



# OPEN Estimating gait parameters from sEMG signals using machine learning techniques under different power capacity of muscle

Shing-Hong Liu<sup>1,5</sup>, Alok Kumar Sharma<sup>1,5</sup>✉, Bo-Yan Wu<sup>1</sup>, Xin Zhu<sup>2</sup>, Chun-Ju Chang<sup>3</sup> & Jia-Jung Wang<sup>4</sup>✉

The gait analysis has been applied in many fields, such as the assessment of falling, force evaluation in sports, and gait disorder detection for neuromuscular diseases. Its main recording techniques include video cameras and wearable sensors. However, the present methods involve measuring surface electromyograms (sEMGs) to analyze muscle activities. The primary goal of this study is to estimate gait parameters under different power capacity of muscle by sEMGs measured from lower limbs. A self-made wireless device recorded sEMGs from two muscles of each foot, and GaitUp Physilog<sup>®</sup> 5 sensors captured gait parameters from 18 participants under running as references. Four features including median frequency (MDF), waveform length (WL), standard deviation (SD), and sample entropy (SampEn), were extracted from the sEMG data. The analysis utilized three machine learning models (Random Forest, CatBoost, XGBoost), evaluated through various evaluation metrics. Additionally, 5-fold cross-validation was conducted to assess the influence of muscle fatigue on the estimation of gait parameters. The results show that all models successfully estimated 20 gait parameters, all showing a Pearson correlation coefficient (PCC) above 0.800. However, the performance of models significantly depends on the condition of muscle fatigue. This study represents a significant advancement in gait analysis, providing a comprehensive method for estimating gait parameters from sEMG signals, with important implications for mobile health applications.

**Keywords** Surface electromyogram, Machine learning, Random forest, XGBoost, CatBoost, Gait parameters

Human gait represents the coordinated lower limb movements observed during walking or running and is a tangible reflection of the human body's dynamic motion characteristics. Gait analysis holds significant potential in diverse applications, encompassing the precise control of prosthetic limbs for individuals with lower limb amputations, the identification of individuals, the detection of fall risks, and the early diagnosis of neuromuscular disorders such as Parkinson's disease<sup>1</sup>. Researchers have employed a variety of methodologies to analyze gait patterns measured by non-wearable and wearable systems<sup>2</sup>. All these systems require the recording of surface electromyograms (sEMG) to assess muscle activities. Non-wearable systems usually use charge-couple device (CCD) cameras to record three-dimensional (3D) kinematic movements over time and space as subjects walk along a designated path<sup>3,4</sup>. The force exerted by a subject's feet on the floor is measured by floor sensors. Wearable systems use inertial sensors placed on several parts of bodies to capture signals characterizing human gaits<sup>5</sup>. Zahradka et al. evaluated video and sensor-based gait analysis systems for measuring terminal contact and initial contact in adults<sup>6</sup>.

sEMG is recorded by electrodes placed on the skin over muscles to capture the electrical activity of skeletal muscles when the muscles perform isotonic or isometric contractions<sup>7</sup>. The analysis of sEMG signals is primarily used to evaluate muscle activity, understand muscle geometry, measure muscle force<sup>8</sup>, and assess muscle fatigue<sup>9</sup>. Specifically, muscle activity measured by sEMG signals is particularly important in clinical gait

<sup>1</sup>Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung City 413310, Taiwan (ROC). <sup>2</sup>Department of AI Technology Development, M&D Data Science Center, Institute of Integrated Research, Institute of Science Tokyo, Tokyo, Japan. <sup>3</sup>Department of Golden-Ager Industry Management, Chaoyang University of Technology, Taichung City 41349, Taiwan (ROC). <sup>4</sup>Department of Biomedical Engineering, I-Shou University, Kaohsiung 82445, Taiwan (ROC). <sup>5</sup>These authors contributed equally: Shing-Hong Liu and Alok Kumar Sharma. ✉email: rbaloksharma@gmail.com; wangjj@isu.edu.tw

analysis<sup>10</sup>. The sEMG can identify muscle activities that do not play a crucial role in gait. These insights, together with clinical assessments, can provide valuable information for monitoring patients over time and designing targeted rehabilitation treatments post-surgery. Thus, unnecessary procedures can be avoided while reinforcing the unaffected structures, and ultimately achieving personalized optimal gait performance<sup>11</sup>. The sEMG for gait analysis usually focuses on force and joint angles. Previous studies utilized artificial neural networks to calculate the angles of knee joints and gait speed with sEMG measured from the muscles of lower limbs<sup>12–14</sup>.

Machine learning (ML) has demonstrated its efficacy in deciphering sEMG signals for various applications such as muscle fatigue<sup>15,16</sup>, motion estimation<sup>17</sup>, classifying gestures<sup>18,19</sup>, and predicting gait parameters<sup>20,21</sup>. Zhang et al.<sup>22</sup> employed multilayer perceptron (MLP) neural networks to continuously detect lower limb angles from sEMG signals. Guo et al.<sup>23</sup> used a combination of long-short term memory (LSTM) and a multilayer perceptron neural network to identify four different gait phases from sEMG signals. Morbidoni et al.<sup>24</sup> explored an ML approach specially formulated for the binary classification of gait events. This approach aims to predict the timing of heel-strike (HS) and toe-off (TO) from sEMG signals in the walking patterns of hemiplegic children. Cheng et al.<sup>25</sup> developed a method that integrates sEMG with accelerometer signals to identify dynamic activities by computing entropy values and acceleration. Morbidoni et al.<sup>26</sup> proposed a deep learning approach using sEMG data for classifying walking phases and predicting foot-floor contact, applicable in natural, non-treadmill environments. Tigrini et al.<sup>27</sup> introduced PHASOR, a phasor-based feature extraction approach that enhanced linear discriminant analysis (LDA) and support vector machine (SVM) performance in gait phase recognition. The study showed PHASOR outperformed state-of-the-art features, achieving 82% accuracy with a five-fold leave-one-trial-out testing approach using SVM. PHASOR, an efficient muscle synergy feature extraction method, also outperformed deep learning approaches with faster processing. Mengarelli et al.<sup>28</sup> investigated estimating the vertical component of ground reaction force (VGRF) using thigh and shank muscle EMG signals. Their study found that EMG signals could reliably estimate VGRF, with shank muscles alone offering a viable solution for reduced recording setups. These findings suggest the feasibility of incorporating GRF information into EMG-driven control schemes for prosthetic devices to improve walking dynamics and comfort. The previous studies primarily focused on predicting HS, TO, speed, stance, and swing; while other critical gait parameters were generally neglected<sup>29,30</sup>. This oversight has led to a significant gap in the research in this area.

This study aims to estimate 23 gait parameters using sEMGs measured from the lower limbs, focusing on two primary muscles, the gastrocnemius and vastus lateralis, in either foot. A customized wireless sEMG device was developed to capture data during treadmill running, while reference gait parameters were measured using shoe-worn GaitUp Physilog® 5 inertial sensors<sup>31</sup>. By using minimal sensors and analyzing sEMG features—median frequency (MDF), standard deviation (SD), simple entropy (SampEn), and waveform length (WL)—with machine learning models (Random Forest, CatBoost, and XGBoost), we demonstrate how a simplified wearable system can provide accurate gait parameter estimation. Additionally, this study evaluates how variations in muscle fatigue affect gait parameters, assessed through 5-fold cross-validation. This focus on power capacity of muscle adds depth to the analysis, highlighting its impact on gait parameter estimation. By minimizing sensor usage while maintaining accuracy, this work presents a practical approach for wearable devices in mobile health applications, offering insights into muscle engagement and real-time gait monitoring.

## Result

After preprocessing, the number of samples for each gait parameter of left and right feet was 1496. These samples were split into training sets and test sets in a ratio of 8:2. This research used an Intel i7-10700 CPU and an NVIDIA GeForce RTX3070 graphics card, and 64 Gbytes of 2933 MHz of RAM.

### Performance evaluation of regression models

Three ML models – RF, CB, and XGB – were used to predict gait parameters. There are 20 gait parameters that demonstrate a high level of consistency, each exhibiting a PCC value above 0.800 for both feet, to ensure the reliability and accuracy of our gait analysis in assessing locomotive function. Table 1 displays their PCC values for both feet. The RF model proved to be the most accurate, achieving PCC values of  $0.898 \pm 0.028$  for the left foot and  $0.895 \pm 0.031$  for the right foot. XGB, while slightly less precise, still delivered robust results with PCC values of  $0.896 \pm 0.032$  for the left foot and  $0.892 \pm 0.024$  for the right foot. CB, although the least accurate of the three, demonstrated substantial accuracy with PCC values of  $0.869 \pm 0.029$  for the left foot and  $0.869 \pm 0.033$  for the right foot.

The training time of three models, RF, CB, and XGB, was 0.49 s, 4.78 s, and 13.74 s, respectively. For prediction times, CB was the quickest at 0.01 s, followed by RF at 0.02 s, and XGB at 0.05 s. These results highlight the efficiency of trade-offs between these models, offering insights into selecting an appropriate algorithm based on time constraints in edge computing.

About the memory sizes of three models. RF model exhibited a moderate memory footprint of 12.76 MB, indicating its suitability for environments with average memory availability. CB, with its notably smaller size of 1.77 MB, is particularly advantageous for memory-constrained environments. Conversely, XGB, at 21.28 MB, has the highest memory requirement, potentially limiting its use in environments with restricted memory resources. The calculation of memory space in this study is crucial for the proposed method's application in edge computing, where memory efficiency is key. Edge computing environments require models that are both accurate and lightweight, as they often process large data volumes in real-time with limited resources. Understanding the memory requirements of these models allows for informed decisions on their suitability in such resource-constrained settings.

This study also calculated MPE, RMSPE, and MAPE for 20 gait parameters using CB model, shown in Table 2. The MPE outcomes revealed slight underestimations, measuring  $-0.94 \pm 2.03$  for the left and  $-0.68 \pm 1.30$  for the right foot, which could suggest a systematic bias in the measurement technique or model used. For RMSPE, the

Name	Left Foot			Right Foot		
	RF	CB	XGB	RF	CB	XGB
GCT	0.91	0.87	0.91	0.92	0.90	0.91
Cadence	0.91	0.87	0.91	0.92	0.89	0.91
Stance	0.89	0.87	0.90	0.86	0.82	0.86
Swing	0.89	0.87	0.90	0.86	0.82	0.87
LDr	0.89	0.86	0.88	0.89	0.86	0.88
FFr	0.92	0.90	0.92	0.90	0.87	0.91
PR	0.90	0.88	0.89	0.91	0.88	0.91
DS	0.90	0.88	0.91	0.88	0.85	0.87
SL	0.90	0.87	0.90	0.91	0.88	0.90
Peak Swing	0.90	0.90	0.91	0.93	0.92	0.92
SMTc	0.91	0.89	0.91	0.92	0.90	0.91
HSP	0.95	0.91	0.94	0.92	0.89	0.91
TOP	0.92	0.90	0.94	0.94	0.91	0.93
SW	0.91	0.87	0.91	0.91	0.89	0.90
Step Length	0.88	0.84	0.88	0.89	0.86	0.88
3D PL	0.91	0.88	0.91	0.91	0.90	0.90
Max HC	0.83	0.80	0.82	0.85	0.83	0.88
Max TC1	0.83	0.80	0.82	0.82	0.80	0.84
Min TC	0.89	0.86	0.86	0.89	0.85	0.89
Max TC2	0.91	0.86	0.90	0.86	0.85	0.86
Mean	0.898	0.869	0.896	0.895	0.869	0.892
SD	0.028	0.029	0.032	0.031	0.033	0.024

**Table 1.** PCC of gait parameters for left and right feet estimated by ML.

Name	MPE (%)		RMSPE (%)		MAPE (%)	
	L	R	L	R	L	R
GCT	0.15	0.09	2.96	2.62	0.02	0.02
Cadence	-0.39	-0.28	2.96	2.66	0.02	0.02
Stance	0.05	0.02	1.65	1.73	0.01	0.01
Swing	-0.20	-0.17	2.34	2.56	0.02	0.02
LDr	-1.41	-0.57	11.88	11.45	0.08	0.08
FFr	-0.80	-0.69	5.17	6.67	0.04	0.04
PR	0.11	-0.49	5.29	5.64	0.04	0.04
DS	-0.80	-1.05	10.88	10.36	0.06	0.06
SL	0.15	0.01	3.06	2.79	0.02	0.02
Peak Swing	0.24	-0.04	4.14	2.86	0.03	0.02
SMTc	0.05	-0.07	1.74	1.55	0.01	0.01
HSP	-0.26	-0.26	9.90	10.41	0.07	0.07
TOP	0.01	-0.19	3.88	3.55	0.03	0.03
SW	-7.47	-2.31	34.9	17.01	0.18	0.12
Step Length	0.04	-0.04	3.56	2.96	0.03	0.02
3D PL	0.01	-0.01	0.50	0.43	0.00	0.00
Max HC	-0.12	-0.03	3.44	3.19	0.03	0.02
Max TC1	-2.82	-1.20	13.39	12.09	0.10	0.09
Min TC	-5.38	-5.65	32.03	31.61	0.15	0.18
Max TC2	-0.03	-0.67	10.04	9.81	0.07	0.07
Mean	-0.94	-0.68	8.19	7.10	0.05	0.05
SD	2.03	1.30	9.43	7.33	0.05	0.04

**Table 2.** Comparative analysis of CB model for both left and right feet.

Group	MDF Left Calf	MDF Right Calf	MDF Left Thigh	MDF Right Thigh
	Mean $\pm$ SD (Hz)	Mean $\pm$ SD (Hz)	Mean $\pm$ SD (Hz)	Mean $\pm$ SD (Hz)
I	49.9 $\pm$ 4.8	60.4 $\pm$ 5.2	49.4 $\pm$ 6.8	49.3 $\pm$ 7.0
II	51.7 $\pm$ 4.6*	52.5 $\pm$ 5.2	48.6 $\pm$ 6.6	48.6 $\pm$ 6.5*
III	49.1 $\pm$ 5.2*	60.3 $\pm$ 5.2	47.7 $\pm$ 7.0*	48.0 $\pm$ 6.3**
IV	49.1 $\pm$ 4.7*	60.3 $\pm$ 5.1	47.3 $\pm$ 7.2**	47.4 $\pm$ 6.6**
V	49.0 $\pm$ 4.4**	60.4 $\pm$ 5.2	46.8 $\pm$ 7.0**	47.0 $\pm$ 7.0**

**Table 3.** The mean and SD of MDF for both left and right feet. \* represents p-value < 0.05, \*\* represents p-value < 0.001.

Fold	Left Foot			Right Foot		
	RF	CB	XGB	RF	CB	XGB
1	0.813 $\pm$ 0.041	0.776 $\pm$ 0.051	0.804 $\pm$ 0.056	0.826 $\pm$ 0.044	0.815 $\pm$ 0.043	0.818 $\pm$ 0.047
2	0.889 $\pm$ 0.043	0.873 $\pm$ 0.049	0.881 $\pm$ 0.049	0.900 $\pm$ 0.042	0.873 $\pm$ 0.049	0.883 $\pm$ 0.048
3	0.904 $\pm$ 0.026	0.894 $\pm$ 0.030	0.898 $\pm$ 0.030	0.891 $\pm$ 0.046	0.882 $\pm$ 0.052	0.883 $\pm$ 0.046
4	0.913 $\pm$ 0.024	0.890 $\pm$ 0.025	0.899 $\pm$ 0.033	0.928 $\pm$ 0.021	0.896 $\pm$ 0.034	0.911 $\pm$ 0.038
5	0.859 $\pm$ 0.042	0.832 $\pm$ 0.045	0.851 $\pm$ 0.044	0.858 $\pm$ 0.060	0.850 $\pm$ 0.057	0.848 $\pm$ 0.051

**Table 4.** Five-fold Pccs of gait Estimation (Mean & SD) for both left and right feet.

values of  $8.19 \pm 9.43$  for the left foot and  $7.10 \pm 7.33$  for the right foot were noted, indicating substantial variability in the gait parameter estimations, possibly reflecting the inherent complexities in gait dynamics or measurement inaccuracies. Finally, the MAPE values, standing at  $0.05 \pm 0.05$  for the left foot and  $0.05 \pm 0.04$  for the right foot, demonstrated a generally high level of accuracy in the absolute terms of the model's estimations.

### Conditions of muscle fatigue

In this study, we conducted an analysis of the MDF of muscle fatigue in both the left and right calves and thighs across various experimental groups. The MDF, a crucial parameter in sEMG, serves as an indicator of muscle fatigue<sup>32</sup>. When the subjects finished a six-minute run, we hypothesized that they would exhibit signs of muscle fatigue.

Table 3 displays both the mean and SD of the MDF values for each muscle across all groups. The mean values provide an average metric of muscle fatigue, essential for understanding the typical performance of the muscles under this study. The SD, on the other hand, shows the variability or consistency of these measurements across each group. We find that, except for the calf muscle of the right foot, the MDFs of all muscles approach a lower frequency from group I to V. Statistical analysis reveals significant differences between several groups for both the calf and thigh muscles of the foot. For the calf muscle of the left foot, the MDF in group I compares with them in groups II, III, and IV, suggesting significant differences with a p-value of less than 0.05. The MDFs in groups I and V indicate a highly significant difference, likely corresponding to a p-value of less than 0.001.

The MDFs in groups I and III show a significant difference in the thigh muscle of the left foot, with a p-value of less than 0.05. Comparing group, I to groups IV and V yields significant results with p-values below 0.001. For the thigh muscle of the right foot, the MDFs in groups I and II differ significantly with a p-value of less than 0.05. With p-values below 0.001, group I comparisons to groups III, IV, and V is highly significant.

### 5-Fold cross validations

Table 4 presents the results of gait parameter estimation using five-fold cross-validation, with Pearson Correlation Coefficients (PCCs) for both left and right feet. Focusing on the fold-specific results, Fold 1 shows relatively lower PCC values compared to the other folds, with the best result for the left foot being 0.813 (RF) and for the right foot being 0.826 (RF). Fold 2 shows a noticeable improvement in performance, with the RF model achieving 0.889 for the left foot and 0.900 for the right foot, marking this fold as one of the better ones for gait estimation. Fold 3 maintains good performance, with RF achieving 0.904 for the left foot and 0.891 for the right foot. Although slightly lower than Fold 2 for the right foot, the results are still strong overall. Fold 4 stands out with the best results across all folds, where RF achieves the highest PCC values of 0.913 for the left foot and 0.928 for the right foot, indicating that this fold provides the most accurate gait parameter estimation. Finally, Fold 5 shows a slight decline in performance compared to Fold 4, with RF achieving 0.859 for the left foot and 0.858 for the right foot. Although these values are still solid, they are lower than those observed in Fold 4. In conclusion, Fold 4 provides the best overall results for both left and right feet, offering the most reliable gait parameter estimation, with Fold 2 also showing strong performance, particularly for the right foot.

### Discussion

The development of wearable devices has become a critical issue in mobile health (mHealth) because these devices handle various health services, including monitoring chronic disease conditions, reducing healthcare

costs, empowering patients and families to manage their daily health, and providing direct access to health services regardless of time and place<sup>33</sup>. Consequently, many studies over the past decade have focused on developing innovative wearable devices for patient care and health management. For example, the Apple Watch not only measures calories and steps but also detects arrhythmias via electrocardiogram and monitors blood oxygen saturation<sup>34</sup>. Liu et al. proposed an sEMG patch that measures the MDF in real-time to evaluate muscle fatigue conditions<sup>35</sup>. In this study, four features extracted from sEMG—MDF, WL, SD, and SampEn—were used to predict twenty gait parameters, achieving PCC values above 0.80, using ML models.

### Comparison of spatiotemporal gait parameter Estimation accuracy based on ML models

The PCC values, as presented in Table 1, indicate strong correlations between the estimated and actual gait parameters for all three models, with most values exceeding 0.85, signifying a high level of agreement. Among the models, XGB consistently showed the highest PCC values, suggesting superior performance in capturing the relationships between the input features and gait outcomes. RF performed similarly to XGB, with only minor differences observed in performance. While CB also provided strong correlations, it demonstrated slightly lower PCC values compared to RF and XGB. However, these differences were not substantial enough to undermine the model's overall effectiveness.

When comparing the models using MPE (Table 2), XGB exhibited some variability, particularly for parameters such as “Step Length” and “Swing,” where it slightly underestimated the true values, as indicated by negative MPE values. Despite this, XGB maintained reliable performance across most parameters. The CB model demonstrated lower MPE values for the right foot, particularly for parameters like “FFr” and “PR,” suggesting it slightly underpredicted these values relative to the true measurements. The RMSPE values further corroborated XGB's superior performance, showing lower errors, especially for parameters like “GCT” and “Cadence.” This suggests that XGB not only captured the patterns more effectively but also provided more precise estimates of gait parameters. Although CB and RF exhibited slightly higher RMSPE values, they still provided reasonable predictions, confirming their utility for gait parameter estimation. Finally, the MAPE values, which quantify the relative percentage error of predictions, were low across all models, reflecting their overall accuracy. CB showed particularly strong results for the left foot, with MAPE values close to zero for parameters such as “3D PL” and “Cadence,” indicating minimal deviation from the true values.

### Analysis of ground clearance parameter Estimation stability of each model

We evaluate the stability of ground clearance parameter estimation (Max HC, Max TC1, Min TC, and Max TC2) across three ML models. The PCC values in Table 1 show strong correlations between the estimated and actual values for all parameters, with most values exceeding 0.80. XGB consistently outperformed RF and CB, particularly for the right foot, where it achieved the highest PCC values, especially for ‘Max HC’ (0.88) and ‘Max TC1’ (0.84). RF and CB also demonstrated strong performance, with PCC values ranging from 0.80 to 0.83, indicating that these models captured the ground clearance parameters well, although XGB showed slightly better consistency.

Regarding MPE (Table 2), all models exhibited slight underestimations of the ground clearance parameters, with negative MPE values, particularly for ‘Min TC,’ where values ranged from −5.38% to −5.65%. XGB displayed greater stability in MPE across both feet, with slightly less underestimation compared to RF and CB. Notably, CB showed a larger underestimation for the ‘Max TC2’ parameter on the left foot (−0.67%). The RMSPE values indicated that XGB generally provided the most precise estimates, with lower errors for parameters such as ‘Max HC’ (3.19% for the right foot). While RF and CB had slightly higher RMSPE values for parameters like ‘Max TC1’ and ‘Min TC,’ their error rates were still within an acceptable range, demonstrating reliable performance in estimating ground clearance. Finally, the MAPE values confirmed low relative errors across all models, further validating their accuracy in estimating ground clearance parameters. XGB again performed the best, with near-zero MAPE values for most parameters, reflecting minimal deviation from actual measurements.

### Impact of muscle fatigue on gait parameter Estimation

Since sEMG signals were used to estimate gait parameters, we analysed the power capacity of muscle to assess its impact on model performance. The experiment segmented the data into five groups to capture muscle contraction dynamics over time. One key parameter, Median Frequency (MDF), used to detect muscular fatigue<sup>32</sup>, showed lower frequencies across most muscles, except for the right calf (Table 3). Significant differences ( $p < 0.001$ ) were observed between groups I and V, indicating changes in muscle function due to fatigue or adaptation.

This study employed 5-fold cross-validation with time-sequenced data splitting to ensure robust evaluation of muscle fatigue during running. The lowest PCC values in groups I and V (Table 4) highlight the complex relationship between muscle fatigue and gait, suggesting that muscle fatigue or adaptation leads to transitions that the model struggles to capture accurately. These findings emphasize the need for more advanced models that account for the dynamic shifts in muscle function and their effect on gait.

### Optimizing sensor and model size for edge computing

A key advantage of our approach is the use of a minimal number of sEMG sensors, which is crucial for improving the feasibility of integrating sEMG-based gait analysis into wearable systems. Previous studies often relied on larger sensor arrays or additional signal sources, increasing the complexity of experimental setups. By minimizing the number of sensors while still achieving high prediction accuracy, our study demonstrates the potential for simplified and practical wearable applications for real-time gait assessment. This reduction in sensor usage is particularly important for practical applications in both clinical and daily-life settings. A similar effort to reduce sensor dependency was presented by Mobarak et al.<sup>36</sup>, who showed that proximal sensor placements can effectively estimate gait kinematics. Our study further supports the idea that sEMG-based gait



analysis can be achieved with fewer sensors while maintaining robust accuracy, making the technology more accessible for real-world deployment.

The findings also underscore the feasibility of developing compact, efficient wearable devices that provide deep insights into muscle fatigue and gait dynamics. Models such as XGB and CB demonstrated close PCC values, but considerations of model size, training, and prediction times are critical for practical use. The XGB model, with a large memory requirement of 21.28 MB, might limit its use in edge computing environments with limited resources. The RF model showed well-rounded performance but required 12.76 MB of memory, which could pose challenges for deployment in edge computing due to its size. In contrast, the CB model, with its minimal 1.77 MB size, emerges as a more feasible choice for applications where memory and processing constraints are a concern. This highlights the balance between performance and computational efficiency needed for wearable devices.

### Comparison of results with previous studies

Furthermore, compared to Liu et al.<sup>37</sup>, who also utilized sEMGs to predict gait parameters but achieved PCC values above 0.80 for only fourteen gait parameters, our method demonstrated improvements in prediction accuracy across a wider range of parameters. The inclusion of WL as a critical feature in our analysis contributed to this improvement. Additionally, because all features in our study were extracted from sEMG signals, the overall signal processing complexity was lower than in previous studies.

Previous study<sup>29</sup> evaluated gait metrics by analyzing EMG signals with the aid of an extreme learning machine (ELM). The validity of the procedure was confirmed by evaluating sEMG data from the Tibialis Anterior and Gastrocnemius Lateral muscles during walking in a group of 18 healthy participants. Inertial sensors and video cameras were used to gather measurements of gait speed and the phases of stance and swing. The results showed RMSPE for the stance percentage of 7.62%, gait speed of 11.86% and swing percentage of 6.07%, respectively. However, in Table 2, our study reports that the RMSPE of left and right feet for the stance parameter are 1.65% and 1.73%, and for swing parameter are 2.34% and 2.56%, which are better than those reported in previous studies.

Zhang et al.<sup>38</sup> utilized wearable sensors to estimate three distinct gait parameters—stride velocity (SV), stride length (SL), and foot clearance (FC)—using support vector regression to analyze both running and walking. The research involved 14 participants across two separate sessions dedicated to these activities. The findings indicated error rates for running at 2.91% for SV, 2.59% for SL, and 5.13% for FC. Hao et al.<sup>39</sup> estimated the SL and stride width (SW) in their gait analysis using an IMU-based sensor. They used nine participants, recording error percentages of -0.24% for SL and -0.02% for SW. In contrast, the current study has demonstrated that the application of the CB model significantly reduces the error percentage of statistical metrics to 0.08% for SL. Previous studies focused on specific aspects of gait analysis for estimation, limiting their scope to certain parameters rather than providing a comprehensive analysis. In contrast, our study successfully estimated 20 gait parameters, offering a more complete analysis.

This study's methodology has the potential benefit for utilizing in clinical environments to evaluate the gait parameters of patients with the gait-related illnesses, including Parkinson's disease, cerebral palsy, and multiple sclerosis. Furthermore, incorporating a compact CB model designed for the edge computing will allow real-time data processing on wearable devices, enabling instant analysis and feedback for the users.

### Conclusions

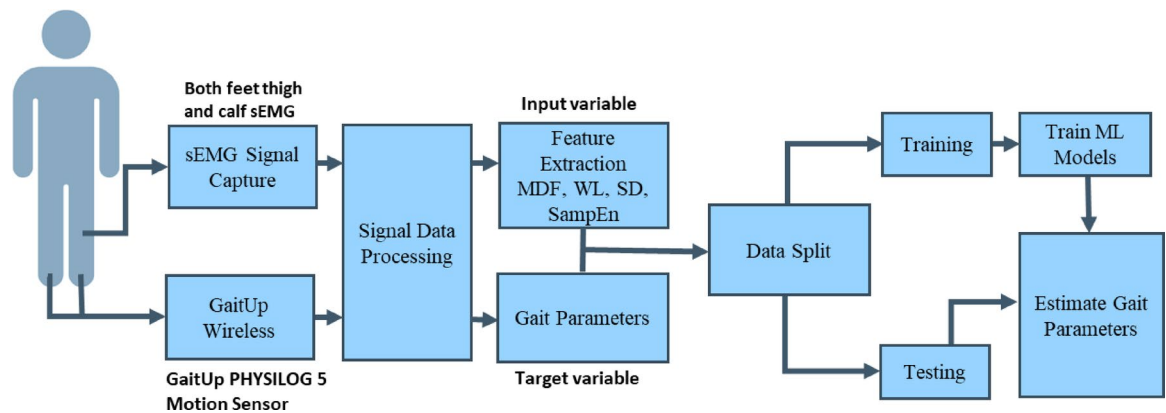
In this research, we propose a sEMG device to collect data from the muscles of both feet, utilizing ML to predict gait parameters. Since gait parameters were estimated using sEMG, the proposed method can be applied not only to gait analysis but also to assess muscle fatigue, motion classification, evaluating muscle fatigue, and measuring muscle force. Moreover, our work contributes to exploring the relationship between muscle fatigue conditions and the performance of gait parameter estimation. This study also has some limitations, such as the inability to accurately predict HS, speed and TA. Additionally, there is a concern about the size of the ML models. In future studies, we will attempt to improve the results and reduce the size of the models. Furthermore, combining the sEMG patch with edge computing techniques will be applied in mHealth.

### Methods

Figure 1 illustrates the workflow of gait-parameter estimation using the sEMG signals from the thigh and calf muscles, captured with GaitUp Physilog® 5 sensors. The method involves the sEMG signals of both feet measured by the self-made boards, gait parameters measured by GaitUp sensors as the target output, extraction of sEMG parameters, MDF, WL, SD, and SampEn as the input features, and data partitioning for model training and testing. This system also demonstrates how sensor technology and ML can be integrated effectively to estimate the gait parameters.

### Experimental protocol

The experiment was conducted with voluntary participants, primarily involving healthy adult females. The study included 18 participants, aged 19 to 23 years, with an average age of  $20 \pm 1$  years. The participants had an average height of  $156 \pm 4.6$  cm, an average weight of  $45.9 \pm 5.7$  kg, and an average shoe size of  $23.9 \pm 0.6$  cm. Prior to participation, each participant completed a self-health assessment to ensure eligibility for the experiment, and informed consent was obtained from all participants before the experiment. This study was approved by the Institutional Review Board of Chung Shan Medical University Hospital, Taichung City, Taiwan, under the reference number CS2-22210.



**Fig. 1.** Workflow diagram for the estimation of gait parameters by sEMG signals.

All methods were performed in accordance with the relevant guidelines and regulations, including the Declaration of Helsinki and ICH-GCP requirements. The first step of the experimental protocol was to place electrodes on the vastus lateralis and gastrocnemius muscles on both feet of participants, as shown in Figs. 2 (a) and (b), for measuring the sEMG signals of calf and thigh muscles. The circuit board was placed in a waist pack, which was positioned in a way that would not interfere with natural movement. We avoided the belly position and shifted the electrodes to a higher position. The surface electrodes used for the EMG recording were Ag/AgCl with a 10-mm diameter on self-adhesive supports. The electrode arrangement ensured negligible crosstalk between adjacent muscles. The positions of the electrodes for each subject were recorded, and the electrodes were placed at the same position. The self-made boards were attached to the participants' thighs. Cardboard was stuck on the back of the board to protect it and decrease uncomfortableness. The second step was to place the GaitUp Physiolog<sup>5</sup> wearable inertial sensors on the tongue of each shoe, ensuring stability and uninterrupted data integrity, as shown in Fig. 2(c).

The third step was to request participants to walk or run on a treadmill at a consistent speed of 5 km per hour for 6 min, as shown in Fig. 2(d). This speed was based on established practices in endurance and gait analysis research<sup>40,41</sup>. This specific speed and duration are frequently employed to strike a balance between acquiring ample data for statistical robustness and mitigating the risk of participant fatigue. Additionally, these parameters closely reflect the typical walking speeds and endurance capabilities observed in the healthy adult population. During this period, subjects could request to decrease the speed when they felt tired. There were four measurements whose recording interval was at least one week to avoid muscle fatigue. The detailed procedure was described in the study of Liu et al.<sup>37</sup>.

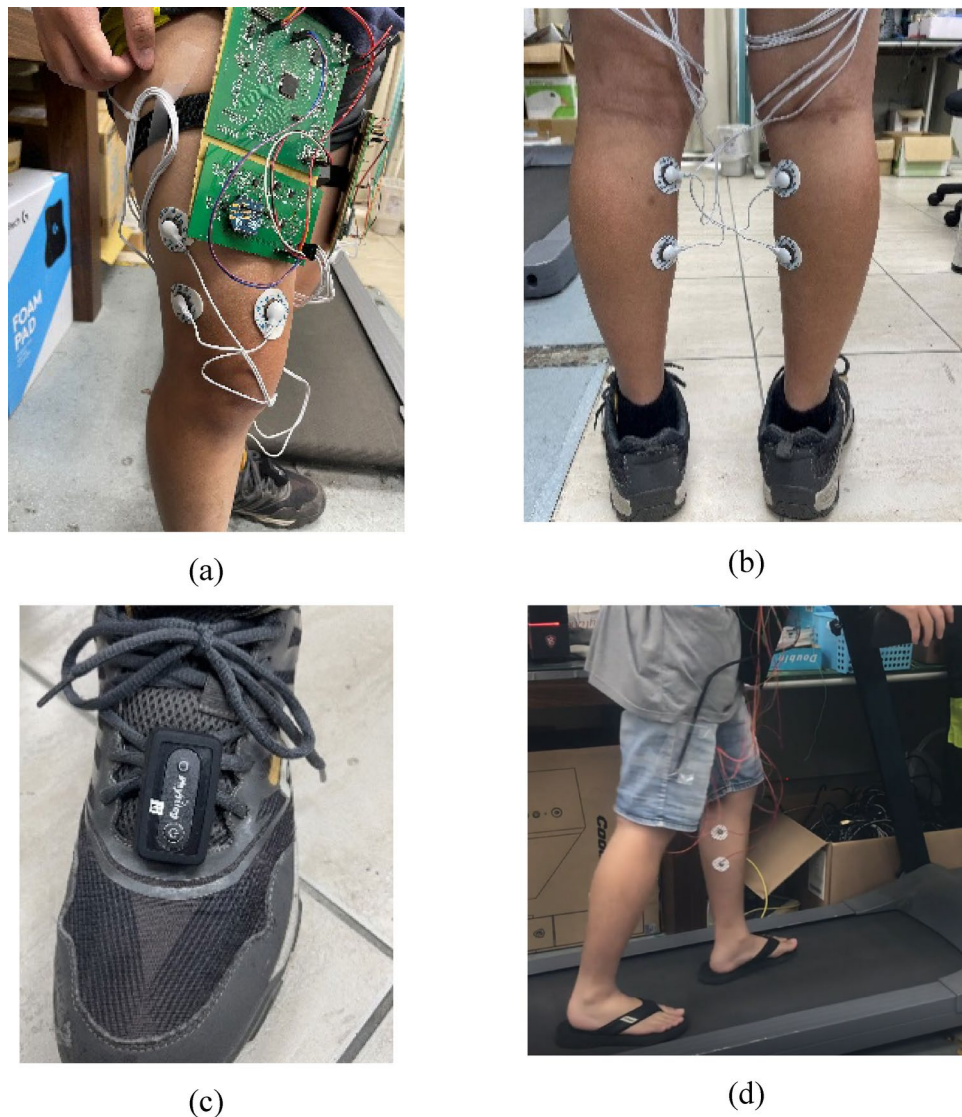
### Experimental device

The self-made wireless device measured the sEMG signals from both feet at a sampling rate of 1000 Hz. This device consists of two boards: a slave board and a master board, each placed on one foot, as shown in Fig. 2(a). The XBee S2C modules (Digi International XBee<sup>®</sup>, Hopkins, USA) were used for data exchange between the two boards, while the HC-05 Bluetooth module was employed to transmit the acquired sEMG data from the master board to a personal computer (PC). This setup enabled the synchronous recording of sEMG signals from both feet. The sEMG circuit was designed following the study by Liu et al.<sup>35</sup>. The passband was 33.9–482.5 Hz, using second-order Butterworth filters, and the total gain was 550 V/V. The motion signals from each foot were measured using both the sEMG and GaitUp sensor signals. To prevent data loss from the slave board to the PC terminal, a down-sampling technique was implemented. The slave board sent data to the master board at a rate of 1000 Hz, and the master board relayed the data to the PC terminal at 500 Hz. This approach successfully prevented any data loss from the slave board.

### Gait parameters

Figure 2(c) depicts the gait analysis device, Physiolog<sup>®</sup> 5 (GaitUp SA, Sweden), utilized in the study. The Physiolog<sup>®</sup> 5 is an advanced auxiliary tool designed to capture objective, quantitative data regarding gait conditions. It enables the collection of pre- and post-analysis data, which supports clinicians in identifying gait deficiencies and assessing potential risks. For this research, the GaitUp Lab gait analysis system (Fig. 3) was used to extract comprehensive gait parameters for both the left and right feet of the subjects.

The gait parameters encompass a total of 23 distinct measurements. These include 9 temporal parameters, 9 spatial parameters, 4 pertaining to ground clearance analysis, and 1 dedicated to turn analysis. Time parameters include 'heel strike time' (HS), 'gait cycle time' (GCT), 'double leg support' (DS), 'Cadence', 'stance phase' (stance), 'swing period' (swing), 'load ratio' (LD<sub>r</sub>), 'foot flat ratio' (FF<sub>r</sub>), and 'push ratio' (PR). Spatial parameters cover 'step length' (SL<sub>length</sub>), 'stride length' (SL), 'gait speed' (Speed), 'maximum angular velocity during swing' (Peak Swing), 'foot speed at minimal toe clearance' (SMTC), 'heel strike angle' (HSP), 'toe off angle' (TOA), 'swing width' (SW), and '3D path length' (3D PL). Turn analysis includes 'turning angle' (TA), and ground clearance analysis features 'maximum heel clearance' (Max-HC), 'maximum toe clearance 1' (Max-TC1), 'minimum toe



**Fig. 2.** (a) Electrode positioned on vastus lateralis muscle. (b) Electrode positioned on gastrocnemius muscle. (c) Gait analyzer attached to shoe tongue. (d) Walking on a treadmill.

clearance' (Min-TC), and 'maximum toe clearance 2' (Max-TC2). The sEMG signal consisted of 180,000 data points recorded over 6 min of running.

### Signal processing

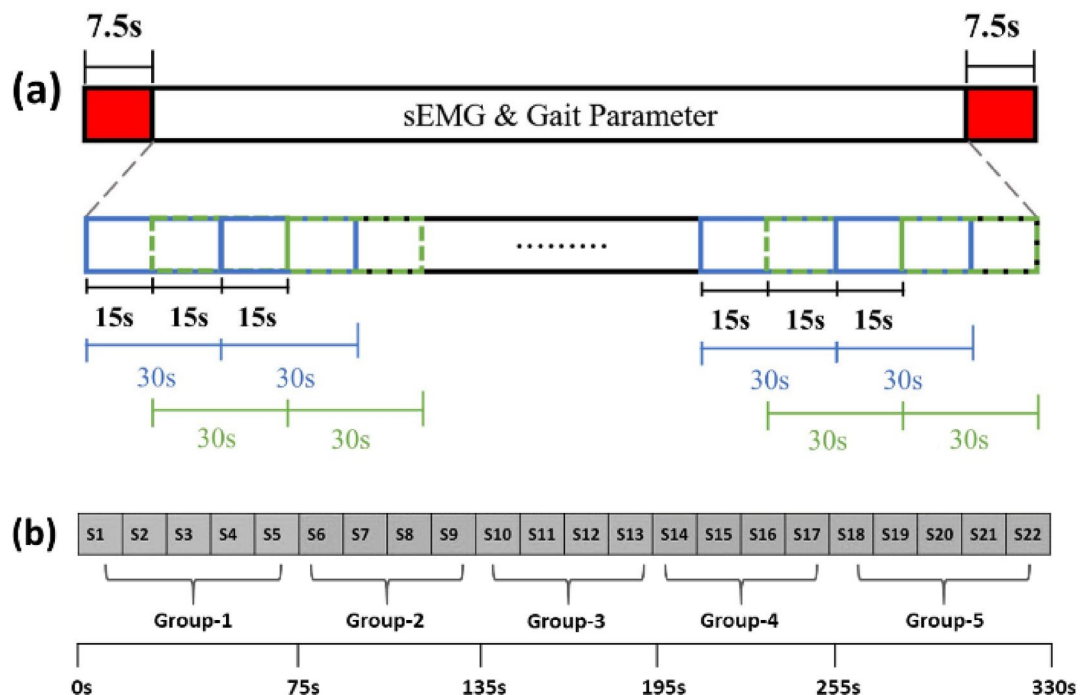
Figure 4 illustrates the schematic diagram of the signal processing procedure. After recording the complete signal, the sEMG measurement device captured a signal length of 180,000 data points over 6 min during the experiment. Since the sEMG multi-channel measurement device used in this study is compatible with the GaitUp Physilog 5 system, which operates differently, the data was cut at specific time points. The starting time for data cutting was determined by the point when the initial force of the sEMG was recorded, aligned with the moment when the activity first became visible in GaitUp Lab.

To eliminate any potential effects of gait variability at the beginning and end of the experiment, 7.5 s were removed from both the start and end of the signal, resulting in a total removal of 15 s. After this adjustment, the signal length was reduced to 172,500 data points. For the signal segmentation, both the sEMG measurement device and GaitUp Physilog<sup>®</sup> 5 were set to segment the data into 30-second window (with blue block), and a 15-second shift (with green block)<sup>42</sup>, shown in Fig. 4(a). This approach generated 22 samples per experiment. Gait parameters from the recorded GaitUp Physilog<sup>®</sup> 5 data were then extracted using the described method in GaitUp Lab. The first and last 7.5 s of sEMG and GaitUp sensor signals were removed. This choice was guided by the need to capture meaningful variations in gait and the power capacity of muscle over time, while maintaining a sufficient number of segments for statistical analysis. The selected window size ensures that transient variations do not dominate the extracted features, while still providing an appropriate temporal resolution for tracking changes in gait parameters across different time intervals.





**Fig. 3.** GaitUp Lab gait analysis system.



**Fig. 4.** (a) Signal data segmentation. (b) Five groups of the segment.

After segmentation, the data were categorized into five groups based on the time sequence. Specifically, Group 1 included segments 1–5, Group 2 included segments 6–9, Group 3 covered segments 10–13, Group 4 contained segments 14–17, and Group 5 comprised segments 18–22, as shown in Fig. 4(b). The gait parameters for both feet in each segment were calculated using GaitUp Lab, leading to a total of 1,496 segments per foot for analysis.

To prevent any data leakage during model evaluation, each segment's data was strictly separated manually during the cross-validation process, ensuring that no data from the testing set overlapped with the training data. This careful partitioning minimizes the risk of data leakage, which could otherwise lead to overly optimistic performance metrics and negatively affect the model's ability to generalize.

### Feature extraction of sEMG signal

Feature extraction from sEMG signal is crucial for maintaining key neural information by removing non-essential elements and interference. This simplifies the original signal into a concise set of features reflecting muscle contraction intensity and neuromuscular activation for depicting muscle contraction physiology<sup>43</sup>. The sEMG features of the proposed method covered both time-domain and frequency-domain components carried

---

**Input:** dataset  $D$ , containing of  $S$  samples and target values

- 1 **Hyperparameters:** number of iterations  $n_{iter}$ , learning rate  $\eta$ , maximum tree depth  $d$ , and loss function  $L$ .
- 2 **Initialization**  
 Initialize the ensemble model as an empty list.  
 Initialize the base model as the mean of the target variable.
- 3 **for**  $t = 1$  to  $n_{iter}$  **do**
- 4     Compute gradients for each sample in  $D$  using the current ensemble model.
- 5      $DT_t \leftarrow \text{TreeBuild}(D, \text{negative gradients}, d)$
- 6     Scale predictions of  $DT_t$  by  $\eta$  and add it to the ensemble model.
- 7     Update the ensemble model by incorporating the predictions of  $DT_t$ .
- 8 **end for**
- 9 **Return** the final ensemble model.

---

**Algorithm 1.** Pseudo Code for CatBoost.

out separately<sup>44</sup>. Previous studies showed that the median frequency (MDF), waveform length (WL), standard deviation (SD), and sample entropy (SampEn) were the common features<sup>45–47</sup>.

MDF represents the point at which half of the signal's power spectrum is distributed, as in (1).

$$\sum_{j=0}^{MDF} p_j = \sum_{j=MDF}^M p_j = \frac{1}{2} \sum_{j=0}^M p_j \quad (1)$$

where  $p_j$  is the  $j^{\text{th}}$  spectrum density, and  $M$  is the Nyquist frequency.

WL quantifies the sEMG signal complexity, representing the total waveform length over a time segment. In (2), 's' denotes sEMG signal amplitude, 'N' the number of data points, and 'i' is an index from 1 to  $N - 1$ , used for calculating amplitude differences between successive points.

$$WL = \sum_{i=1}^{N-1} |s_{i+1} - s_i| \quad (2)$$

SD is calculated as Eq. (3).

$$SD = \sqrt{\frac{1}{2} \sum_{i=1}^N (s_i - \bar{s})^2} \quad (3)$$

Sample Entropy (SampEn) serves as a measure for evaluating consistency and complexity in a time series of signals. A greater entropy value signifies more complexity in the time series. To compute SampEn, define the dimension  $m$  and select a suitable value for  $r$ .

The signal segments represented  $S_m$ , according to the following process:

$$S_m(i) = \{s_i, s_{i+1}, s_{i+2}, \dots, s_{i+m-1}\} \quad (4)$$

Equation (5) represents the expression of SampEn

$$\text{SampEn} = -\ln \frac{A}{B} \quad (5)$$

$$A = d[S_{m+1}(i), S_{m+1}(j)] < r \quad (6)$$

$$B = d[S_m(i), S_m(j)] < r \quad (7)$$

where  $A$  represents the count of pairs of sequences of length  $n + 1$  that are within a tolerance  $r$  and  $B$  represents the count of pairs of sequences of length  $n$  that are within the same tolerance  $r$ . These counts are used to calculate SampEn, which measures the complexity of the dataset.  $r = SD \times 0.2$ ,  $m = 2$ , and  $d$  is the Euclidean distance.

In this research, sEMG signals were collected from both the thigh and calf muscles, and a total of 8 features were extracted from the sEMG signals for each foot.

---

```

1  Input: Dataset of size  $S$ , number of trees  $n_{tree}$ , number of features  $N$ ,
    leaf minimum  $n_{leaf}$ 
2  procedure RandomForest( $S, n_{tree}, N, n_{leaf}$ )
3      for  $j=1$  to  $n_{tree}$  do
4          Extract a bootstrap subset with size  $L \leq S$ .
5          Build decision tree  $DT_j$  from the bootstrap subset as follows:
6               $DT_j \leftarrow \text{TreeBuild}$  (node = 1, samples = bootstrap,  $N$ )
7      end for
8      return ensemble tree  $\{DT_j\}_{j=1}^{n_{tree}}$ 
9  end procedure
10 procedure TreeBuild (node, subset,  $N$ )
11     if length(subset) <  $n_{leaf}$  then
12         Set the node as a leaf.
13     else
14         Randomly select  $P = \frac{N}{3}$  features from the  $N$  total features.
15         Determine the optimal attribute and the division threshold.
16         Divide node into two child nodes (left and right).
17         leftSubset  $\leftarrow$  subset meeting division criteria.
18         rightSubset  $\leftarrow$  subset failing to meet division criteria.
19         TreeBuild (left node, leftSubset,  $N$ )
20         TreeBuild (right node, rightSubset,  $N$ )
21     end if
22 end procedure
23 To predict a value at a new data point  $x$ :
24      $\{DT_j\}_{j=1}^{n_{tree}} \leftarrow \text{RandomForest}(S, n_{tree}, N, n_{leaf})$ 
25      $y(x) = \frac{1}{n_{tree}} \sum_{j=1}^{n_{tree}} DT_j(x)$ 

```

---

**Algorithm 2.** Pseudo Code for Random Forest.

---

```

1  Input: feature matrix  $x$ , target variable  $y$ 
2  Set the number of base learners  $n_{estimators}$ 
3  Initialize ensemble prediction  $\hat{y}$  with 0 value:  $\hat{y} = 0$ 
4  for  $i=1$  to  $n_{estimators}$  do
5      Calculate residuals  $R = y - \hat{y}$ .
6      Learn the base learner on residuals and  $x$ .
7      Get the prediction of the current base learner:
           $\hat{y}_i = \text{basic}_{\text{learner\_predictor}}(x)$ 
8      Update the ensemble model prediction  $\hat{y} = \hat{y} + l_{rate} \times \hat{y}_i$ 
9  end for
10 Return ensemble prediction  $\hat{y}$ 

```

---

**Algorithm 3.** Pseudo Code for XGBoost.

## Machine learning models

Three distinct types of ensemble ML regression models – CB, RF, and XGB – were used to estimate various gait parameters. These models were chosen for their proven efficacy in handling complex datasets, and their application aimed at providing more nuanced insights into the dynamics of human gait.

### Catboost

CatBoost (CB)<sup>48,49</sup> is particularly adept at processing categorical data seamlessly with minimal preprocessing, owing to its unique approach to transforming categorical values into numerical inputs. This algorithm enhances robustness against overfitting through its use of ordered boosting, a methodology that introduces randomness in the way training data is fed to the model, thus improving generalization. Additionally, CB algorithm begins by accepting a dataset  $D$  comprising  $S$  samples and their corresponding target values, alongside hyperparameters such as the number of iterations,  $\eta$ , learning rate,  $\gamma$ , maximum tree depth,  $d$ , and choice of loss function,  $L$ . Following initialization, where the ensemble model is initialized as an empty list and the base model as the mean of the target variable, the algorithm iterates through each iteration from  $t = 1$  to  $T$ . Within each iteration, gradients are computed for each sample in  $D$  based on the current ensemble model's predictions, guiding the construction of a decision tree  $DT_t$  aimed at predicting these gradients.

After scaling the predictions of  $DT_t$  by the learning rate and incorporating them into the ensemble model, the algorithm continues to update the ensemble by integrating the predictions of the newly added tree. Upon completion of all iterations, the final output is the ensemble model consisting of decision trees,  $DT_t$ , which collectively represent the combined knowledge extracted from the dataset. Algorithm 1 presents the pseudo-code implementation of the CB algorithm discussed.

### Random forest

Random Forest (RF)<sup>50,51</sup> is prominent for its straightforward application and robust performance across diverse datasets and problem types. This algorithm aggregates predictions from a multitude of decision trees to mitigate the overfitting issues typical of individual trees. The robustness of RF extends to its handling of both classification and regression tasks, making it a versatile tool in predictive analytics. Nevertheless, the model's complexity can lead to substantial memory consumption and slow prediction times when the ensemble includes a large number of trees.

**Algorithm 2** outlines the process of building RF tasks. Initially, it accepts a dataset with samples, along with parameters such as the number of trees ( $n_{tree}$ ), the number of features ( $m$ ), and the minimum samples in a leaf node ( $n_{leaf}$ ). It then proceeds to build each DT in the RF by iteratively selecting bootstrap subsets from the dataset. Within each tree-building process, a subset of features is randomly chosen, and the tree is recursively partitioned based on optimal attribute selection until leaf nodes contain a minimum number of samples. To predict a new data point, the algorithm averages the predictions of all decision trees in the ensemble.

### XGBoost

XGBoost (XGB)<sup>52</sup> stands out for its scalability and execution speed, driving its popularity in the ml community, particularly for competitive data science. It incorporates advanced regularization techniques, which help in reducing overfitting, thereby enhancing the predictive accuracy of the model<sup>53</sup>.

XGB employs a series of decision trees, termed base learners, to sequentially refine its predictions in a ml context. Initiated with an ensemble prediction ( $\hat{y}$ ) set to zero, the algorithm proceeds to calculate the residuals ( $R$ ), which represent the discrepancies between the observed values ( $y$ ) and the current predictions ( $\hat{y}_i$ ). Each tree is trained on these residuals, targeting the specific errors of the preceding model. The predictions from each base learner are then scaled using a predefined learning rate ( $\eta$ ) and subsequently integrated into the overall ensemble prediction ( $\hat{y}$ ). This iterative process, repeated for a defined number of iterations ( $n_{estimators}$ ), cumulatively adjusts the model, aiming to minimize the residuals and enhance accuracy. The final ensemble prediction ( $\hat{y}$ ) represents the cumulative adjustments made by all the base learners. Algorithm 3 outlines the process of building XGB tasks.

### Evaluation metrics

The study conducted a comprehensive evaluation of ML model, utilizing metrics such as Pearson correlation coefficient (PCC), Mean Percentage Error (MPE), Root Mean Squared Percentage Error (RMSPE), and Mean Absolute Percentage Error (MAPE).

- a) MPE calculates the average of percentage discrepancies between the estimated values and the actual values. Equation (8) represents the MPE equation.

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{x_i - y_i}{x_i} \times 100\% \quad (8)$$

- b) RMSPE measures estimation accuracy by calculating the square root of the average squared percentage errors between estimated and actual values.

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - y_i}{x_i} \right)^2} \times 100\% \quad (9)$$

- c) MAPE evaluates the mean of the absolute percentage errors through a comparison between the predicted and actual values.



$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \times 100\%. \quad (10)$$

In (8), (9) and (10),  $x_i$  is the actual value and  $y_i$  is the estimated value.

- d) The PCC, commonly represented as 'r' or Pearson's  $r$ , was employed in the testing data to determine the connection between predicted gait parameter values and the actual target gait parameter values. The PCC is a statistical measure used to assess the linear association between two continuous variables, providing insight into both the strength and direction of their relationship. Equation (11) for calculating the PCC between two variables,  $y_i$  and  $\hat{y}_i$ , with  $n$  data points,  $m_y$  and  $m_{\hat{y}}$  represent the respective average values of the  $y_i$  and  $\hat{y}_i$  variables.

$$r = \frac{\sum_{i=1}^n (y_i - m_y) (\hat{y}_i - m_{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - m_y)^2 \sum_{i=1}^n (\hat{y}_i - m_{\hat{y}})^2}} \quad (11)$$

## Data availability

Data available on request from the author, Shing-Hong Liu (shliu@cyut.edu.tw).

Received: 8 November 2024; Accepted: 25 March 2025

Published online: 12 April 2025

## References

1. Alam, M. N., Garg, A., Munia, T. T. K., Fazel-Rezai, R. & Tavakolian, K. Vertical ground reaction force marker for Parkinson's disease. *PLoS One*. **12**, e0175951 (2017).
2. Muro-de-la-Herran, A. & Garcia-Zapirain, B. Mendez-Zorrilla, A. Gait analysis methods: an overview of wearable and Non-Wearable systems, highlighting clinical applications. *Sensors* **14**, 3362–3394 (2014).
3. Victoria, L. & Chester, Edmund, N. Biden; Maureen tingley. Gait analysis. *Biomed. Instrum. Technol.* **39**, 64–74 (2005).
4. Chambers, H. G. & Sutherland, D. H. A practical guide to gait analysis. *J. Am. Acad. Orthop. Surg.* **10**, 222–231 (2002).
5. Carroll, K., Kennedy, R. A., Koutoulas, V., Bui, M. & Kraan, C. M. Validation of shoe-worn gait up Physilog<sup>5</sup> wearable inertial sensors in adolescents. *Gait Posture*. **91**, 19–25 (2022).
6. Zahradka, N. et al. An evaluation of three kinematic methods for gait event detection compared to the kinetic-based 'gold standard'. *Sens. (Switzerland)*. **20**, 1–15 (2020).
7. Merletti, R. & Parker, P. A. *Electromyography: Physiology, Engineering, and Noninvasive Applications* (Wiley, 2004). <https://doi.org/10.1002/0471678384>
8. Hof, A. L. The relationship between electromyogram and muscle force. *Sport · Sport*. **11**, 79–86 (1997).
9. Cechetto, A., Parker, P. & Scott, R. The effects of four time-varying factors on the mean frequency of a myoelectric signal. *J. Electromyogr. Kinesiol.* **11**, 347–354 (2001).
10. Sutherland, D. H. The evolution of clinical gait analysis part I: Kinesiological EMG. *Gait Posture*. **14**, 61–70 (2001).
11. Papagiannis, G. I., Roupelakis, I. M., Triantafyllou, A. I., Makris, I. N. & Babis, G. C. No differences identified in transverse plane biomechanics between medial Pivot and rotating platform total knee implant designs. *J. Arthroplasty*. **31**, 1814–1820 (2016).
12. Kumar, A., Godiyal, A. K., Joshi, P. & Joshi, D. A. New force Myography-Based approach for continuous Estimation of knee joint angle in lower limb amputees and Able-Bodied subjects. *IEEE J. Biomed. Heal Inf.* **25**, 701–710 (2021).
13. Hahn, M. E. & O'Keefe, K. B. A neural network model for Estimation of net joint moments during normal gait. *J. Musculoskelet. Res.* **11**, 117–126 (2008).
14. Rabe, K. G., Lenzi, T. & Fey, N. P. Performance of sonomyographic and electromyographic sensing for continuous Estimation of joint torque during ambulation on multiple terrains. *IEEE Trans. Neural Syst. Rehabil Eng.* **29**, 2635–2644 (2021).
15. Liu, J., Tao, Q. & Wu, B. Dynamic muscle fatigue state recognition based on deep learning fusion model. *IEEE Access*. **11**, 95079–95091 (2023).
16. Liu, G., Dobbins, C., D'Souza, M. & Phuong, N. A machine learning approach for detecting fatigue during repetitive physical tasks. *Pers. Ubiquitous Comput.* **27**, 2103–2120 (2023).
17. Tang, Z., Yu, H. & Cang, S. Impact of load variation on joint angle Estimation from surface EMG signals. *IEEE Trans. Neural Syst. Rehabil Eng.* **24**, 1342–1350 (2016).
18. Zou, X. et al. High-Fidelity sEMG signals recorded by an on-skin electrode based on AgNWs for hand gesture classification using machine learning. *ACS Appl. Mater. Interfaces*. **15**, 19374–19383 (2023).
19. Wang, Y., Zhao, P. & Zhang, Z. A deep learning approach using attention mechanism and transfer learning for electromyographic hand gesture Estimation. *Expert Syst. Appl.* **234**, 121055 (2023).
20. Vijayvargiya, A., Dhanka, B., Gupta, V. & Kumar, R. Implementation of machine learning algorithms for automated human gait activity recognition using sEMG signals. *Int. J. Biomed. Eng. Technol.* **42**, 150–166 (2023).
21. Haufe, S., Isaías, I. U., Pellegrini, F. & Palmisano, C. Gait event prediction using surface electromyography in parkinsonian patients. *Bioengineering* **10**, 212 (2023).
22. Zhang, F. et al. sEMG-based continuous Estimation of joint angles of human legs by using BP neural network. *Neurocomputing* **78**, 139–148 (2012).
23. Luo, R., Sun, S., Zhang, X., Tang, Z. & Wang, W. A. Low-Cost End-to-End sEMG-Based gait Sub-Phase recognition system. *IEEE Trans. Neural Syst. Rehabil Eng.* **28**, 267–276 (2020).
24. Morbidoni, C. et al. Machine-Learning-Based prediction of gait events from EMG in cerebral palsy children. *IEEE Trans. Neural Syst. Rehabil Eng.* **29**, 819–830 (2021).
25. Juan Cheng, X. Minfen Shen. A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals. *IEEE J. Biomed. Heal Inf.* **17**, 38–45 (2013).
26. Morbidoni, C., Cucchiarelli, A., Fioretti, S. & Di Nardo, F. A deep learning approach to EMG-Based classification of gait phases during level ground walking. *Electronics* **8**, 894 (2019).
27. Tigrini, A. et al. Phasor-Based myoelectric synergy features: A fast Hand-Crafted feature extraction scheme for boosting performance in gait phase recognition. *Sensors* **24**, 5828 (2024).
28. Mengarelli, A. et al. Myoelectric-Based Estimation of vertical ground reaction force during unconstrained walking by a stacked One-Dimensional convolutional long Short-Term memory model. *Sensors* **24**, 7768 (2024).

29. Lim, H. L., Ho, J. H. & Lee, K. Estimation of gait parameters using EMG signal with extreme learning machine. *J. Telecommun. Electron. Comput. Eng.* **10**, 61–65 (2018).
30. Ye, Y. et al. Online pattern recognition of lower limb movements based on sEMG signals and its application in real-time rehabilitation training. *Robotica* **42**, 389–414 (2024).
31. Gait Up, S. A. *Physilog - Digital Motion Analysis Platform*. (2013). <https://www.gaitup.com/>
32. Molinari, F., Knaflitz, M., Bonato, P. & Actis, M. V. Electrical manifestations of muscle fatigue during concentric and eccentric isokinetic knee Flexion-Extension movements. *IEEE Trans. Biomed. Eng.* **53**, 1309–1316 (2006).
33. Silva, B. M. C., Rodrigues, J. J. P. C., de la Torre Díez, I., López-Coronado, M. & Saleem, K. Mobile-health: A review of current state in 2015. *J. Biomed. Inf.* **56**, 265–272 (2015).
34. How to use the Blood Oxygen app on Apple Watch. <https://support.apple.com/en-us/HT211027>
35. Liu, S. H. et al. An EMG patch for the Real-Time monitoring of Muscle-Fatigue conditions during exercise. *Sensors* **19**, 3108 (2019).
36. Mobarak, R. et al. A minimal and Multi-Source recording setup for ankle joint kinematics Estimation during walking using only proximal information from lower limb. *IEEE Trans. Neural Syst. Rehabil. Eng.* **32**, 812–821 (2024).
37. Liu, S. H. et al. Estimation of gait parameters for adults with surface electromyogram based on machine learning models. *Sensors* **24**, 734 (2024).
38. Zhang, H., Guo, Y. & Zanotto, D. Accurate ambulatory gait analysis in walking and running using machine learning models. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 191–202 (2020).
39. Hao, M., Chen, K. & Fu, C. Smoother-Based 3-D foot trajectory Estimation using inertial sensors. *IEEE Trans. Biomed. Eng.* **66**, 3534–3542 (2019).
40. van Hedel, H. J. A., Tomatis, L. & Müller, R. Modulation of leg muscle activity and gait kinematics by walking speed and bodyweight unloading. *Gait Posture* **24**, 35–45 (2006).
41. Crozara, L. F. et al. Utility of electromyographic fatigue threshold during treadmill running. *Muscle Nerve* **52**, 1030–1039 (2015).
42. Liu, S. H., Chang, K. M. & Cheng, D. C. The progression of muscle fatigue during exercise Estimation with the aid of High-Frequency component parameters derived from ensemble empirical mode decomposition. *IEEE J. Biomed. Heal. Inf.* **18**, 1647–1658 (2014).
43. Asghari Oskoei, M. & Hu, H. Myoelectric control systems—A survey. *Biomed. Signal. Process. Control* **2**, 275–294 (2007).
44. Phinyomark, A. et al. EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Syst. Appl.* **40**, 4832–4840 (2013).
45. Phinyomark, A., Hirunviriya, S., Limsakul, C. & Phukpattaranont, P. Evaluation of EMG feature extraction for hand movement recognition based on Euclidean distance and standard deviation. *ECTI-CON 2010–2010 ECTI int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol.* 856–860 (2010).
46. Putro, N. A. S., Avian, C., Prakosa, S. W., Mahali, M. I. & Leu, J. S. Estimating finger joint angles by surface EMG signal using feature extraction and transformer-based deep learning model. *Biomed. Signal. Process. Control* **87**, 105447 (2024).
47. Zhang, X. & Zhou, P. Sample entropy analysis of surface EMG for improved muscle activity onset detection against spurious background spikes. *J. Electromyogr. Kinesiol.* **22**, 901–907 (2012).
48. Prokhorenkova, L. et al. Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **31**, 6638–6648 (2018).
49. Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. *A Comparative Analysis of Gradient Boosting Algorithms. Artificial Intelligence Review* vol. 54 Springer Netherlands, (2021).
50. Breiman, L. *Random Forests*. vol. 45 (2001).
51. Biau, G. Analysis of a random forest model. *J. Mach. Learn. Res.* **13**, 1063–1095 (2012).
52. Chen, T., Guestrin, C. & XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* vol. 42 785–794 ACM, New York, NY, USA, (2016).
53. Pan, S., Zheng, Z., Guo, Z. & Luo, H. An optimized XGBoost method for predicting reservoir porosity using petrophysical logs. *J. Pet. Sci. Eng.* **208**, 109520 (2022).

## Author contributions

S-H.L.: Writing – review & editing, validation, methodology, conceptualization, supervision. A.K.S.: Writing – original draft, formal analysis, visualization, methodology. B-Y.W.: Data curation, software. C.-J. C.: Data curation. X. Z. & J.-J. W.: Writing – review & editing. All authors reviewed the manuscript.

## Funding

This study was supported by the National Science and Technology Council of Taiwan under the grant NSTC 111-2221-E-324 -003 -MY3 and NSTC 113-2221-E-214 -005.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to A.K.S. or J.-J.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025