETUP

For this study, the data collection and processing were approved by the Norwegian Centre for Research Data (ID: 216680), and the collection conducted in accordance with the Declaration of Helsinki. Participation in the study was voluntary and the participants were informed about the possibility to withdraw from the study, its purpose and risks before they signed the consent form.

A. Experiment Protocol

Each participant performed three testing days within a two-week time span. Participants were asked to refrain from high intensity activities or heavy strength training, and alcohol consumption 24 hours prior, caffeine 12 hours prior and food three hours prior to testing. Each test day started with a 10 minutes lying rest period followed by 10 minutes of seated rest, followed by a 5-minute warm-up/familiarization with the setup, and finally three 4-minute stages of wheelchair propulsion on a motorized treadmill (Forcelink Technology, Culemborg, Netherlands). These three stages were performed at increasing speed with 2-3 minutes of rest in between.

The incline-speed setup for the wheelchair propulsion stages are reported in Table I. A total of 15 recordings were available per participant (two resting periods and three exercise stages per test day). The speeds-incline combinations covered low to moderate/high exercise intensity, which are the intensities MWU typically experience during activities of daily life.

B. Data Collection: Physiological Signals, Personal Characteristics and Movement Data

On the first test day, information on sex, age, body mass, body height and type of impairment (MWU group) was collected from each participant. Moreover, participants completed a self-reported short version of the International Physical Activity Questionnaire (IPAQ) to assess habitual physical

activity levels [31], [32]. HR data were recorded during the experiments using a Polar chest strap connected to a Polar M400 HR monitor (Polar Electro Oy, Finland). Movement data were recorded by five IMUs (Gait Up Physilog5® inertial sensor, 3D accelerometer, 3D gyroscope and barometric pressure, 256Hz, GaitUp; Lausanne, Switzerland) attached to chest, upper back, wrist, seat and wheel.[2]

In addition, movement trajectories were captured by mounting retroreflective markers on the body and the wheelchair, and using a Qualisys motion capture system (measurement accuracy is millimetres and capturing frequency 120Hz). This system is considered as the gold standard for validating the estimated movement patterns from IMU sensor data. A Vyntus ergospirometer (Vyntus CPX, Vyaire, Medical GmbH, Germany) was used to measure Oxygen uptake ($\dot{V}O_2$, measured in *L/min*) and Carbon dioxide production ($\dot{V}CO_2$, measured in *L/min*) continuously during the test. Energy expenditure (EE, *kcal/min*) was then calculated according to Weir's formula [33], i.e.,

$$\mathrm{EE} = 3.94\ \dot{V}O_2 + 1.106\ \dot{V}CO_2\ .$$

III. Preprocessing Steps and Feature Extraction

The primary aim of the current study is to develop EE estimation algorithms based on the information obtained from body worn sensors during rest and wheelchair propulsion in MWU. Additionally, a classifier is developed to distinguish between resting and wheelchair propulsion modes. Combined with EE estimations, this offers valuable insights for developing a personalised biofeedback platform to promote physical activity in MWU. The performance of Machine Learning (ML) algorithms is closely tied to the quality of the collected data and informativity of the defined feature space. Thus, it is essential to implement specific preprocessing steps before training classification and regression models. This section focuses on the data processing and feature extraction steps.

A. Preprocessing Steps for Accelerometry Data

In the current study, ML algorithms were initially developed based on the features extracted from datastreams of the IMU sensor (acceleration and gyroscope time series). The features derived from IMU data are referred to as movement features. Our experimental set-up is similar to the one in [34], which showed that most human movements fall between $[0.3 - 3.5]$ Hz.

The acceleration signals (X, Y and Z) include both slow movement components due to gravity forces acting on the body, and fast components due to body movement. It is thus beneficial to remove the gravity components first before extracting movement features. As the initial and final positions of the participant on the treadmill were the same geographically, the average value of the acceleration (Acc) norm (i.e., $\|Acc\| = \sqrt{Acc_x^2 + Acc_y^2 + Acc_z^2}$) during each individual stage was assumed to be the gravitational force and subtracted. After removing the gravity component, the features were extracted from $\|Acc\|$.

[2]The upper back sensor was mounted on the C7 cervical vertebra, the chest sensor on top of the heart rate band sensor, the wrist sensor above the wrist bone, the seat sensor towards the front under the seat, and the wheel sensor between two of the spokes at a standardized distance to the center of the wheel.

TABLE II
Features Extracted From the Signals and Used for Our Estimation Purposes

Domain	Sources	Features
Time	HR, $\|Acc\|$, $Gyro_X$,$Gyro_Y$,$Gyro_Z$	Mean, Max, Min, SD, MAD[3], Energy Entropy, IQR[4], Kurtosis, Skewness
Frequency	$\|Acc\|$,Acc_X[5] $Gyro_X$,$Gyro_Y$,$Gyro_Z$	Max PSD, Min PSD, Mean PSD, Freq of Max PSD, Number of cycles, Mean cycle time, Median cycle time
Personal Characteristics	Self Reported: Measured:	Sex, IPAQ, Age, Height Body Mass, HR peak[6], HR peak percentage[7]

B. Preprocessing Steps for EE Measurements

As for pre-processing the EE values, it should be noted that these were provided as $10s$ average values. Since the increase in EE during activity often occurs with a delay, we only considered steady-state measurements during the last minute of each measured EE time series. Thus, after removing the final 10 seconds of each recorded EE time series, the last 60 seconds samples of each stage was considered as steady-state EE values for that stage. Our groundtruth EE was then defined as the mean value of these steady-state values.

C. Feature Extraction and Model Order Reduction

In the field of human activity recognition and EE tracking studies using wearable sensor technology, features derived from statistical analyses in both the time and frequency domains are well-established and commonly used [35]. In our previous study, we proposed an algorithm for EE estimation in the control group based on a list of time and frequency domain features extracted from wearable sensors, together with participants' personal characteristics [25]. While promising results confirmed that the collected data and the defined feature space are informative, we here refine the previously used set of features by introducing new movement features extracted from IMU measurements. The process of extracting and validating these new features based on the data from gold standard Qualisys motion capture system is described in detail in the following section. The complete set of features used for the estimation of EE are summarized in Table II.

Furthermore, to reduce the dimension of the feature space and computational cost, and to aid the interpretability of the results (especially highlighting the most contributing features), we implemented a Principal Component Analysis (PCA) prior to model training. The number of principal components was chosen dynamically to explain 95% of the variance of the whole dataset.

[3]Median absolute deviation.
[4]Interquartile range of the signal.
[5]Cycle features were extracted from Acc_X.
[6]Highest recorded HR value throughout all the experiments.
[7]The percentage of the highest HR recorded per stage divided by HR peak.

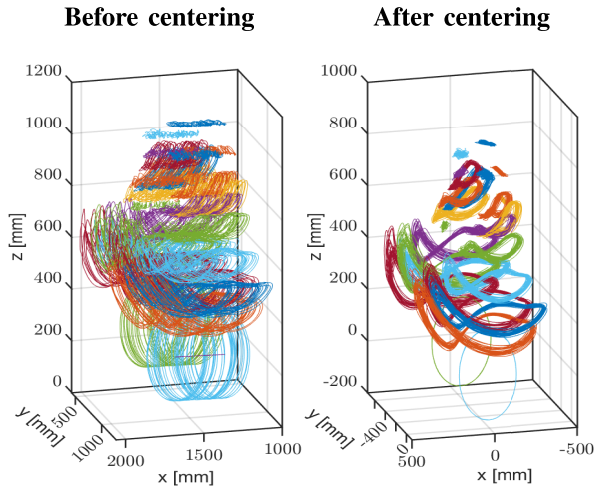

Fig. 1. Example of preprocessing movement trajectories recorded by Qualisys motion capture system by removing the variation along the treadmill.

D. Extracting Cycle Features

As described in the experiment protocol, the Qualisys motion capture system was used to track the movement of the retroreflective markers that were placed on both the participant and the wheelchair. We used the marker trajectories to calculate selected movement-related features as a reference and then extracted the same features from IMU data to establish the accuracy of the latter.

The Qualisys data was captured twice per stage for 30 seconds, roughly at one and three minutes into the four minute stages, respectively referred to as the first and second Qualisys window. The IMU data was downsampled to have the same frequency as the Qualisys data, and the high-frequency noise components of IMU measurements were removed by applying a Gaussian filter, as outlined in [36]. The left forearm marker was chosen to perform the analysis as it was the closest to the wrist IMU sensor.

The first step in the preprocessing of Qualisys data, was to remove the movement along the treadmill by centering the data. This was done by subtracting the position of the wheelchair's left wheel center from the position of the markers for each sample. An example of movement trajectories before and after centering is shown in Figure 1.

The IMU measurements had to be synchronized and cut to match the available Qualisys data within the specific time frames. Due to an inherent offset between the left forearm marker recordings and the IMU measurements, manually cutting and aligning the signals based on recording time alone was not feasible. Thus, an ad-hoc synchronizing process was performed on the IMU and Qualisys respective coordinate frame. We computed the acceleration of the wrist from Qualisys wrist marker measurements by double-differentiating, applying a moving average filter to remove noise and finding the overlap of the signals. Finally, since the Qualisys marker trajectories and IMU acceleration measurements contained propulsion cycles, we defined and extracted the following features:

- ◇ Number of cycles (i.e., how many times the participant performed an entire cyclic movement of her/his arm while pushing the wheels of the wheelchair). These have been calculated by extracting extremal points (maxima) from the acceleration data (Acc_x) within a time window.
- ◇ Average cycle time, calculated by extracting the average of the temporal differences between extremal points.
- ◇ Median cycle time, calculated by extracting the median of the temporal differences between extremal points.

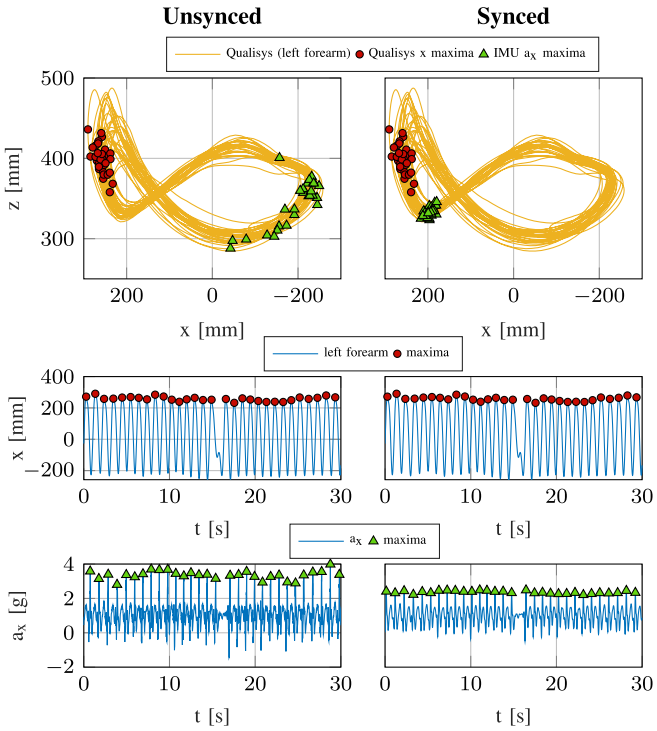

Fig. 2. Example of left forearm marker trajectories recorded by the Qualisys motion capture system and acceleration measurements from an IMU sensor mounted above the wrist, along with their respective extremal points shown before and after applying the ad-hoc synchronization process.

More specifically, to identify cycles, the *findpeaks* algorithm from the Signal Processing Toolbox in MATLAB (MathWorks, 2022) was used to detect local maximas. Figure 2 demonstrates an example of local maximas and their respective timestamps derived from X-axis of the left forearm marker trajectories and the wrist acceleration data (Acc_x). The graphs on the left show the location of detected maximas after manually cutting and aligning the signals based on the recording time, while the graphs on the right illustrate the effect of using the ad-hoc synchronization process prior to detecting local maximas. In this example, the Acc_X maximas in the unsynced variant are spread out in the recovery phase of the propulsion cycle, whereas in the synced variant all detected maximas are gathered towards the end of the push phase, which provides a more accurate number of cycles. Note that, for simplicity, the extremal seeking algorithm focused on single axis acceleration data. After examining the locations of the detected extremal points from each axis in all recorded propulsion stages (using Qualisys marker data as a reference), the X-axis of acceleration data was selected for extracting cycle features. The results of the validation of extracted features are briefly presented in section V.

IV. Models and Algorithms Structure

A. Algorithmic Choices for the Classification of Movement

The first step towards developing an EE estimation algorithms is to correctly detect the activity state of the participant. In this study, we differentiate between lying, sitting, and performing wheelchair propulsion ranging from low to moderate/high intensity. For this purpose we use Support vector machines (SVM) [37] for classification based on the extracted features from our IMU and HR measurements, as discussed in Section III-A. Prior to classification, PCA was applied to the feature dataset. Note that three records for lying, three for sitting and nine for wheelchair propulsion were available for classification per participant. For investigating which kernel function best serves this classification task, we used linear, exponential, sigmoid and radial basis function (RBF) kernels and a grid search over the hyperparameters of such kernels was performed. The available data showed that the RBF kernel, i.e.,

$$K_{RBF} = \exp(-\gamma \left\| x - x' \right\|^2)$$

provided the best classification performance, even though it did not outperform the other kernel structures. Thus, in the remainder of the paper we only report the results obtained from the RBF kernel.

B. Algorithmic Choices for the Regression of EE

Another intermediate objective of this manuscript was to find an opportune observer of EE during rest and wheelchair propulsion stages. Based on our previous work [25] and to validate the robustness of our results with regards to model selection, we focused primarily on comparing how well this task was accomplished by means of Support vector regression (SVR) with a RBF kernel [38], and Gaussian process regression (GPR) with a Matern kernel [39], i.e.,

$$\begin{aligned} K_{Matern} &= k_N + k_C \; k_{Ma}, \\ k_{Ma} &= \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}r}{l}\right), \\ r &= \left\| x - x' \right\|. \end{aligned}$$

with k_C a constant kernel, k_N a white noise kernel parameterized by the noise level σ, and the Matern kernel k_{Ma} (a general form of RBF kernel parameterized with a length-scale parameter $l > 0$). Furthermore, to provide more comprehensive analyses, Random Forest (RF), Neural Networks (NN) and a boosting XGBoost regression algorithms were also applied for EE estimation [40], [41], [42]. Hyperparameter optimization for each regressor was conducted using a grid search across model parameters.

C. Cross Validating the Results

To validate the estimation performance, to avoid selection bias and over-fitting, and to explore the generalization capabilities of the estimators on unseen data, we employed a set of cross validation (CV) and personalisation scenarios. More specifically, we considered the following CV strategies (each partitioning data into approximately $70 - 30\%$ training and holdout splits).

- ◇ *Record-wise CV*: each fold contained data from random participants and random wheelchair propulsion stages (i.e., a repeated k-fold (RKF) for maximal randomness). This strategy enabled the investigation of how informative the overall dataset and the experiment design were.
- ◇ *Participant-wise CV*: we shuffle split over participants identifiers. In other words, each fold contained data "per participant" - thus for each iteration the entire data of some participants was contained in the training set, while the data of the remaining participants comprised the test set. This strategy enabled exploring prediction capabilities on unseen participants.

Note that RKF and shuffle split CV were performed with ten repetitions and the average over all repetitions is reported. To investigate the level of necessary personalisation in model training based on personal characteristics of individuals, we analysed the prediction accuracy of the models by dividing the data based on sex and IPAQ factor before CV split.

As performance indexes we considered: ◇ the accuracy of the classifier, i.e., the ratio of the number of correct predictions and the total number of predictions on the test data in SVM; ◇ the average coefficient of determination R^2, the root mean squared error (RMSE) for the EE regressors, and the corresponding standard deviations (SD; averaged across all validation repetitions).

V. Results

This section summarizes the classification and regression results from the data analyses described above, and discusses the estimated prediction capabilities of the trained models on unseen data for the MWU and control group. We present the validation results of the introduced movement features, in addition to investigating the impact of personalising the models based on sex and IPAQ factor (a physical activity level index). Furthermore, we briefly present and discuss the results related to identifying an optimal IMU sensor placement. Lastly, given that it is challenging to collect data of a sufficient number of MWU, mixture models and cross testing scenarios were defined to investigate the possibility of using data collected of the control group to train models for the MWU group.[8]

A. Participants

The experimental data included in this study was from 20 MWU [11 men and 9 women; age $= 37 \pm 13$ years (mean $\pm$ S.D.); body mass (kg): 74.5 ± 18.6; height (cm): 172.5 ± 12.2] and 20 participants without a disability (control group) [11 men and 9 women; age $= 33 \pm 11$ years (mean $\pm$ S.D.); body mass (kg): 75.2 ± 11.4; height (cm):176.2 ± 9.9]. The MWU had one of the following impairments: SCI (n $=$ 11), Spina bifida (n $=$ 2), Cerebral palsy (n $=$ 2) and other neurological or musculo-skeletal impairments (n $=$ 5).

B. Movement Features Analysis

This section presents the accuracy of IMU extracted cycle features compared to the gold standard Qualisys marker

[8]Note that some records were excluded from the overall dataset due to missing data or not completed wheelchair propulsion stages.

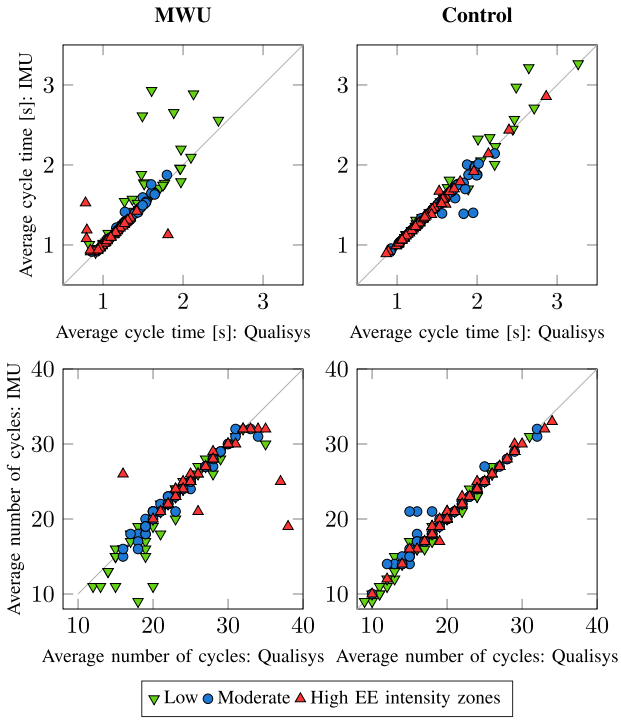

Fig. 3. Comparing cycle features extracted from IMU and Qualisys (ground-truth) data in control and MWU group labeled by EE intensity zones.

trajectories, and investigates the importance of these features by how much they correlate with EE. Figures 3 visualizes examples of cycle features extracted from the left forearm marker trajectories and the wrist IMU measurements in the MWU and control datasets (note that this part of the analysis includes only the wheelchair propulsion stages). Here steady state EE values from all wheelchair propulsion stages are sorted and split equally into three exercise intensity zones (low, moderate, and high) and the corresponding features extracted per stage are labeled by these intensity zones. For each stage, the estimated features and the error—calculated as the difference between the estimated features from both sources—were averaged across participants and summarized in Table III. The results indicate that the cycle features identified from the IMU measurements are close to the reference line, demonstrating high accuracy compared to the left forearm Qualisys marker data. However, the error rate analysis reveals greater accuracy in the control group than in the MWU group, which is unexpected. In addition, outliers in slower propulsion stages (low intensity zones) could be explained by reduced acceleration and smaller motion ranges in both MWU and control group participants, which can make it challenging for algorithms to detect extremal points accurately. When comparing the features between the MWU and control group, a lower cycle rate was found in the control group. Although this warrants further investigation, it may be attributed to increased trunk range of motion in the control group at least in comparison to participants with higher-level impairments in the MWU group.

In conclusion, the high accuracy of cycle feature detection algorithms based on IMU data is promising but needs to be confirmed in non-standardized daily-life settings.

TABLE III
SUMMARY OF MOVEMENT FEATURES EXTRACTED FROM QUALISYS (GROUND-TRUTH) AND IMU MEASUREMENTS IN MWU AND CONTROL DATASETS

			Cycle Times (seconds)			Number of Cycles (pushes/min)		
	Incline	Stage	Qualisys	IMU	Error	Qualisys	IMU	Error
MWU Group	0.5%	Stage 1	1.42	1.55	0.12	41.36	39.58	2.00
		Stage 2	1.20	1.21	0.01	49.80	49.20	1.40
		Stage 3	1.05	1.11	0.06	56.14	53.28	3.42
	2.5%	Stage 1	1.42	1.57	0.15	41.76	39.30	2.70
		Stage 2	1.23	1.24	0.01	48.12	48.22	1.00
		Stage 3	1.08	1.10	0.03	54.80	53.74	2.14
	5.0%	Stage 1	1.64	1.71	0.10	36.36	35.42	1.88
		Stage 2	1.29	1.26	0.04	46.10	47.36	1.48
		Stage 3	1.06	1.10	0.03	54.62	53.54	1.38
Control Group	0.5%	Stage 1	1.82	1.82	0.02	33.00	33.00	0.66
		Stage 2	1.46	1.47	0.01	42.00	41.78	0.42
		Stage 3	1.29	1.30	0.01	49.06	49.06	0.46
	2.5%	Stage 1	1.62	1.59	0.04	37.18	38.00	1.06
		Stage 2	1.46	1.46	0.01	41.12	41.88	0.78
		Stage 3	1.33	1.33	0.00	44.66	45.00	0.34
	5.0%	Stage 1	1.80	1.84	0.19	33.26	34.38	2.62
		Stage 2	1.47	1.45	0.02	41.30	41.52	0.46
		Stage 3	1.25	1.25	0.00	47.74	48.00	0.26

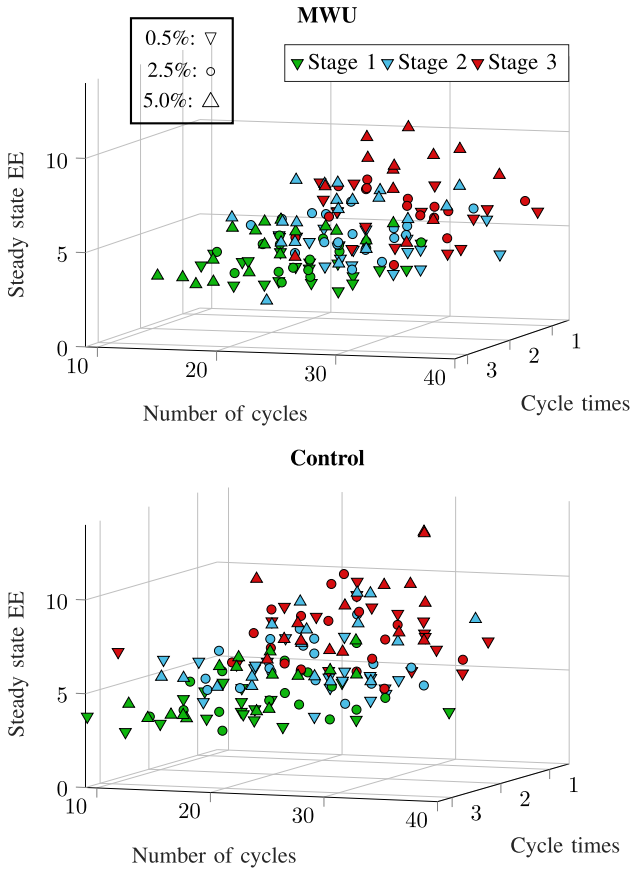

Fig. 4. Relationship between extracted cycle features and steady state energy expenditure, labeled by wheelchair propulsion stages (color mapped) and test day (incline 0.5%: down-pointing triangle, incline 2.5%: circle and incline 5%: up-pointing triangle).

Figure 4 shows the interaction between cycle features extracted from Left forearm marker measurements and steady state EE values in MWU and control group datasets. The feature space is color-labeled by wheelchair propulsion stages and shape-labeled by incline. Cycle features were found to be significantly positively related with EE, which highlights the importance of including these features in the EE estimation

TABLE IV

TOP 10 CONTRIBUTING FEATURES FOR EACH PRINCIPAL COMPONENT, WITH THEIR RESPECTIVE EXPLAINED VARIANCE INDICATED IN PARENTHESES. COLORS REPRESENT THE SOURCE OF EACH FEATURE: GYROSCOPE (X, Y, Z) IN DARK GRAY, NORM OF ACCELEROMETER DATA IN LIGHT GRAY, AND HEART RATE AND PERSONAL CHARACTERISTICS IN BLUE (DEFINED IN TABLE II)

PC1 (59.33%)		**PC2** (9.34%)		**PC3** (4.97%)		**PC4** (3.82%)		**PC5** (3.22%)	
Entropy (Z axis) :	3.5%	Sex:	21.7%	IPAQ:	15.4%	Median Cycle Time:	3.7%	Entropy (Norm):	3.6%
Entropy (Y axis) :	3.3%	Height:	9.8%	HR Peak:	12.5%	IPAQ:	3.5%	HR Max:	3.5%
Num of Cycles :	3.3%	Body Mass:	9.4%	Sex:	10.9%	HR Max:	3.1%	Sex:	3.4%
Entropy (X axis) :	3.2%	Age:	8.1%	Age:	10.4%	Entropy (Z axis):	3.1%	PSD Max (X axis):	3.1%
Min (Norm) :	2.7%	HR Peak:	4.8%	Body Mass:	4.2%	HR Peak%:	2.9%	HR Mean:	3.0%
MAD (Norm) :	2.5%	IPAQ:	4.8%	HR Peak%:	3.6%	Energy (Norm):	2.7%	HR Peak:	2.7%
Min (Z axis):	2.4%	HR Entropy:	2.5%	HR Kurtosis:	3.2%	Entropy (Y axis):	2.6%	HR Peak%:	2.7%
Max (Z axis) :	2.4%	HR Max:	1.8%	Median Cycle Time:	2.8%	PSD Min (X axis):	2.5%	HR Min:	2.6%
SD (Z axis):	2.4%	MAD (Norm):	1.6%	Max (X axis):	1.7%	HR SD:	2.4%	Entropy (Z axis):	2.6%
IQR (Norm):	2.4%	PSD Max (X axis):	1.4%	Mean Cycle Time:	1.6%	HR Mean:	2.4%	Entropy (Y axis):	2.4%

algorithms. The figures also indicate that the participants have the highest EE at incline 5%.

C. PCA, Classification and Regression Analysis

1) PCA: To aid model interpretability and understand which of the extracted features contribute the most to explaining the variability in our data, we present a detailed analysis of the PCA results for the MWU group dataset. Table IV reports the top 10 contributing features for each PC, with their respective explained variance. The explained variance of PC1 (59.33%) highlights the importance of the defined movement features (mainly entropy, number of cycles, Min, Max and SD values from IMU measurements). Entropy and cycle features are likely indicators of propulsion intensity and consistency. PC2, with a variance of 9.34%, is a combination of demographic factors like sex, height, and body mass, age and IPAQ factor. These features likely contribute towards adjusting the model to interindividual physiological differences. HR metrics, including HR Peak, HR Kurtosis, Min, Max and Entropy, which reflect cardiovascular response, contribute in PC3 (4.97%) and PC4 (3.82%). HR features are indicators of individual physiological responses to increased intensity. Similar loadings were achieved for the control group (with slightly differences in the percentage of contributing features; PC1 (57%), PC2 (11%) and PC3(6%)).

2) Classification: Table V shows the SVM classification performance on the test data resulting from the defined CV and sex/IPAQ based personalisation scenarios. Furthermore, Figure 5 illustrates the confusion matrices from participant-wise CV, resulting from the overall MWU and control group datasets. The diagonal and off-diagonal values represent the number of correctly classified and misclassified records (in light green), with the percentage below indicating their proportion relative to the total dataset. The last row and column summarize the total predicted and actual records per class, along with the sensitivity and precision (in green), and error rates (in red). As expected, the wheelchair propulsion stages were separable without confusion from resting stages (i.e., lying and sitting). Minor confusion was found for classifying lying and sitting stages. However, a binary classification between wheelchair propulsion and resting (combined lying and sitting stages), resulted in perfect accuracy for our datasets. Different CV and sex/IPAQ-based personalisation

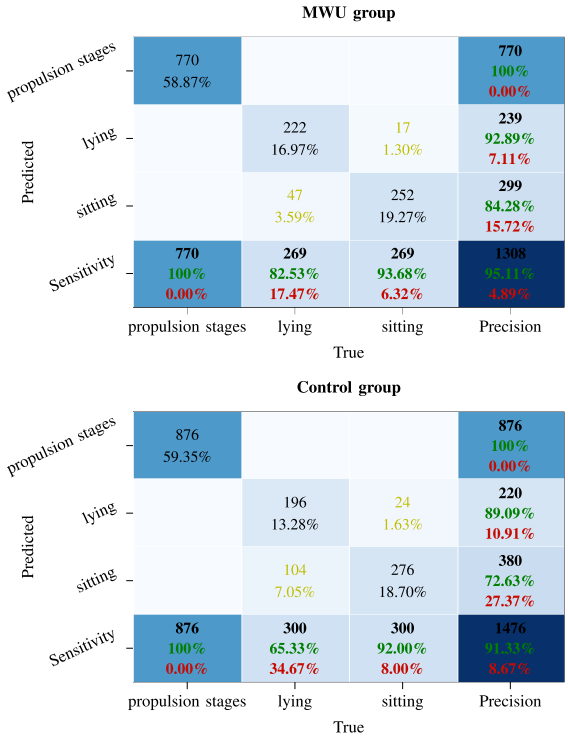

Fig. 5. Confusion matrices explaining the overall precision and sensitivity of the SVM model in classifying lying, sitting and wheelchair propulsion stages.

scenarios were found to have no significant impact on the performance of the classifier.

3) Regression: The regression results are reported in Table V and include the mean and SD of repeated CV scenarios. Results confirm the validity of the proposed EE estimation approaches. Comparing the overall performance indexes in record-wise and participant-wise CV proves the informativity of the extracted features and significant generalisation capabilities of the trained models on unseen stages and participants. Note that to assess the impact of PCA on model performance, we compared models trained with and without PCA. The results showed that applying PCA improved model accuracy (specifically in participant-wise CV) while reducing computational complexity and overfitting.

Comparing SVR and GPR results trained on the MWU group indicate that both models succeeded, since an acceptable level of precision and consistency was achieved. In addition, EE estimations are robust as indicated by small SDs

TABLE V
CLASSIFICATION AND REGRESSION RESULTS, MODELS TRAINED SEPARATELY ON CONTROL AND MWU GROUP

Data	Classifier	CV	Overall Accuracy: mean±SD	Sex: Male Accuracy: mean±SD	Sex: Female Accuracy: mean±SD	IPAQ: LOW Accuracy: mean±SD	IPAQ: MOD Accuracy: mean±SD	IPAQ: HIGH Accuracy: mean±SD
MWU	SVR	Record-wise	0.94±0.03	0.92±0.03	0.89±0.04	0.82±0.10	0.92±0.05	0.84±0.07
		Participant-wise	0.95±0.02	0.95±0.04	0.94±0.03	0.93±0.00	0.93±0.05	0.88±0.05
Control	SVR	Record-wise	0.90±0.03	0.89±0.05	0.85±0.06	-	0.86±0.04	0.88±0.05
		Participant-wise	0.91±0.03	0.91±0.04	0.92±0.03	-	0.88±0.04	0.89±0.02

Data	Regressor	CV	Overall R^2	Overall RMSE	Sex: Male R^2	Sex: Male RMSE	Sex: Female R^2	Sex: Female RMSE	IPAQ: LOW R^2	IPAQ: LOW RMSE	IPAQ: MOD R^2	IPAQ: MOD RMSE	IPAQ: HIGH R^2	IPAQ: HIGH RMSE
MWU	GPR	Record-wise	0.94±0.02	0.52±0.06	0.96±0.01	0.47±0.08	0.91±0.03	0.50±0.08	0.96±0.02	0.45±0.13	0.93±0.03	0.57±0.10	0.91±0.03	0.59±0.08
		Participant-wise	0.90±0.02	0.70±0.06	0.92±0.05	0.64±0.15	0.84±0.06	0.61±0.12	0.49±0.31	1.42±0.31	0.91±0.04	0.70±0.16	0.84±0.03	0.78±0.07
	SVR	Record-wise	0.92±0.01	0.63±0.06	0.91±0.02	0.70±0.11	0.87±0.03	0.60±0.12	0.81±0.07	1.01±0.32	0.86±0.06	0.81±0.24	0.86±0.03	0.73±0.10
		Participant-wise	0.90±0.01	0.70±0.06	0.89±0.02	0.80±0.08	0.83±0.03	0.66±0.10	0.61±0.03	1.40±0.31	0.81±0.07	1.01±0.27	0.81±0.04	0.86±0.11
	RFR	Record-wise	0.79±0.06	0.99±0.14	0.83±0.04	0.98±0.13	0.71±0.12	0.89±0.19	0.63±0.22	1.34±0.26	0.75±0.07	1.08±0.22	0.77±0.06	0.94±0.13
		Participant-wise	0.74±0.08	1.09±0.16	0.82±0.06	1.01±0.16	0.62±0.13	0.96±0.12	0.31±0.36	1.72±0.17	0.65±0.11	1.34±0.26	0.68±0.12	1.11±0.17
	XGBoost	Record-wise	0.89±0.02	0.71±0.07	0.90±0.02	0.76±0.09	0.84±0.05	0.65±0.11	0.76±0.23	1.06±0.32	0.85±0.04	0.83±0.16	0.85±0.05	0.76±0.14
		Participant-wise	0.86±0.04	0.79±0.11	0.86±0.05	0.88±0.18	0.76±0.09	0.75±0.09	0.41±0.19	1.66±0.13	0.78±0.06	1.05±0.19	0.79±0.06	0.90±0.11
	NN	Record-wise	0.93±0.02	0.56±0.08	0.94±0.02	0.57±0.11	0.88±0.03	0.56±0.09	0.93±0.04	0.59±0.18	0.89±0.05	0.70±0.13	0.88±0.03	0.69±0.08
		Participant-wise	0.89±0.04	0.72±0.11	0.92±0.03	0.66±0.12	0.82±0.06	0.67±0.11	0.44±0.29	1.52±0.21	0.88±0.03	0.78±0.13	0.82±0.05	0.85±0.12
Control	GPR	Record-wise	0.93±0.01	0.65±0.06	0.93±0.02	0.74±0.10	0.90±0.03	0.56±0.09	-	-	0.88±0.03	0.85±0.11	0.96±0.01	0.49±0.07
		Participant-wise	0.87±0.05	0.89±0.15	0.91±0.03	0.86±0.11	0.78±0.08	0.82±0.15	-	-	0.80±0.04	1.10±0.17	0.94±0.01	0.59±0.07
	SVR	Record-wise	0.90±0.02	0.78±0.09	0.90±0.02	0.88±0.12	0.88±0.04	0.63±0.11	-	-	0.84±0.05	1.03±0.22	0.91±0.02	0.73±0.14
		Participant-wise	0.89±0.02	0.85±0.11	0.89±0.03	0.92±0.17	0.83±0.07	0.73±0.14	-	-	0.80±0.05	1.11±0.25	0.90±0.04	0.77±0.22
	RFR	Record-wise	0.78±0.03	1.17±0.09	0.76±0.07	1.37±0.17	0.81±0.06	0.78±0.12	-	-	0.68±0.10	1.43±0.20	0.81±0.06	1.06±0.19
		Participant-wise	0.70±0.09	1.37±0.23	0.77±0.06	1.34±0.14	0.75±0.09	0.89±0.16	-	-	0.65±0.14	1.44±0.23	0.75±0.09	1.20±0.28
	XGBoost	Record-wise	0.89±0.02	0.84±0.07	0.89±0.03	0.94± 0.12	0.86±0.04	0.68±0.08	-	-	0.82±0.04	1.07±0.14	0.89±0.03	0.79±0.13
		Participant-wise	0.84±0.05	1.00±0.17	0.86±0.03	1.05±0.15	0.78±0.08	0.83±0.13	-	-	0.77±0.07	1.18±0.23	0.82±0.07	1.02±0.26
	NN	Record-wise	0.91±0.01	0.76±0.06	0.90±0.03	0.89±0.13	0.88±0.05	0.63±0.11	-	-	0.84±0.06	0.99±0.13	0.92±0.05	0.67±0.20
		Participant-wise	0.84±0.07	0.99±0.19	0.86±0.04	1.06±0.14	0.76±0.08	0.86±0.15	-	-	0.71±0.12	1.29±0.19	0.89±0.03	0.78±0.14

across randomized CV repeats. Comparing additional regression results indicates that while RF and XGBoost generally displayed lower accuracy, NN performed similarly to SVR in most cases but occasionally showed slightly lower accuracy in participant-wise CV. As SVR and GPR models outperformed or matched the alternatives, and due to space constraints, we present simulation results for only these two models in the remainder of this study.

Figure 6 illustrates an example of GPR estimation results trained on overall MWU data. The top figure shows the results of testing over a randomly selected fold from RKF CV. Here, the EE measurements are sorted and divided to represent resting stages (left), lower intensity wheelchair propulsion stages (middle), and higher intensity wheelchair propulsion stages (right). The bottom figure is an example of participant-wise CV in which EE estimations in all records are shown for six MWU participants (i.e., nine records for the exercise stages and six records for the resting stages per participant). The visualizations show that the accuracy of the model prediction is acceptable within all EE intensity divisions and unseen participants. Table V also presents details of the estimation results while training personalised models based on sex and IPAQ factor divisions. Exploring overall (combined male/female data) and personalised results showed consistency with respect to selected regression models, which suggests that experiment design, collected data and feature extraction steps possibly have more impact on the accuracy than model selection. Within the control group, we see a clear distinction in estimation accuracy related to sex. The models trained on male data showed comparatively weaker performance compared to the models trained on a combination of male and female data (overall dataset). In contrast, models trained on female data and IPAQ divisions provided higher

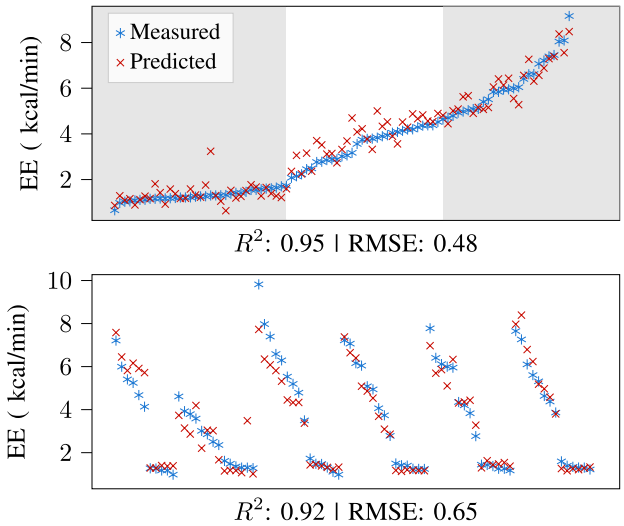

Fig. 6. GPR model predictions, trained on the overall dataset from MWU (including rest and wheelchair propulsion stages), results are randomly selected from one cross-validation repeat (top: RKF, bottom: shuffle-split).

estimation accuracy compared to overall dataset. However, compared to overall dataset, grouping based on female and IPAQ in MWU resulted in lower prediction capabilities and separation based on male participants provided similar results achieved by the overall dataset. This can be explained by a higher variability in body mass within the female participants' data in the MWU group, since body mass is known to be highly correlated with EE. Furthermore, the self-reported IPAQ factor is originally developed for people without a disability, and may thus not be tailored for these users. In other words, a fair comparison may require modifying the IPAQ factor to account for inherent differences between MWU and control group participants.

TABLE VI
REGRESSION RESULTS, MODEL TRAINED ON THE DATASET FROM MWU AND VARIOUS IMU SENSOR LOCATIONS ON BODY AND THE WHEELCHAIR

Scenario	IMU Location	GPR				SVR			
		Record-wise		Participant-wise		Record-wise		Participant-wise	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
1	Wrist	0.88±0.02	0.58±0.06	0.80±0.05	0.74±0.08	0.79±0.03	0.79±0.09	0.73±0.04	0.87±0.09
2	Chest	0.88±0.02	0.58±0.07	0.72±0.09	0.87±0.13	0.78±0.06	0.81±0.14	0.67±0.07	0.95±0.11
3	Wheel	0.85±0.04	0.66±0.09	0.64±0.13	0.99±0.16	0.76±0.04	0.84±0.09	0.63±0.12	1.00±0.17
4	Back	0.85±0.03	0.65±0.07	0.58±0.21	1.05±0.26	0.76±0.04	0.84±0.10	0.61±0.19	1.02±0.23
5	Wrist + Back	0.89±0.02	0.56±0.07	0.71±0.15	0.87±0.22	0.78±0.03	0.80±0.09	0.64±0.15	0.98±0.20
6	Wrist + Wheel	0.90±0.02	0.53±0.07	0.82±0.09	0.70±0.16	0.81±0.03	0.75±0.08	0.74±0.09	0.85±0.15
7	Wrist + Chest	0.90±0.02	0.53±0.06	0.80±0.07	0.73±0.11	0.80±0.04	0.77±0.11	0.73±0.05	0.86±0.09
8	Chest + Wheel	0.89±0.03	0.57±0.07	0.76± 0.09	0.81±0.13	0.79±0.04	0.78±0.10	0.71±0.08	0.89±0.13
9	Wrist + Chest + Wheel	0.91±0.02	0.52±0.07	0.81±0.08	0.72±0.13	0.81±0.03	0.76±0.09	0.73±0.08	0.86±0.12
10	Wrist + Back + Wheel	0.90±0.02	0.54±0.07	0.74±0.13	0.82±0.21	0.79±0.04	0.78±0.10	0.67±0.17	0.94±0.23

Finally, clear signs of overfitting were found in low IPAQ MWU (comprising only 3 participants). This is likely due to that dividing the data into smaller and smaller datasets (e.g., dividing MWU participants based on their physical activity levels) increases the risks of overfitting.

D. IMU Sensor Placement

Table VI shows the results of exploring the dependency of EE estimation accuracy on the placement of the IMU sensor. This table considers various single sensor measurements in addition to sensor fusion scenarios. Table VI presents GPR and SVR regression results applied on the overall MWU dataset. Note that lying and sitting rest measurements are not included in this analysis since movement features do not contribute to estimating EE during rest periods. However, excluding resting samples lead to a smaller number of measurements in the model training phase, and caused a slight decrease in the overall estimation accuracy. Furthermore, the data from the sensor located under the seat was excluded from the analysis due to noisy measurements.

The MWU dataset corresponding to each sensor location was divided into train (70%) and test (30%) splits and previously described RKF and SS CV were applied. In order to have a fair comparison, regressor kernel functions were set the same through all sensor placement simulations.

The findings from single-sensor-based EE estimations reveal that the wrist sensor, when compared to chest, wheel and back-mounted sensors, resulted in the highest accuracy for EE estimations. However, considering sensor fusion scenarios and a trade-off between fit level and estimation error, the wrist and wheel combination seems to provide the most informative dataset and cost-effective solution. The combination of the wrist and wheel sensor may also be feasible for non-standardized daily life settings, where reduced accuracy (due to an increased variability in arm movements) is expected for the wrist sensor. Although there is a slight performance improvement with the chest-wrist-wheel datasets, this improvement is not large enough to justify the increased cost and data management overhead that come with a three-sensor solution.

Fig. 7. Illustration of train, validation and test sets in mixture model scenarios where each block represents 5 participants (dark gray: MWU, dark gray diagonal cross hatching: MWU, light gray: control group).

In conclusion, a single above-the-wrist mounted IMU or a combination of wrist and wheel placement are most likely the best in terms of cost-effectiveness. In addition, instead of pursuing additional IMU sensors, it might be more advantageous to explore the integration of alternative measures such as heart rate variability, body temperature, Galvanic skin response, power meters, food diaries and more-in-detail body composition features to enhance EE estimation accuracy and MWU PA monitoring.

E. Mixture Models

An additional objective of this study is to explore the possibilities of training regression models on a mixture of data collected from the MWU and control group. We investigate this by cross testing and validation scenarios schematized in Figure 7, report the corresponding results in Table VII, and comment on them below (note that the results of the mixture models validation are based on the wrist IMU sensor).

- ◇ *S1*: trained and tested on control (20 participants) and validated on MWU (20 participants),
- ◇ *S2*: trained and tested on MWU (15 participants), validated on MWU (5 participants),
- ◇ *S3*: trained and tested on 15 MWU + 5 control, validated on MWU (5 participants),

TABLE VII
REGRESSION RESULTS, MODEL TRAINED ON SEVERAL VALIDATION SCENARIOS DEFINED IN FIGURE 7

Regression	CV	S 1		S 2		S 3		S 4		S 5	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
GPR	Record-wise	0.74±0.04	1.10±0.06	0.87±0.05	0.79±0.07	0.87±0.02	0.79±0.06	0.87±0.03	0.80±0.06	0.88±0.02	0.76±0.06
	Participant-wise	0.68±0.12	1.22±0.15	0.88±0.04	0.74±0.10	0.86±0.06	0.82±0.11	0.87±0.03	0.78±0.11	0.86±0.04	0.81±0.08
SVR	Record-wise	0.80±0.02	0.97±0.09	0.85±0.02	0.85±0.08	0.88±0.01	0.74±0.08	0.87±0.01	0.77±0.08	0.88±0.01	0.75±0.08
	Participant-wise	0.78±0.05	1.00±0.11	0.86±0.03	0.81±0.14	0.88±0.02	0.77±0.18	0.87±0.01	0.77±0.12	0.88±0.02	0.76±0.05

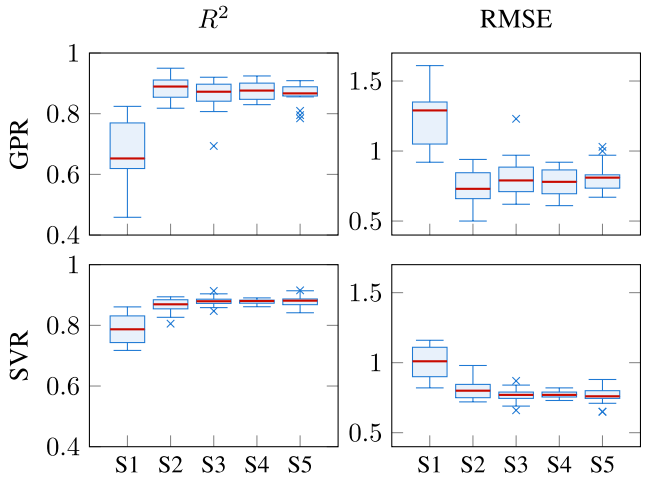

Fig. 8. Comparison between GPR and SVR results within participant-wise cross validation reporting mixture models definitions according to Figure (7).

- ◇ *S4*: trained and tested on 15 MWU + 10 control, validated on MWU (5 participants),
- ◇ *S5*: trained and tested on 15 MWU + 20 control, validated on MWU (5 participants).

As shown in Figure 8, performance metrics in the defined mixture regression models confirm the inherent differences between MWU and control group participants. More specifically, the added information collected from the control group did not improve EE estimation performance in MWU group. Moreover, training the model solely on control group and testing it on the MWU data resulted in the lowest fit and highest estimation errors. Consistent results from both regressors support this conclusion. This is unfortunate since it indicates that augmenting the data from MWUs with data from people without a disability is not a suitable solution.

VI. CONCLUSION

We investigated the capabilities of Support vector machines for classifying different activities (lying, sitting and wheelchair propulsion), and Support vector regression, Gaussian process regression, Random Forest, Neural Networks and XGBoost models for providing estimations of energy expenditure (EE). The proposed approach was feasible in classifying the type of activity with overall accuracy of 95% in MWU and 91% in control group. Similar conclusions were reached when testing the capabilities of regression models for estimating steady-state EE. The performance of the different cross-validation scenarios indicated that these models work well on unseen data (i.e., unseen participants with overall accuracy of 90% in MWU and 89% in control group). Similarities and differences among and between manual wheelchair users (MWU) and people without a disability (control group) were discussed, and the dependency of the models' performance on the characteristics of individuals for personalising purposes were explored.

Concerning the IMU sensor placement, important aspects to consider are accuracy, comfort, ease of use and wear compliance. This study concluded that the classification and regression results based on the data collected from a HR sensor and one single IMU sensor mounted above the wrist are sufficiently accurate, at least in the laboratory setting. Additionally, wearing a sensor on the wrist has shown high wear compliance rates and provides the possibility of integrating our algorithms in already existing wearable wrist-worn devices. In contrast, the back and chest sensors are uncomfortable to use during a full day and only a few commercially available devices are available for these locations. We thus do not recommend using sensors located on back or chest when collecting data over extend periods of time.

While this requires further investigation, in non-standardized daily life settings the inclusion of an extra sensor (i.e., on the wheel) may be necessary due to the potential variability in arm movements which can reduce the accuracy of wrist data. Finally, the results of this study suggest that combining data from the MWU and control group does not lead to estimating EE more accurately for MWU, likely due to the inherent differences between these groups.

In summary, the experimental data in the current lab settings is informative to develop a preliminary EE estimation framework. However, it is important to highlight that ecological validity was reduced since data was collected in a standardized laboratory setting. Therefore, collecting data from MWU performing unstandardized daily life activities will improve ecological validity of the algorithms. Furthermore, collecting data from a larger group of participants to further develop individualized models that account for inherent differences in MWU physiology is a necessity. Once the algorithms are valid for daily life settings, integrating them in a personalized biofeedback framework for MWU is the next step. Such biofeedback frameworks could be used in individualized rehabilitation programs to track patients' balance between EE and intake, while also taking other parameters such as physical activity, sleep and stress levels into account. Ultimately, and when integrated in the health care system, biofeedback systems can provide valuable insight for practitioners to adjust guidelines and recommendations related to physical activity, health and well-being of MWU.

ACKNOWLEDGMENT

The funding bodies had no role in the design of this study nor its execution, analyses, interpretation of the data, or decision to submit results.

REFERENCES

[1] M. L. Wilby, "Physical mobility impairment and risk for cardiovascular disease," *Health Equity*, vol. 3, no. 1, pp. 527–531, Oct. 2019.

[2] N. J. Holm, T. Møller, L. Adamsen, L. T. Dalsgaard, F. Biering-Sorensen, and L. H. Schou, "Health promotion and cardiovascular risk reduction in people with spinal cord injury: Physical activity, healthy diet and maintenance after discharge—Protocol for a prospective national cohort study and a preintervention- postintervention study," *BMJ Open*, vol. 9, no. 12, Dec. 2019, Art. no. e030310.

[3] E. Weil et al., "Obesity among adults with disabling conditions," *Jama*, vol. 288, no. 10, pp. 1265–1268, 2002.

[4] K. R. Westerterp, "Doubly labelled water assessment of energy expenditure: Principle, practice, and promise," *Eur. J. Appl. Physiol.*, vol. 117, no. 7, pp. 1277–1285, Jul. 2017.

[5] G. P. Kenny, S. R. Notley, and D. Gagnon, "Direct calorimetry: A brief historical review of its use in the study of human metabolism and thermoregulation," *Eur. J. Appl. Physiol.*, vol. 117, no. 9, pp. 1765–1785, Sep. 2017.

[6] H. A. Haugen, L. Chan, and F. Li, "Indirect calorimetry: A practical guide for clinicians," *Nutrition Clin. Pract.*, vol. 22, no. 4, pp. 377–388, Aug. 2007.

[7] G. Hajj-Boutros, M. Landry-Duval, A. S. Comtois, G. Gouspillou, and A. D. Karelis, "Wrist-worn devices for the measurement of heart rate and energy expenditure: A validation study for the Apple Watch 6, Polar Vantage V and Fitbit Sense," *Eur. J. Sport Sci.*, vol. 23, no. 2, pp. 165–177, Feb. 2023.

[8] R. O'Driscoll et al., "How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies," *Brit. J. Sports Med.*, vol. 54, no. 6, pp. 332–340, Mar. 2020.

[9] K. L. Dannecker, N. A. Sazonova, E. L. Melanson, E. S. Sazonov, and R. C. Browning, "A comparison of energy expenditure estimation of several physical activity monitors," *Med. Sci. Sports Exerc.*, vol. 45, no. 11, pp. 2105–2112, 2013.

[10] H. Vathsangam, E. T. Schroeder, and G. S. Sukhatme, "Hierarchical approaches to estimate energy expenditure using phone-based accelerometers," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1242–1252, Jul. 2014.

[11] D. Moreno et al., "Validity of caloric expenditure measured from a wheelchair user smartwatch," *Int. J. Sports Med.*, vol. 41, no. 8, pp. 505–511, Jul. 2020.

[12] M. L. Danielsson, M. Vergeer, G. Plasqui, and J. K. Baumgart, "Accuracy of the apple watch series 4 and fitbit versa for assessing energy expenditure and heart rate of wheelchair users during treadmill wheelchair propulsion: Cross-sectional study," *JMIR Formative Res.*, vol. 8, May 2024, Art. no. e52312.

[13] S. V. Hiremath, S. S. Intille, A. Kelleher, R. A. Cooper, and D. Ding, "Estimation of energy expenditure for wheelchair users using a physical activity monitoring system," *Arch. Phys. Med. Rehabil.*, vol. 97, no. 7, pp. 1146–1153, Jul. 2016.

[14] T. E. Nightingale, J. P. Walhin, D. Thompson, and J. L. J. Bilzon, "Predicting physical activity energy expenditure in wheelchair users with a multisensor device," *BMJ Open Sport Exerc. Med.*, vol. 1, no. 1, 2015.

[15] A. Pande, J. Zhu, A. K. Das, Y. Zeng, P. Mohapatra, and J. J. Han, "Using smartphone sensors for improving energy expenditure estimation," *IEEE J. Transl. Eng. Health Med.*, vol. 3, pp. 1–12, 2015.

[16] Z. Ni, T. Wu, T. Wang, F. Sun, and Y. Li, "Deep multi-branch two-stage regression network for accurate energy expenditure estimation with ECG and IMU data," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 10, pp. 3224–3233, Oct. 2022.

[17] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sensors J.*, vol. 17, no. 2, pp. 386–403, Jan. 2017.

[18] A. Narayanan, F. Desai, T. Stewart, S. Duncan, and L. Mackay, "Application of raw accelerometer data and machine-learning techniques to characterize human movement behavior: A systematic scoping review," *J. Phys. Activity Health*, vol. 17, no. 3, pp. 360–383, Mar. 2020.

[19] K. Lankhorst, M. Oerbekke, R. Van Den Berg-Emons, T. Takken, and J. de Groot, "Instruments measuring physical activity in individuals who use a wheelchair: A systematic review of measurement properties," *Arch. Phys. Med. Rehabil.*, vol. 101, no. 3, pp. 535–552, Mar. 2020.

[20] S. M. Ceesay et al., "The use of heart rate monitoring in the estimation of energy expenditure: A validation study using indirect whole-body calorimetry," *Brit. J. Nutrition*, vol. 61, no. 2, pp. 175–186, Mar. 1989.

[21] S. Levikari et al., "Improving energy expenditure estimation in wrist-worn wearables by augmenting heart rate data with heat flux measurement," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[22] H. Hiilloskorpi et al., "Factors affecting the relation between heart rate and energy expenditure during exercise," *Int. J. Sports Med.*, vol. 20, no. 7, pp. 438–443, Oct. 1999.

[23] L. R. Keytel et al., "Prediction of energy expenditure from heart rate monitoring during submaximal exercise," *J. sports Sci.*, vol. 23, no. 3, pp. 289–297, Mar. 2005.

[24] K. L. Rennie, S. J. Hennings, J. Mitchell, and N. J. Wareham, "Estimating energy expenditure by heart-rate monitoring without individual calibration," *Med. Sci. Sports Exerc.*, vol. 33, no. 6, pp. 939–945, Jun. 2001.

[25] R. Doshmanziari, H. S. Aandahl, M. L. Danielsson, J. K. Baumgart, and D. Varagnolo, "Experiment design considerations for estimating energy expenditure during wheelchair propulsion," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 6504–6509, 2023. [Online]. Available: https://folk.ntnu.no/royado/Papers/IFACWC2023.pdf

[26] Y. C. Learmonth, D. Kinnett-Hopkins, I. M. Rice, J. L. Dysterheft, and R. W. Motl, "Accelerometer output and its association with energy expenditure during manual wheelchair propulsion," *Spinal Cord*, vol. 54, no. 2, pp. 110–114, Feb. 2016.

[27] T. E. Nightingale, J.-P. Walhin, D. Thompson, and J. L. J. Bilzon, "Influence of accelerometer type and placement on physical activity energy expenditure prediction in manual wheelchair users," *PLoS ONE*, vol. 10, no. 5, May 2015, Art. no. e0126086.

[28] O. Kossi, J. Lacroix, B. Ferry, C. S. Batcho, A. Julien-Vergonjanne, and S. Mandigout, "Reliability of ActiGraph GT3X+ placement location in the estimation of energy expenditure during moderate and high-intensity physical activities in young and older adults," *J. Sports Sci.*, vol. 39, no. 13, pp. 1489–1496, Jul. 2021.

[29] X. García-Massó, P. Serra-Añó, L. M. García-Raffi, E. A. Sánchez-Pérez, J. López-Pascual, and L. M. Gonzalez, "Validation of the use of actigraph GT3X accelerometers to estimate energy expenditure in full time manual wheelchair users with spinal cord injury," *Spinal Cord*, vol. 51, no. 12, pp. 898–903, Dec. 2013.

[30] M. Altini, J. Penders, R. Vullers, and O. Amft, "Estimating energy expenditure using body-worn accelerometers: A comparison of methods, sensors number and positioning," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 219–226, Jan. 2015.

[31] P. H. Lee, D. J. Macfarlane, T. Lam, and S. M. Stewart, "Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review," *Int. J. Behav. Nutrition Phys. Activity*, vol. 8, no. 1, pp. 1–11, Dec. 2011.

[32] A. Bauman et al., "Progress and pitfalls in the use of the international physical activity questionnaire (IPAQ) for adult physical activity surveillance," *J. Phys. Activity Health*, vol. 6, no. 1, pp. 5–8, Jan. 2009.

[33] J. B. D. V. Weir, "New methods for calculating metabolic rate with special reference to protein metabolism," *J. Physiol.*, vol. 109, nos. 1–2, pp. 1–9, Aug. 1949.

[34] T. E. Nightingale, J.-P. Walhim, D. Thompson, and J. L. J. Bilzon, "Predicting physical activity energy expenditure in manual wheelchair users," *Med. Sci. Sports Exerc.*, vol. 46, no. 9, pp. 1849–1858, 2014.

[35] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[36] O. Rindal, T. Seeberg, J. Tjønnås, P. Haugnes, and Ø. Sandbakk, "Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data," *Sensors*, vol. 18, no. 1, p. 75, Dec. 2017. [Online]. Available: https://www.mdpi.com/1424-8220/18/1/75

[37] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020.

[38] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.

[39] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*, vol. 1. Berlin, Germany: Springer, 2006.

[40] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.

[41] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[42] J. Brownlee, *XGBoost With Python: Gradient Boosted Trees With XGBoost and Scikit-Learn*. Machine Learning Mastery, 2016.