

GMAC—A Simple Measure to Quantify Upper Limb Use From Wrist-Worn Accelerometers

Sivakumar Balasubramanian¹, *Member, IEEE*

Abstract—Various measures have been proposed to quantify upper-limb use through wrist-worn inertial measurement units. The two most popular traditional measures of upper-limb use – thresholded activity counts (TAC) and the gross movement (GM) score suffer from high sensitivity and low specificity, and vice versa. We previously proposed a hybrid version of these two measures – the GMAC – that showed better overall detection performance than TAC and GM. In this paper, we answer two critical questions to improve the GMAC measure’s usefulness: (a) can it be implemented using only the accelerometer data? (b) what are its optimal parameter values? Here, we propose a modified GMAC using only the accelerometer data and optimize its parameters to develop: (a) a generic measure that is both limb- and subject-independent, and (b) limb-specific measures that were only subject-independent. The optimized GMAC showed better detection performance than the previous GMAC and surprisingly had comparable performance to the best-performing machine learning-based measure (random forest inter-subject model). In hemiparetic data, its performance was similar to the previous GMAC and the random forest inter-subject model; the limb-specific GMAC measure, however, had a better performance than the generic measure. The optimized limb-specific GMAC is a simple, interpretable alternative to a machine learning-based inter-subject model. The optimized GMAC can be a valuable measure for offline or real-time detection and feedback of upper limb use. The preliminary results of this study, based on a small dataset, need to be validated on a larger dataset to evaluate its generalizability.

Index Terms—Neurorehabilitation, upper-limb assessment, wearable sensor, upper-limb functioning, accelerometers, inertial measurement units.

I. INTRODUCTION

THERE is a growing interest in using wearable sensors for tracking upper limb (UL) movement behavior to quantify participation in daily life [1], [2], [3]. This interest is fuelled by the need to go beyond conventional measures that

Manuscript received 27 November 2023; revised 22 May 2024; accepted 19 June 2024. Date of publication 24 June 2024; date of current version 15 July 2024. This work was supported in part by the Fluid Research Grant from CMC Vellore under Grant IRB Min. 11303; in part by the National Hub for Healthcare Instrumentation Development, Anna University, under Grant TDP/BTD/11/2018/General; in part the Joint Grant through the “Artificial Intelligence for Advancing Healthcare across India and Sweden”, Department of Biotechnology, India, under Grant BT/PR40873/Swdn/135/8/2020; and in part by VINNOVA, Sweden, under Grant 2020-03619.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by CMC Vellore’s Institutional Review Board under IRB Min. No. 12321.

The author is with the Department of Bioengineering, Christian Medical College, Vellore 632002, India, and also with the School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, QLD 4072, Australia (e-mail: siva82kb@cmcvellore.ac.in).

Digital Object Identifier 10.1109/TNSRE.2024.3417964

rely on self-reported questionnaires and patient interviews [4], which lack objectivity and sensitivity. An ideal system for this assessment must consist of (a) an unobtrusive measurement system to seamlessly record movement-related information, and (b) an automated data analytics pipeline to extract the relevant information. Such a system can provide accurate and reliable quantitative answers to clinically relevant questions, such as, how much the two ULs are used, how symmetric is their use, how “good” are the movements, etc.

Micro-electromechanical systems-based inertial measurement unit (IMU) is a compact, wearable sensor that measures linear acceleration and angular velocity of the rigid body to which it is attached. Ideally, a single sensor per limb is preferred to track UL movements in daily life. The most popular choice for sensor location is around the distal forearm just proximal to the wrist joint [4], [5], [6], [7], [8], for the following reasons: (a) the forearm’s linear and angular kinematics are sensitive to both shoulder and elbow movements, (b) this location has the largest moment arm about the shoulder and elbow joints, thus, registering relatively large linear acceleration signals resulting from shoulder/elbow joint rotations, and (c) the ease of donning/doffing the sensor on this location,

The most fundamental construct of interest in UL functioning is *UL use* [4]. This is a binary construct indicating the presence or absence of a voluntary, meaningful UL movement or posture [9]. An accurate estimation of this complex construct requires access to information on the complete UL kinematics and kinetics, and the context in which the movement/posture is performed. In practice, a single wrist-worn IMU only provides the linear acceleration \mathbf{a}_S and angular velocities of the ω_S of the forearm in the local sensor reference frame, which is problematic for multiple reasons: (a) it cannot dissociate useful shoulder-elbow movements from unwanted movements, such as whole body movements, (b) it cannot detect finger movements, (c) it cannot ascertain if a movement/posture is voluntary, and (d) it is devoid of contextual information.

Nevertheless, several measures have been proposed in the literature to detect UL use from a single IMU [5], [7], [8], [10]. These measures can be broadly categorized into traditional [5], [10], [11] and machine learning (ML)-based measures [6], [10]; we use the terms measure, model, and algorithm interchangeably in the rest of the manuscript. The traditional measures are simple, hand-crafted algorithms with pre-specified parameter values that use specific signal features to detect UL use. For instance, the thresholded activity counts (TAC) measure [5] uses the magnitude of the gravity-subtracted acceleration, while the gross movement (GM) score [4], [8] uses the

orientation of the forearm and the amount of forearm movement. On the contrary, ML-based measures are algorithms trained on a set of labeled data to detect UL use from IMUs. Random forests, support vector machines, and multilayer perceptrons have been reported previously for detecting UL use [6], [10], with the random forests [6], [10] offering the best performance to date. Additionally, intra-subject (i.e. subject-specific) ML models perform better than inter-subject (i.e. one model trained across different participants) models [6], [10]. Although the ML-based measures perform better than the traditional methods, the latter has some advantages, such as: (a) they are simple and easy to interpret, and (b) they can be implemented efficiently in firmware for real-time detection and feedback of UL use (e.g. like the step count feedback from pedometers).

The TAC and the GM measures are the two most popular measures for quantifying UL use. Previous studies have shown that the TAC is a highly sensitive measure, while the GM is highly specific [10], [12]. We recently proposed a hybrid measure, called the *GMAC*, that combines the TAC and GM measures to balance out their respective high sensitivity and specificity [10]. The GMAC showed a better overall performance than TAC or GM, as quantified using the Youden index, but had a lower performance than the inter- and intra-subject ML measures. Our previous work also showed that the best-performing ML measures used the mean and variance of the accelerometer signal to detect UL use; interestingly, the accelerometer’s mean and variance are related to the orientation of the forearm (used by the GM), and the variance is related to the amount of forearm movement (related to the GM and TAC). Thus, in principle, the GMAC and ML measures use similar information but different decision boundaries for deciding UL use. Given, that GMAC is a simple and reasonable alternative to the ML measures, a more detailed investigation of the GMAC algorithm and the optimization of its parameters to work effectively for both healthy and hemiparetic participants are warranted. Thus, this study aimed to find answers to two important questions about the GMAC algorithm:

- 1) Can the GMAC algorithm be implemented using only a wrist-worn accelerometer? This is an important question because: (a) some popular wearable sensors (e.g. from ActiGraph, USA) only contain an accelerometer, (b) gyroscopes do not add any values to UL use detection [10], and (c) gyroscopes are power-hungry sensors, and avoiding them can result in more efficient UL-use trackers.
- 2) What are the optimal parameters for the GMAC algorithm that work well for both healthy and hemiparetic participants? The parameters of the GMAC algorithm were previously chosen based on TAC and GM, which might not be optimal.

The paper starts with a description of the GMAC algorithm proposed by Subash et al. [10], followed by the description of the newly proposed GMAC that works only with accelerometer data. The optimization of the parameters of this new algorithm and its comparison with existing algorithms is presented subsequently. The paper ends with a discussion of

TABLE I
DETAILS OF THE HEALTHY AND HEMIPARETIC PARTICIPANTS THAT PARTICIPATED IN THE STUDY [12]

Subject	Age/Sex	Condition	Paretic side	Time since injury
H1	33, F	-	-	-
H2	24, M	-	-	-
H3	24, M	-	-	-
H4	23, F	-	-	-
H5	24, M	-	-	-
H6	23, M	-	-	-
H7	22, M	-	-	-
H8	22, M	-	-	-
H9	25, F	-	-	-
H10	25, F	-	-	-
P1	30-40, M	TBI	Right	204
P2	40-50, M	CVA	Left	3
P3	30-34, M	TBI	Right	12
P4	<30, F	TBI	Left	7
P5	50-60, M	CVA	Right	3

its results, its implications for clinical use, and the limitations of the current study.

II. METHODS

This work used data from our previous study [12] which is openly available as part of a Github repository. The data was collected from 10 healthy and 5 hemiparetic participants, using a custom-built wearable IMU sensor that samples accelerometer and gyroscope data at 50 Hz. Each IMU sensor contained an SEN-14001 board (Spark Fun Inc.) with a SAMD21 microprocessor, a real-time clock, a 9-DOF IMU (MPU9250, InvenSense-TDK Co.), a MicroSD card slot, and a battery charging circuit. The IMU data and the real-time clock’s timestamp were logged at 50 Hz to an 8 Gb microSD card. Each participant wore two IMU sensors one on each arm – whose real-time clocks were synchronized to GMT+5.5h. The details of the 10 healthy and 5 hemiparetic participants are provided in Table I.

The participants performed various tabletop and non-tabletop tasks (listed in Table II) chosen from the Motor Activity Log [13] while wearing the IMU sensors on both forearms, proximal to the wrist joints. These tasks were chosen such that they included both functional and non-functional movements of the upper limbs. The movements performed by these participants were simultaneously recorded using a video camera connected to a PC that was time-synchronized with the IMU sensors. A custom software (using OpenCV and Python) was written to record the video data along with its PC’s timestamp. Two trained occupational therapists annotated the recorded videos to label UL use employing the Functional Arm Activity Behavioural Observation System framework (FAABOS) [14]. More details about the data and the protocol can be found in [10], [12].

A. The Previous GMAC Measure

The accelerometer and gyroscope signals are given by $\mathbf{a}_S[n]$ and $\boldsymbol{\omega}_S[n]$, respectively, at the sampling time instant $n \in \mathbb{Z}$; both signals are sampled at $f_s = 50\text{Hz}$. The previously proposed GMAC measure (referred to as the “Old-GMAC”) computes the UL use u_{gmac}^{old} every second (i.e. every f_s samples) using the activity counts α_{vm} obtained from the

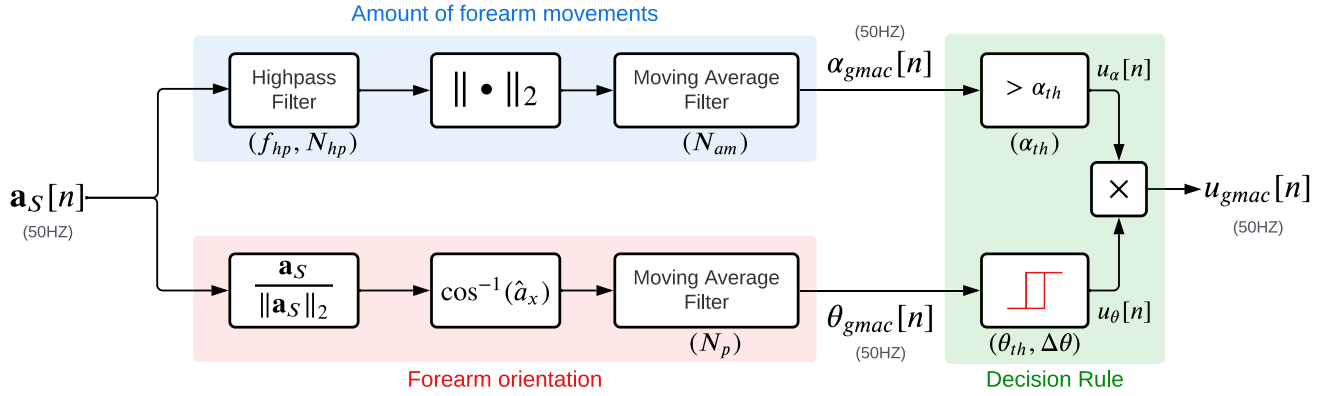


Fig. 1. Schematic of the proposed GMAC algorithm to work with only accelerometer data. The proposed algorithm has three subblocks: (a) Forearm orientation (red background), (b) Amount of forearm movements (blue background), and (c) Decision rule (green background). The different parameters associated with the three blocks are shown in green colored text in the figure.

TABLE II

LIST OF TASKS PERFORMED BY THE PARTICIPANTS IN [12]

Tasks performed	
Eat from a bowl using a spoon	Turn on a switch
Eat from a plate using finger	Walk 25m
Open a bottle and pour water to a glass	Open a door and move in
Drink from a glass	Write using a pen
Button a shirt	Comb your hair
Drink from a cup holding its handle	Type on a keyboard
Fold a towel	Call using from a mobile
Wipe off a tabletop	

vector magnitude algorithm [5], and the mean pitch angle $\bar{\theta}_p$ of the forearm,

$$u_{gmac}^{old}[k] = \begin{cases} 1, & \alpha_{vm}[k] > 0 \text{ \& } |\bar{\theta}_p[k]| < 30 \text{ deg} \\ 0, & \text{Otherwise} \end{cases}$$

$$\bar{\theta}_p[k] = \frac{1}{f_s} \sum_n \theta_p[n], \quad (k-1) \cdot f_s < n \leq k \cdot f_s \quad (1)$$

where, $k \in \mathbb{Z}$ represents the sampling time instants of u_{gmac}^{old} , α_{vm} , and $\bar{\theta}_p$ (all computed every 1s), $\alpha_{vm}[k]$ and $\bar{\theta}_p[k]$ are the output of the vector magnitude algorithm and the mean forearm pitch angle using the IMU data over the time window where $(k-1) \cdot f_s < n \leq k \cdot f_s$, respectively. This Old-GMAC algorithm uses both the accelerometer and the gyroscope signals to compute α_{vm} and $\bar{\theta}_p$ [10].

B. The New GMAC Measure: Using Only the Accelerometer Data

If we only had the acceleration data $\mathbf{a}_S[n]$, we could still estimate information about the amount of forearm movement $\alpha_{gmac}[n]$ and forearm orientation $\theta_{gmac}[n]$. A block diagram representation of the estimation procedure for $\alpha_{gmac}[n]$ and $\theta_{gmac}[n]$ from $\mathbf{a}_S[n]$ is shown in Figure 1, which also shows the various associated parameters (details in Table III). The forearm orientation is computed as the *arccos* of the normalized component of the acceleration signal along the length of the forearm (which is taken as the x axis in Figure 1). The amount of forearm movement is computed by first highpass filtering the accelerometer data to remove the slow varying contribution from gravity, followed by computing the 2-norm (Figure 1). Both of these signals are smoothed using moving

average filters. The decision rule consists of two rules as shown in Eq. 1 for detecting UL use,

$$u_{gmac}[n] = u_\alpha[n] \cdot u_\theta[n] \quad (2)$$

where, $u_\alpha[n], u_\theta[n] \in \{0, 1\}$ are obtained through thresholding rules applied on $\alpha_{gmac}[n]$ and $\theta_{gmac}[n]$, respectively, similar to Eq. 1. The thresholding rule on $\alpha_{gmac}[n]$ is given as follows,

$$u_\alpha[n] = \begin{cases} 1, & \alpha_{gmac}[n] > \alpha_{th} \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

While, the second thresholding rule on $\theta_{gmac}[n]$ is a hysteresis rule, where the output at the time instant n depends on the current input $\theta_{gmac}[n]$ and the past value of the output $u_\theta[n-1]$,

$$u_\theta[n] = \begin{cases} 1, & \theta_{gmac}[n] > \theta_{th} \\ 0, & \theta_{gmac}[n] < \theta_{th} - \Delta\theta \\ u_\theta[n-1], & \text{Otherwise} \end{cases} \quad (4)$$

Figure 2 depicts the above hysteresis rule, where the shaded red region represents the range of forearm pitch angles where the previous state of the output is retained ($u_\theta[n] = u_\theta[n-1]$). Forearm pitch angles above θ_{th} are considered as UL use ($u_\theta[n] = 1$), while angles below $\theta_{th} - \Delta\theta$ are considered as no UL use $u_\theta[n] = 0$.

The choice of a simple thresholding rule for $\alpha_{gmac}[n]$ and a hysteresis rule for $\theta_{gmac}[n]$ was based on preliminary optimization work, indicating that a hysteresis rule on $\alpha_{gmac}[n]$ did not improve the detection performance. Eqs. 2, 3, and 4 constitute the new GMAC algorithm that uses only the accelerometer data $\mathbf{a}_S[n]$ to detect UL use. The optimization of the parameters associated with the different components of this algorithm is described in the next section.

C. Optimization of the New GMAC Parameters

The optimization of these parameters was carried out through a grid search approach with the parameter ranges for the search shown in Table III. The choice of the parameter ranges was based on what we believed to be a reasonable range for the parameters:

TABLE III

DESCRIPTION OF THE DIFFERENT PARAMETERS OF THE PROPOSED GMAC ALGORITHM DEPICTED IN FIGURE 1

GMAC Algorithm Subblock	Parameter	Description	Grid Search Range
Amount of Forearm Movement	f_{hp}	Cut-off frequency of the highpass filter.	$\{0.01, 0.1, 1\}$ Hz
	N_{hp}	Order of the highpass filter.	$\{1, 2\}$
	N_{am}	Window size of the moving average filter.	$\{1, 25, 50\}$
Forearm Orientation	N_p	Window size of the moving average filter for estimating the pitch angle of the forearm.	$\{1, 25, 50\}$
GMAC Decision Rule	α_{th}	Threshold for the amount of forearm movement.	$\{0, 0.1\}$
	θ_{th}	Pitch angle threshold.	$\{-90, -80, \dots, +90\}$ deg
	$\Delta\theta$	Hysteresis size on the pitch angle.	$\{0, 20, 40, \dots, 80\}$ deg

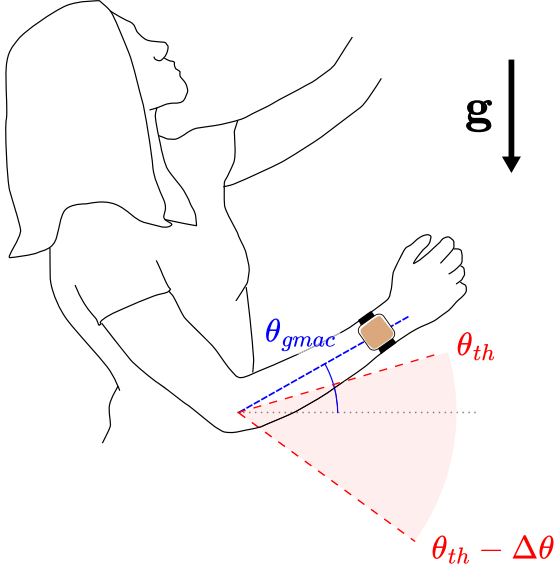


Fig. 2. Depiction of the pitch threshold rule with hysteresis. The shaded red region represents the range of forearm pitch angles where the previous state of the output is retained ($u_\theta[n] = u_\theta[n-1]$). The pitch angles above θ_{th} are considered as UL use ($u_\theta[n] = 1$), while angles below $\theta_{th} - \Delta\theta$ are considered as no UL use $u_\theta[n] = 0$. g represents the acceleration due to gravity, the brown colored box on the forearm represents the IMU sensor, and the angle of the blue dashed line (forearm axis) represents the pitch angle of the forearm. In this particular example, $\theta_{gmac}[n] > \theta_{th} \implies u_\theta[n] = 1$.

- $f_{hp} \in \{0.01, 0.1, 1\}$ Hz: Frequencies below 0.01 Hz are most likely to correspond to static postures, and movements with frequency components 1 Hz are likely to correspond to movements of interest.
- $N_{hp} \in \{1, 2\}$ were chosen to keep the high pass filter simple.
- $N_p, N_{am} \in \{1, 25, 50\}$ were chosen to limit the window size of the moving average filters to less than 1 s, to avoid the past beyond 1 sec from influencing the upper-limb use output. The sampling frequency is $f_s = 50$ Hz.
- $\alpha_{th} \in \{0, 0.1, 0.25\}$: This threshold was based on the values previously used in the GMAC algorithm [10].
- $\theta_{th} \in \{-90, -80, \dots, 90\}$ deg covers the entire range of possible angles of interest.
- $\Delta\theta \in \{0, 20, \dots, 80\}$ deg covers a wide enough range of hysteresis values between 0 and 90.

For each parameter combination \mathbf{p} , the Youden index [15] was computed for each limb for each subject (healthy participants and patients), as the following:

$$J(\mathbf{p}) = \text{Sensitivity}(\mathbf{p}) + \text{Specificity}(\mathbf{p}) - 1$$

where, $\mathbf{p} = [f_{hp}, N_{hp}, N_{am}, N_p, \alpha_{th}, \theta_{th}, \Delta\theta]^T$ is a parameter combination, the sensitivity and specificity are computed from the confusion matrix generated from the UL use detected using the new GMAC algorithm for the given parameter combination u_{gmac} (Figure 1), and the ground truth obtained from the FAABOS framework.

The optimal parameter combination for the new GMAC algorithm was defined as the one that maximizes the overall detection accuracy, consistently (in terms of the Youden index). This was defined as the following optimization problem,

$$\mathbf{p}^* = \arg \max_{\mathbf{p}} f(\mathbf{p})$$

$$f(\mathbf{p}) := J_{50}(\mathbf{p}) \cdot [1 - (J_{75}(\mathbf{p}) - J_{25}(\mathbf{p}))] \quad (5)$$

where, $f(\mathbf{p})$ is the performance measure of a particular parameter combination \mathbf{p} , $J_q(\mathbf{p})$ is the q^{th} percentile of the Youden index computed for a given parameter combination \mathbf{p} , and \mathbf{p}^* is the optimum parameter combination. The median Youden index $J_{50}(\mathbf{p})$ in $f(\cdot)$ is a measure of the detection accuracy, while the term $[1 - (J_{75}(\mathbf{p}) - J_{25}(\mathbf{p}))]$ is a measure of the consistency of the detection accuracy.

The new GMAC algorithm's parameters were optimized for developing two types of models:

- A **single generic model** was obtained by maximizing $f(\mathbf{p})$ on the entire dataset (15 participants) involving both limbs of healthy and hemiparetic patients. This model will be referred to as the “generic” model in the rest of the manuscript. Although simple, a single generic model can miss inter-subject and inter-limb differences and thus compromise performance.
- **Limb specific models** obtained by maximizing $f(\mathbf{p})$ for each limb of the 10 healthy (right and left), and 5 hemiparetic (affected and unaffected) participants. This results in four models corresponding to the two limbs of healthy and hemiparetic patients. These models will be referred to as “limb-specific” models. Notice that these are still inter-subject models that are limb-specific; a single model is still employed across participants for each limb. This approach is expected to perform better than the single generic model by accounting for inter-limb differences.

Leave one subject out validation: An estimate of the expected performance of the optimized GMAC models was computed by employing a leave-one-subject-out cross-validation

TABLE IV

OPTIMAL PARAMETER COMBINATION FOR THE GMAC ALGORITHM FOR THE DIFFERENT MODELS. THE OPTIMAL PARAMETER COMBINATION IS THE ONE THAT MAXIMIZES THE PERFORMANCE MEASURE $f(\rho)$ DEFINED IN EQ. 5. THE FIRST ROW CORRESPONDS TO THE SINGLE MODEL FOR ALL PARTICIPANTS, WHILE THE REST FOUR ARE MODELS TRAINED FOR THE FOUR LIMBS - LEFT AND RIGHT FOR THE 10 HEALTHY PARTICIPANTS, AND UNAFFECTED AND AFFECTED LIMBS FOR THE 5 PATIENTS

	N_p	f_{hp}	N_{hp}	N_{am}	α_{th}	θ_{th}	$\Delta\theta$	$f(\mathbf{p}^*)$
Generic	50	0.01 Hz	2	25	0.1	10°	40°	0.409
Right	50	0.1 Hz	1	50	0.1	10°	20°	0.550
Left	50	0.01 Hz	2	25	0.1	10°	40°	0.431
Unaff	1	0.1 Hz	2	50	0.1	20°	60°	0.620
Aff	1	0.1 Hz	2	25	0.1	20°	40°	0.429

approach. This cross-validation approach was slightly different for the generic and the limb-specific models.

- 1) **Validation of the generic model:** We have a total of 15 participants. The grid search optimization process was performed on 14 participants by randomly leaving out one participant $i \in \{1, 2, \dots, 15\}$; let the optimal parameter when leaving out the i^{th} participant be \mathbf{p}_i^* . The Youden index $J(\mathbf{p}_i^*)$ of these 15 optimal parameters $\{\mathbf{p}_i^*\}_{i=1}^{15}$ were obtained to compute the expected Youden index from generic optimal GMAC model on unseen data.
- 2) **Validation of limb-specific models:** The procedure is similar to that of the generic model, except that limb-specific models for the left and right arms result in 10 optimal parameter sets (corresponding to leaving one of the 10 healthy participants out), and 5 estimates for affected and unaffected limb-specific models (corresponding to the 5 patients). The Youden indices were similarly computed for the optimal parameters identified for each of the four limb-specific by leaving a subject out.

The expected Youden index obtained from the cross-validation procedure for the different models was compared against that of the Old-GMAC, inter-subject random forest (RF-Inter), and intra-subject random forest (RF-Intra) measures from our previous work [10], through linear mixed effects models; the statistical significance was set at $p < 0.05$.

The data used in the current study was made available as part of our previously published work in [10]. This data can be found at Upper-limb Assessment GitHub repository. The code used in the current study is available at GMAC GitHub repository.

III. RESULTS

All analyses in this work were carried out in Python using the Jupyter Notebook environment [16] and the linear mixed-effects modeling was performed using the ‘statsmodels’ package [17]. The results for the generic model are presented first followed by that of the limb-specific models.

A. Performance of the Generic Model

The optimal parameter combination for the generic model is shown in the first row of Table IV which results in the

TABLE V

COMPARISON OF THE YODEN INDEX, SENSITIVITY, AND SPECIFICITY OF THE OPTIMIZED GENERIC GMAC WITH THE OLD-GMAC, RF-INTER, AND RF-INTRA MEASURES FROM SUBASH ET AL. [10]. THE MEAN DIFFERENCES WERE OBTAINED THROUGH A LINEAR MIXED EFFECT MODEL WITH THE DIFFERENT MEASURES AS THE FIXED EFFECT, AND THE PARTICIPANTS AS A RANDOM EFFECT. THE CELLS HIGHLIGHTED IN LIGHT RED INDICATE NON-SIGNIFICANT DIFFERENCES

All participants			
Comparison with Subash et al. 2022	Youden	Sensitivity	Specificity
Old GMAC	+0.129	+0.079	+0.05
RF Inter	-0.012	-0.035	+0.023
RF Intra	-0.254	-0.08	-0.174
Healthy			
Old GMAC	+0.192	+0.071	+0.121
RF Inter	-0.005	-0.092	+0.088
RF Intra	-0.238	-0.148	-0.09
Patients			
Old GMAC	+0.003	+0.095	-0.091
RF Inter	-0.027	+0.078	-0.105
RF Intra	-0.285	+0.058	-0.343

maximum performance $f(\mathbf{p}^*) = 0.409$ across both limbs of the 10 healthy and 5 hemiparetic participants.

Figure 3(A) shows the comparison of the Youden index of the optimized GMAC model to that of the three different measures from [10]: the Old-GMAC, the RF-Inter and RF-Intra models. The mean Youden indices and the 95% confidence interval shown in Figure 3(A) are bootstrap estimates obtained for both limbs across all participants. The figure indicates that the optimized generic GMAC is better than the Old-GMAC measure, but is not different from the RF-Inter subject model. Figure 3(B) shows the receiver operating characteristics plot depicting the sensitivity and specificity of the optimized GMAC algorithm and the three measures from Subash et al. [10]. Figure 3(C) and (D) show the scatter plot of the sensitivity and specificity of the optimized GMAC algorithm for different parameter combinations for the healthy and hemiparetic participants, respectively. The mean and 95% confidence interval estimated through a bootstrap procedure for the sensitivity and specificity are shown in these plots.

The mean differences in the Youden index, sensitivity, and specificity between the optimized GMAC and the other three measures from [10] are shown in Table V; these values were obtained through a linear mixed effect model with the different measures as the fixed effect, and the participants as a random effect. Separate linear mixed-effects models were fit for all participants, healthy participants alone, and hemiparetic participants alone (Table V). The cells highlighted in light red indicate non-significant differences. The table reveals the following:

- 1) When we consider all participants (10 healthy and 5 hemiparetic) or just the healthy participants, the generic optimized GMAC measure has a significantly greater Youden index (Table V) than Old-GMAC and is not different from the RF-Inter subject model.
- 2) For hemiparetic participants, optimized GMAC is not significantly different from both the Old-GMAC and RF-Inter participants models.

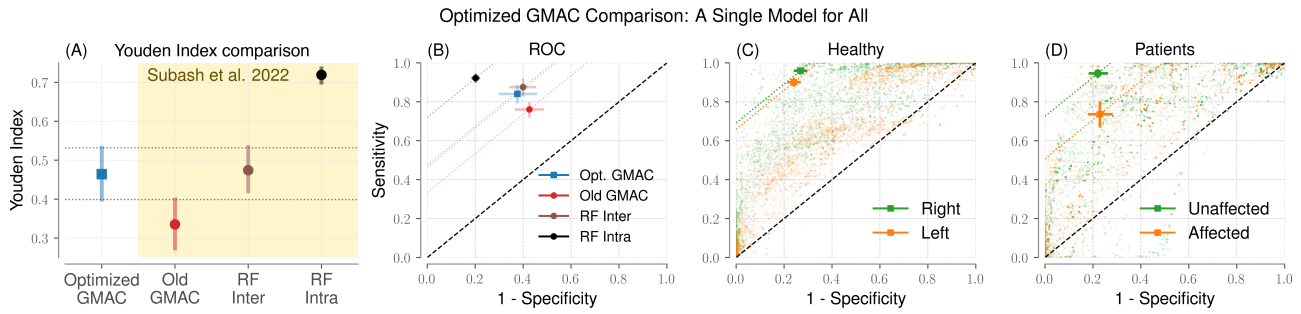


Fig. 3. Comparison of the Youden index of the generic model with that of the Old-GMAC, RF-Inter, and RF-Intra measures from Subash et al. [10]. (A) Plot of the bootstrap estimates of the mean and its 95% confidence interval of the Youden index for the different measures. (B) The receiver operating characteristic plot depicts the sensitivity and specificity of the generic model and the three measures from Subash et al. [10]. The dashed black line represents the performance of a random classifier with a Youden index of 0. The lighter-colored dotted lines passing through the markers represent the constant Youden index line corresponding to the different measures. The closer the dotted line is to the top-left corner, the higher its Youden index. (C) The scatter plot of the sensitivity and specificity of the proposed GMAC algorithm for different parameter combinations \mathbf{p} for the healthy participants. The scatter plots in green and orange represent the data for the right and left limbs, respectively, along with the mean and 95% confidence interval of the sensitivity and specificity. (D) The scatter plot of the sensitivity and specificity of the proposed GMAC algorithm for different parameter combinations \mathbf{p} for the hemiparetic participants. The scatter plots in green and orange represent the data for the unaffected and affected limbs, respectively, along with the mean and 95% confidence interval of the sensitivity and specificity.

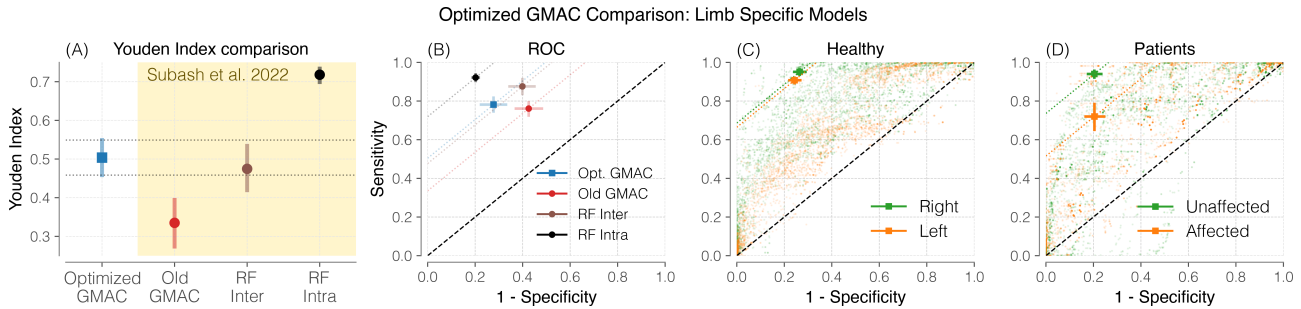


Fig. 4. Comparison of the Youden index of limb-specific model with that of the Old-GMAC, RF-Inter, and RF-Intra measures from Subash et al. [10]. (A) Plot of the bootstrap estimates of the mean and its 95% confidence interval of the Youden index for the different measures. (B) The receiver operating characteristic plot depicts the sensitivity and specificity of the limb-specific model and the three measures from Subash et al. [10]. The dashed black line represents the performance of a random classifier with a Youden index of 0. The lighter-colored dotted lines passing through the markers represent the constant Youden index line corresponding to the different measures. The closer the dotted line is to the top-left corner, the higher is its Youden index. (C) The scatter plot of the sensitivity and specificity of the proposed GMAC algorithm for different parameter combinations \mathbf{p} for the healthy participants. The scatter plots in green and orange represent the data for the right and left limbs, respectively, along with the mean and 95% confidence interval of the sensitivity and specificity. (D) The scatter plot of the sensitivity and specificity of the proposed GMAC algorithm for different parameter combinations \mathbf{p} for the hemiparetic participants. The scatter plots in green and orange represent the data for the unaffected and affected limbs, respectively, along with the mean and 95% confidence interval of the sensitivity and specificity.

- 3) The optimized GMAC measure is significantly worse than the RF-Intra model, across all participants, only healthy or only hemiparetic participants.

B. Performance of the Limb-Specific Models

The optimal parameter combination for the four limb-specific models is shown Table IV (last four rows); note the maximum performance (last column of Table IV) for all four limb-specific models is greater than that of the generic model. The models for healthy participants (left and right) have similar optimal parameter combinations to each other and to that of the generic model. However, these are different for the hemiparetic participants compared to those of the models for healthy participants and the generic model. The major differences are the moving average filter N_p , the pitch threshold θ_{th} and pitch hysteresis $\Delta\theta$.

Figure 4(A) shows the comparison of the Youden index of the limb-specific optimized GMAC models, similar to Figure 3. The figure indicates that the limb-specific optimized GMAC is better than the Old-GMAC measure, but is not different from the RF-Inter subject model; it has a slightly higher mean Youden index than the generic

model. Figure 4(B) depicts the sensitivity and specificity of the limb-specific GMAC algorithm and the three measures from [10]. Figure 4(C) and (D) show the scatter plot of the sensitivity and specificity of the optimized GMAC algorithm for different parameter combinations \mathbf{p} for the healthy and hemiparetic participants, respectively.

Table VI shows similar results to that of Table V, and the main results are similar to that of the generic model.

- 1) When considering all participants or just healthy participants, the limb-specific optimized GMAC measure has a significantly greater Youden index (Table VI) than the Old-GMAC and is not different from the RF-Inter subject model.
- 2) Interestingly, for patients, the limb-specific optimized GMAC is still not significantly different from the Old-GMAC and RF-Inter participants models. Although the mean value of the Youden index is higher for the limb-specific models (0.119), compared to the generic model (0.003).
- 3) The RF-Intra measure performs better than limb-specific optimized GMAC.

TABLE VI

COMPARISON OF THE YODEN INDEX, SENSITIVITY, AND SPECIFICITY OF THE OPTIMIZED LIMB-SPECIFIC GMAC WITH THE OLD-GMAC, RF-INTER, AND RF-INTRA MEASURES FROM SUBASH ET AL. [10]. THE MEAN DIFFERENCES WERE OBTAINED THROUGH A LINEAR MIXED EFFECT MODEL WITH THE DIFFERENT MEASURES AS THE FIXED EFFECT, AND THE PARTICIPANTS AS A RANDOM EFFECT. THE CELLS HIGHLIGHTED IN LIGHT RED INDICATE NON-SIGNIFICANT DIFFERENCES

All participants			
Comparison with Subash et al. 2022	Youden	Sensitivity	Specificity
Old GMAC	+0.170	+0.021	+0.148
RF Inter	+0.029	-0.093	+0.121
RF Intra	-0.213	-0.137	-0.076
Healthy			
Old GMAC	+0.195	+0.061	+0.134
RF Inter	-0.001	-0.102	+0.101
RF Intra	-0.235	-0.158	-0.077
Patients			
Old GMAC	+0.119	-0.058	+0.176
RF Inter	+0.089	-0.074	+0.163
RF Intra	-0.170	-0.095	-0.075

The individual Youden index for each participant for both limbs from the leave-one-subject-out cross-validation procedure for the generic and limb-specific models are shown in Table VII. The corresponding values for the three measures from [10] are also provided in the table for comparison. The corresponding mean (μ) and standard deviation (σ) of the Youden index for the healthy and hemiparetic participants for each limb are also shown in this table. In general, the right and the unaffected limbs have a higher mean Youden index than the left and affected limbs, respectively. The affected limb has the lowest mean Youden index compared to the other limbs.

IV. DISCUSSION

This paper presented preliminary work on a simple measure for quantifying UL use employing a wrist-worn accelerometer. The present work demonstrated that the GMAC measure can be computed entirely from the accelerometer data, and with optimal parameters, it performs better than the Old-GMAC measure [10] on the entire dataset (healthy and hemiparetic participants) in terms of the Youden index. Surprisingly, the performance of this optimized GMAC measure was as good as that of the RF-inter model [10]. However, when the hemiparetic data are analyzed separately, the optimized GMAC measure is not significantly different from the Old-GMAC measure or the RF-Inter (Table V). This is most likely due to the large inter-limb and inter-subject variability in the small hemiparetic dataset used in this study. Additionally, four limb-specific optimized GMAC models were developed to address the inter-limb differences between healthy and hemiparetic participants. The limb-specific models performed similarly to the generic optimized GMAC model with slightly better performance (although not significant) for the hemiparetic dataset. The RF-Intra model [10] outperformed both the generic and limb-specific optimized GMAC models. The results of this study have some important implications for the UL use detection problem and its use in clinical research and practice.

A. Optimized GMAC Versus Old GMAC

Given that the parameters of the new GMAC measure were optimized in this study, it is expected that both the generic and limb-specific optimized GMAC measure performed better than the Old-GMAC measure [10] on the entire dataset (healthy participants and patients) and healthy participants only. This improved Youden index was due to increased sensitivity (Table V and Table VI) without compromising specificity. There are two possible reasons for this improvement:

- 1) **Modified pitch angle thresholds.** It was previously noted that the GM pitch angle thresholds of $\pm 30^\circ$ might be conservative, as many functional movements require one to lift his/her forearm by more than $+30^\circ$ [12]. This range of $\pm 30^\circ$ originally proposed by Leuenberger et al. [8] was based only on visual observations of reaching movements and ADL. The proposed GMAC addresses this issue by making all pitch angles greater than θ_{ih} to be marked as functional if it satisfies the acceleration magnitude criterion (decision rule in Figure 1). This could have helped improve the sensitivity of the optimized GMAC compared to that of the Old-GMAC.
- 2) **Hysteresis in the pitch angle decision rule.** Another reason for the improved performance of the optimized GMAC could be the use of the hysteresis rule on the forearm pitch angle instead of the simple rule $|pitch| < +30^\circ$ (Eq. 1). The hysteresis adds memory to the pitch angle decision rule, which captures the intuition that if a UL is in a functional (or non-functional) state at the current time instant, it is likely to stay in that state unless there is a drastic change in its forearm pitch angle or the acceleration magnitude. All previously reported measures have been purely “feedforward” in nature, i.e., the current output does not impact the future output. It might be worth exploring the use of output feedback in other measures for UL use detection to improve their performance.

Nevertheless, the generic and limb-specific optimized GMAC measures were not significantly different from the Old-GMAC measure for the hemiparetic dataset. This could be attributed to the small size of the hemiparetic dataset and the large variability in the Youden index for the affected limb for the hemiparetic (Table VII). Although not significant, the limb-specific model has a higher mean Youden index (0.119 in Table VI) than the generic model (0.003 in Table V) for the hemiparetic dataset. This larger Youden index is due to a significant increase in the specificity of the measure compared to the Old-GMAC.

B. Optimized GMAC Versus Random Forest Models

The more interesting result of the current study is the similar performance of the optimized GMAC measure to that RF-Inter model from [10]; this was observed for the entire dataset, and for both healthy and hemiparetic dataset individually (Table V and Table VI). There are two potential explanations for this: (a) The proposed GMAC uses similar information to the two most important features identified for the RF-Inter model from [10]: mean x -component of the wrist

TABLE VII

YOUDEN INDEX OF THE DIFFERENT MEASURES FOR EACH PARTICIPANT FOR BOTH LIMBS. THE VALUES FOR THE GENERIC AND LIMB-SPECIFIC GMAC WERE OBTAINED FROM THE CURRENT STUDY, WHILE THE CORRESPONDING VALUES FOR THE OLD-GMAC, RF-INTER, AND RF-INTRA MEASURES ARE FROM SUBASH ET AL. [10]. THE MEAN (μ) AND STANDARD DEVIATION (σ) OF THE YOUDEN INDEX FOR THE HEALTHY AND HEMIPARETIC PARTICIPANTS FOR EACH LIMB ARE ALSO SHOWN

Subject	Right					Left				
	GMAC (Generic)	GMAC (Limb-Specific)	Old GMAC	RF Inter	RF Intra	GMAC (Generic)	GMAC (Limb-Spec.)	Old GMAC	RF Inter	RF Intra
H1	0.594	0.506	0.280	0.567	0.525	0.337	0.332	0.188	0.395	0.724
H2	0.406	0.600	0.213	0.482	0.679	0.269	0.455	0.245	0.419	0.792
H3	0.767	0.659	0.493	0.741	0.788	0.464	0.464	0.229	0.403	0.728
H4	0.700	0.570	0.484	0.704	0.806	0.419	0.415	0.145	0.369	0.707
H5	0.639	0.589	0.440	0.652	0.704	0.331	0.512	0.110	0.437	0.805
H6	0.458	0.396	0.250	0.617	0.756	0.443	0.496	0.321	0.442	0.724
H7	0.470	0.528	0.329	0.440	0.707	0.317	0.317	0.166	0.281	0.695
H8	0.593	0.393	0.457	0.534	0.724	0.410	0.500	0.179	0.411	0.677
H9	0.884	0.679	0.592	0.885	0.922	0.765	0.765	0.492	0.732	0.874
H10	0.444	0.458	0.377	0.417	0.737	0.323	0.464	0.196	0.198	0.722
$\mu(\sigma)$	0.60±0.16	0.54±0.10	0.39±0.12	0.60±0.15	0.73±0.10	0.41±0.14	0.47±0.12	0.23±0.11	0.41±0.14	0.74±0.06
	Unaffected					Affected				
S1	0.503	0.613	0.644	0.477	0.837	0.149	0.378	0.232	0.236	0.556
S2	0.651	0.674	0.462	0.574	0.765	0.220	0.196	0.101	0.202	0.507
S3	0.420	0.544	0.428	0.372	0.842	0.042	0.343	0.033	0.465	0.687
S4	0.465	0.656	0.687	0.392	0.810	0.546	0.460	0.260	0.574	0.816
S5	0.227	0.452	0.511	0.448	0.703	0.636	0.698	0.469	0.387	0.191
$\mu(\sigma)$	0.45±0.15	0.59±0.09	0.55±0.11	0.45±0.08	0.79±0.06	0.32±0.26	0.42±0.18	0.22±0.17	0.37±0.16	0.55±0.23

acceleration (related to forearm pitch) and the acceleration variance. (b) The average population-level UL use behavior has a relatively simple structure which is captured well by the simple GMAC algorithm proposed in the current study.

However, the RF-Intra model outperforms all the other models including the optimized GMAC, which is expected as the intra-subject models are tuned to each subject's characteristics. In general, the RF-Intra was found to have high sensitivity and specificity, except for the limb-specific optimized GMAC models for hemiparetic participants (Table VI).

C. Generic Versus Limb-Specific GMAC Models

The generic and limb-specific models investigated in the current study are inter-subject models. The generic model uses a single set of parameters for both limbs of all participants (healthy or hemiparetic). This is the simplest solution to the UL use detection problem requiring no subject-specific or limb-specific tuning. However, such models may have inferior performance than a subject-specific model due to their inability to capture inter-subject and inter-limb variability. The generic optimized GMAC model, which performed well on the entire dataset and healthy participants only, did not perform better than the Old-GMAC measure for the hemiparetic dataset. Table VII indicates that the Youden index is different for the affected and unaffected limbs (≈ 0.3) for hemiparetic participants, pointing to inter-limb differences that could have impacted the generic GMAC model's performance. The generic model performs better than the Old-GMAC model for the affected limb (0.32 versus 0.22) and worse for the unaffected limb (0.45 versus 0.55). This observation prompted the investigation of the limb-specific GMAC models, which were expected to perform better than the generic model, by capturing inter-limb differences.

The limb-specific model does not produce drastically different results from the generic model for healthy participants. This is not surprising since the optimal parameters for the right and left limb-specific models (Table IV) are similar to

that of the generic model. This similarity is explained by the large representation of healthy participants in the full dataset used for optimizing the generic model. On the contrary, for hemiparetic participants the limb-specific model results in a higher mean Youden index than the generic GMAC model. The optimal parameters for the affected and unaffected limb-specific models are very different from that of the generic model (Table IV). The pitch threshold θ_{th} is 20 deg for the affected and unaffected limbs, which is 10 deg higher than that of the healthy participants. This could be due to the annotating clinicians following a conservative strategy with hemiparetic participants when assigning functional status to a limb when it was in a non-functional state previously; they possibly required the arm to move by a larger amount than healthy individuals to deem it functional. The other big difference is the moving average filter size N_p , which was equal to 1 for the affected and unaffected limbs, i.e., no filtering is applied while computing the pitch angle compared to a 1sec (50 samples) long filter for the right and left limbs. This could be due to patients performing slow movements, especially moving the forearm against gravity, which removes the need for a filter for pitch angle estimation. The pitch hysteresis band $\Delta\theta$ was 40 deg and 60 deg for the affected and unaffected limbs, respectively (Table IV). This implies that when the affected and unaffected forearms drop below -20 deg and -40 deg, UL use is set to non-functional, respectively. This difference might be related to the relative use of the affected and unaffected limbs in patients. Due to reduced use of the affected limb, a smaller drop in forearm pitch angle is marked as non-functional, while a larger drop is required for the unaffected limb.

D. Where Will the GMAC Be Used?

The current study indicates that the optimized GMAC is a superior alternative to the existing traditional measures of UL use and the previously proposed by [10]. It is a simpler and easily implementable alternative to the random forest

inter-subject model, which is currently the best-performing ML model for population-level data [6], [10]. However, limb-specific optimized GMAC models are better alternatives to a generic optimized GMAC model. Notably limb-specific models are still inter-subject models and once trained they can be employed without any subject-specific tuning. The limb-specific models are a good compromise between a single, generic inter-subject model for both limbs of all participants and a fully personalized limb-specific intra-subject model.

Intra-subject ML algorithms produce better UL use detection accuracy than traditional approaches [6], [10]. If available, optimal trained intra-subject models are currently the best option for offline UL use detection from previously recorded wearable sensor data. However, such ML-based algorithms may not be best suited for real-time UL use detection and feedback. The proposed GMAC measure (Figure 1) is an attractive alternative when, (a) trained ML intra-subject models are unavailable, (b) there is no annotated dataset to train new ML models, or (c) if real-time detection of UL use is required in an application. The GMAC measure can be efficiently implemented in the wearable sensor to detect UL use and intermittently transmit average UL use information to a mobile app for regular feedback. Given that the GMAC only involves simple linear filtering and thresholding rules (Figure 1), it is well suited for highly efficient firmware-level implementation in a wearable device; we note that relatively simple random-forest algorithms are also amenable to efficient firmware implementation. Future work must explore these possibilities of using GMAC to provide regular feedback to patients about UL use, which could encourage a hemiparetic subject to incorporate their affected limb in daily life.

E. Limitation of the Current Study

The main limitation of the study is the limited size of data involving a small number of participants (10 healthy participants and 5 patients). Thus, the study outcomes should be considered preliminary. However, the outcomes warrant future validation with a large dataset involving more patients with a wide range of impairments, performing a broader set of tasks.

V. CONCLUSION

The paper demonstrated how the GMAC can be derived from just the accelerometer data and showed that an optimized choice of the GMAC measure's parameters leads to better performance than the Old-GMAC measure [10]. Surprisingly, the optimized GMAC had a similar performance to the random forest inter-subject measure [10], indicating that at the population level, the UL use behavior has a simple average structure. A generic GMAC model is a simple solution for detecting UL use, but limb-specific inter-subject models are a better alternative, especially for hemiparetic participants. Limb-specific optimized GMAC is a very attractive alternative when trained machine learning models are unavailable. The proposed GMAC algorithm can also be efficiently implemented in firmware for real-time detection and feedback of UL use, which is an important step towards encouraging UL use in hemiparetic patients. Future work involving a large dataset,

verifying the outcomes of the current study, and efficient real-time implementation and evaluation are warranted.

ACKNOWLEDGMENT

The authors would like to thank the members of the Biological Learning and Rehabilitation Group (CMC Vellore), Tanya Subash, and Sadhana Yadav for their critical comments and feedback about the manuscript.

REFERENCES

- [1] A. Kaku, A. Parnandi, A. Venkatesan, N. Pandit, H. Schambra, and C. Fernandez-Granda, "Towards data-driven stroke rehabilitation via wearable sensors and deep learning," in *Proc. Mach. Learn. Healthcare Conf.*, vol. 126, 2020, pp. 143–171.
- [2] I. Boukhenoufa, X. Zhai, V. Utti, J. Jackson, and K. D. McDonald-Maier, "Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103197. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1746809421007941>
- [3] D. Johansson, K. Malmgren, and M. A. Murphy, "Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: A mixed-methods systematic review," *J. Neurol.*, vol. 265, no. 3, pp. 1740–1752, 2018, doi: [10.1007/s00415-018-8786-y](https://doi.org/10.1007/s00415-018-8786-y).
- [4] A. David, T. Subash, S. K. M. Varadhan, A. Melendez-Calderon, and S. Balasubramanian, "A framework for sensor-based assessment of upper-limb functioning in hemiparesis," *Frontiers Human Neurosci.*, vol. 15, Jul. 2021, Art. no. 667509. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnhum.2021.667509/full>
- [5] R. R. Bailey, J. W. Klaesner, and C. E. Lang, "An accelerometry-based methodology for assessment of real-world bilateral upper extremity activity," *PLoS ONE*, vol. 9, no. 7, Jul. 2014, Art. no. e103135.
- [6] P. S. Lum et al., "Improving accelerometry-based measurement of functional use of the upper extremity after stroke: Machine learning versus counts threshold method," *Neurorehabilitation Neural Repair*, vol. 34, no. 12, pp. 1078–1087, Dec. 2020.
- [7] D. S. de Lucena, O. Stoller, J. B. Rowe, V. Chan, and D. J. Reinkensmeyer, "Wearable sensing for rehabilitation after stroke: Bimanual jerk asymmetry encodes unique information about the variability of upper extremity recovery," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 1603–1608.
- [8] K. Leuenberger, R. Gonzenbach, S. Wachter, A. Luft, and R. Gassert, "A method to qualitatively assess arm use in stroke survivors in the home environment," *Med. Biol. Eng. Comput.*, vol. 55, no. 1, pp. 141–150, Jan. 2017.
- [9] N. Hogan and D. Sternad, "On rhythmic and discrete movements: Reflections, definitions and implications for motor control," *Exp. Brain Res.*, vol. 181, no. 1, pp. 13–30, Jul. 2007, doi: [10.1007/s00221-007-0899-y](https://doi.org/10.1007/s00221-007-0899-y).
- [10] T. Subash et al., "Comparing algorithms for assessing upper limb use with inertial measurement units," *Frontiers Physiol.*, vol. 13, Dec. 2022, Art. no. 1023589, doi: [10.3389/fphys.2022.1023589](https://doi.org/10.3389/fphys.2022.1023589).
- [11] G. Uswatte, W. H. R. Miltner, B. Foo, M. Varma, S. Moran, and E. Taub, "Objective measurement of functional upper-extremity movement using accelerometer recordings transformed with a threshold filter," *Stroke*, vol. 31, no. 3, pp. 662–667, Mar. 2000, doi: [10.1161/01.STR.31.3.662](https://doi.org/10.1161/01.STR.31.3.662).
- [12] A. David et al., "Quantification of the relative arm use in patients with hemiparesis using inertial measurement units," *J. Rehabil. Assistive Technol. Eng.*, vol. 8, Jan. 2021, Art. no. 205566832110196, doi: [10.1177/20556683211019694](https://doi.org/10.1177/20556683211019694).
- [13] E. Taub, K. McCulloch, G. Uswatte, D. M. Morris, M. Bowman, and J. Crago, "Motor activity log (MAL) manual," *UAB Training CI Therapy*, vol. 1, p. 18, Jan. 2011.
- [14] G. Uswatte and L. H. Qadri, "A behavioral observation system for quantifying arm activity in daily life after stroke," *Rehabil. Psychol.*, vol. 54, no. 4, p. 398, 2009.
- [15] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
- [16] T. Kluyver et al., "Jupyter notebooks—A publishing format for reproducible computational workflows," in *Proc. 20th Int. Conf. Electron.*, 2016, pp. 87–90.
- [17] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with Python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 92–96. [Online]. Available: <https://conference.scipy.org/proceedings/scipy2010/seabold.html>