Machine Learning for VO_2max Predictions: A Comparison of Methods using Wearable Sensor Data

Aron B. Syversen*[1], Alexios Dosis[2], Zhiqiang Zhang[3], David Jayne[2], and David Wong[2]

Abstract— **Cardiopulmonary exercise testing is the gold standard for assessing VO_2max, but it is costly in terms of time and personnel. Its limitations drive the need for alternative methods of assessment. Using physiological measurements such as heart rate, several machine learning prediction models have been developed to estimate VO_2max. This paper provides the first direct comparison of multiple modelling approaches in a clinical population using wearable sensor data. Wearable ECG and accelerometer data were first pre-processed. We used a signal quality index for ECG data and ML-based physical activity classification. We then extracted known useful features, based on previous literature. Five models (Multiple-Linear Regression (MLR), Support Vector Regression, Random Forest, XGBoost, Multi-layer Perceptron) were compared using 5-fold cross-validation, with performance evaluated via RMSE, R^2, correlation, and SEE. MLR outperformed other models in predicting VO_2max (R = 0.68±0.09, RMSE = 3.35±0.32). Overall performance in this clinical population was lower than in studies using exercise-derived features in a healthy individuals, but shows that wearable sensor data, including heart rate variability features, can still provide meaningful insight for VO_2max estimations.**

Clinical relevance— **This study shows how a linear model can estimate VO_2max from ECG and accelerometer data. This model offers better interpretability to more sophisticated machine learning approaches with no cost in performance in this case.**

I. INTRODUCTION

Cardiopulmonary exercise testing is the gold-standard method to measure cardio-respiratory fitness (CRF) [1]. Specifically, it measures the maximal oxygen uptake (VO_2) attained during exhaustive exercise. Research has established VO_2max as one of the strongest predictors of morbidity and mortality across various populations [2]. Consequently, CPET has become an essential tool in clinical settings.

The perioperative setting, particularly thoraco-abdominal surgery, has widely adopted CPET as a standard preoperative assessment tool [3]. Combined with clinical measures, CPET predicts postoperative outcomes, aiding risk stratification and resource allocation. However, its high cost, need for trained clinicians and controlled environments limit accessibility and widespread use [4]. Consequently, research has focused on estimating VO_2max without CPET.

One approach involves sub-maximal testing, such as the 6-minute walk test. However, these methods are less accurate and still require visits to a preoperative clinic for structured assessment [5]. Recent advances in wearable sensor technology offer a promising alternative that allows continuous monitoring of vital signs in free-living environments [4]. This approach provides a more comprehensive view of the daily movements and physiological responses of patients compared to the snapshot provided by short-term tests.

Research has demonstrated the potential of features from wearable sensors, including resting heart rate (RHR) and physical activity measures, to predict VO_2max [6][7]. Evidence also suggests that heart rate variability (HRV) may be associated with VO_2max [8]. While HRV analysis traditionally required precise beat detection, novel methods have been developed to overcome this challenge in wearable sensors [9].

Various analytical approaches have been employed to predict VO_2max [10]. Multiple linear regression (MLR) is a commonly used effective method. More recently, machine learning methods have gained popularity, with multi-layer perceptrons (MLP), support vector machines (SVM) and random forests (RF) emerging as potentially superior models. Although these studies have reported high accuracy, they generally use healthy participant data. A comprehensive comparison of these methods, particularly in clinical populations where VO_2max is commonly assessed, is lacking.

We address this gap by conducting a comparison of machine learning methods for predicting VO_2 in a sample of preoperative patients. Specifically, we combine demographic data with features extracted from wearable sensor recordings to evaluate the performance of five machine learning models.

II. MATERIALS AND METHODS

A. Dataset

This research used data from the REMOTES study, an observational clinical trial completed at Leeds Teaching Hospitals NHS Trust (December 2022-September 2024). The experimental procedures involving human subjects described in this paper were approved by the UK HRA (REC 22/SS/0050).

All participants were adults scheduled for elective major abdominal surgery and attending preoperative CPET. Major abdominal surgery was defined as procedures classified as Major 1 to Major +5 according to the British United Provident Association (BUPA) classification of procedures (e.g. colectomy, total pancreatectomy) [11]. Eligible participants were undergoing surgery for a range of benign and malignant conditions and were referred for CPET as part of routine preoperative assessment. They received trial information and

*This work was supported by EPSRC grant number EP/S024336/1.
[1]School of Computing, University of Leeds, UK a.b.syversen@leeds.ac.uk
[2]School of Medicine, University of Leeds, UK
[3]School of Electrical Engineering, University of Leeds, UK

provided written informed consent. They wore the Ubiqvue-Lifesignals LX1550E chest sensor continuously for 72 hours, which included a 2-lead electrocardiogram (ECG) (244.14 Hz) and a triaxial accelerometer (25 Hz). Baseline demographic and physiological data were collected including age, gender, body mass index (BMI), and VO_2max (ml/kg/min). Participants with less than 24 hours of recorded data were excluded from analysis.

B. Signal Processing

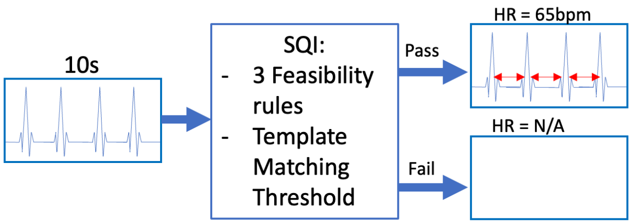

Fig. 1. Figure to show the implementation of the SQI. The ECG is portioned into 10s-segments before passing each segment through the SQI tool. If passed and labelled as acceptable, HR and RR-intervals were extracted from the segment.

1) ECG signals: Wearable ECG sensor data is prone to motion artefacts. To minimise noise, we used Orphanidou's signal quality index (SQI) to extract periods of 'acceptable' signal after testing a range of approaches [12][13]. Of the two ECG leads, the ECG lead with the most 'acceptable' segments across participants was used. Heart rate (HR) was extracted from 'acceptable' windows after the signal was passed through the SQI (Figure 1); the final HR in beats-per-minute was computed as the mean value across segments containing only 'acceptable' ECG data. This feature extraction process ensured that only high-quality physiological data were used.

2) Accelerometer signals: Accelerometer data was processed using the Biobank Accelerometer Analysis Toolkit, categorising physical activity (PA) into five behavioural states: Sleep, Sedentary Behaviour, Light PA (LPA), and Moderate-to-Vigorous PA (MVPA) and Vigorous PA [14].

Step count extraction employed an open-source wavelet-based algorithm aggregating to per-minute step-counts [15].

C. Feature Extraction

1) Movement Features:

- Daily Activity Distribution - We computed the average daily time spent across PA classifications. Total duration was divided by valid device wear resulting in mean daily values for LPA, MVPA and VPA.
- Daily Step Metrics - We calculated the mean number of steps taken per day, considering only valid wear time.
- MVPA Walking Duration - We identified periods of MVPA walking by isolating minutes where steps surpassed 100 per minute, indicative of MVPA intensity, before computing the average daily duration [16].

2) HR Features:

- Resting HR - We estimated RHR by calculating the mean HR recorded between 3 AM-7 AM, during periods classified as 'sleep' or 'sedentary', adhering to established protocols for deriving RHR [17].
- Average active HR - Average HR recorded during LPA, MVPA and VPA.
- Minimum and Maximum HR - Minimum and maximum HR recorded by device.
- Step-to-Heart Rate Analysis - We aligned HR data with per-minute step counts, focusing on active periods (step count>0). We calculated step count-to-heart rate ratios for these minutes, known to predict VO_2max [7]. We derived the 25th, 50th, 75th, and 95th percentiles of these ratios to extract four features.

3) HRV features: HRV features were categorised into two groups: short-term and long-term HRV [18].

- Short-term HRV - We developed a pipeline to extract HRV features from noisy wearable ECG data. We extracted ECG between 3 a.m. to 7 a.m. sleep to minimise external influences [19]. From this, The 5-minute ECG segment with most 'acceptable' labels was identified. After preprocessing to handle ectopic beats (and removing those with >50 ectopic beats), we calculated ten short-term HRV measures using the hrv-analysis package: four time-domain features (SDNN, RMSSD, pNN50%, MeanNN), four frequency-domain features (VLF, LF, HF, LF/HF) and two non-linear features (SD1, SD2) [20].
- Long-term HRV - We analysed recordings from day 1 due to minimal data loss (17.5% average). Participants required >16 hours of valid HR data to ensure comprehensive circadian rhythm assessment, aligning with recommendations for capturing full day-night autonomic activity cycles [18]. From this we extracted two long-term HRV features:
  SDNN24 - Calculated from all RR intervals in day 1:

$$\mathrm{SDNN24} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(RR_i - \overline{RR})^2} \tag{1}$$

  Where N is the total number of RR intervals, RR_i is the i-th RR interval, and $\overline{RR}$ is the mean of all RR intervals.
  $\mathrm{SDANN_{HR}24}$ - To account for noise and missing data, a modified version of SDNN was calculated using 60 minute segment lengths [9]:

$$\mathrm{SDANN_{HR}24} = \sqrt{\frac{\sum_{i=1}^{N_{\text{segments}}}\left(60/\mathrm{HR}_i - \frac{\sum_{i=1}^{N_{\text{segments}}} 60/\mathrm{HR}_i}{N_{\text{segments}}}\right)^2}{N_{\text{segments}}}} \tag{2}$$

D. Prediction Model Development

Using these 27 features combined with age, gender and BMI, we selected five different models to predict VO_2max based on models commonly implemented [10]:

- Multiple Linear Regression (MLR)
- Support Vector Regression (SVR)
- Random Forest (RF)
- XGBoost
- Multi-layer Perceptron (MLP)

E. Model Development

To prevent multi-collinearity, highly correlated features were removed (>0.9). We employed participant-level 5-fold cross-validation for model development and evaluation. The dataset was partitioned into five folds, with approximately 20% of participants held out for testing in each fold, and each participant appearing in the test set once. For each iteration, models were trained on four folds and tested on the remaining fold to allow independent evaluation on unseen participants. This allowed us to maximize the use of the small dataset while maintaining data separation between training and testing. Within each training set, feature selection (for Linear Regression and SVR) and hyper parameter tuning (for Random Forest, XGBoost, SVR, and MLP) were performed using internal cross-validation to prevent data leakage.

The five models were developed as follows:

- **Linear Regression:** LASSO-selected features.
- **Random Forest:** All features, parameters: estimators (100-500), max depth (10-50 or None).
- **XGBoost:** All features, parameters: estimators (100-500), max depth (3-8), learning rate (0.01-0.2).
- **SVR:** LASSO-selected features, parameters: C (0.1-100).
- **MLP:** All features, parameters: hidden layer sizes (3-11 neurons), learning rate (0-1), momentum (0-1), logistic activation.

For each model (excluding MLR), we performed 20 iterations of RandomizedSearchCV with 3-fold cross-validation.

F. Model Evaluation

Performance was assessed using four metrics and averaged across folds [10]:

1) Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

2) Standard Error of the Estimate (SEE):

$$\text{SEE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}} \tag{4}$$

3) R-squared (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

4) Pearson Correlation Coefficient (r): Measures the linear correlation between predicted and observed values.

Where y_i are the observed values, $\hat{y}_i$ are the predicted values, $\bar{y}$ is the mean of observed values, n is the number of samples, and k is the number of features in the model.

III. Results

198 participants were recruited into the REMOTES trial, of which 169 participants had sufficient data for analysis.

TABLE I
DESCRIPTIVE STATISTICS (MEAN $\pm$ STANDARD DEVIATION) BY GENDER.

Variable	Men (n=125)	Women (n=44)
Age (years)	68.78 $\pm$ 10.17	67.16 $\pm$ 13.37
BMI (kg/m^2)	28.26 $\pm$ 5.05	30.27 $\pm$ 7.49
Collected data (hours)	72.51 $\pm$ 14.34	73.73 $\pm$ 18.28
VO_2max (ml/kg/min)	18.73 $\pm$ 4.72	15.25 $\pm$ 3.70
Average Daily Step Count	3836$\pm$ 3043	2778 $\pm$ 2215

A. Model Development

HR during MVPA and LPA were removed due to insufficient data. RMSSD, SD1, SDNN and SDNN24 were removed due to high multi-collinearity, leaving 25 features for analysis. Daily average steps, gender, BMI, age, SDANN_{HR}24 and LF/HF had non-zero coefficients from the LASSO regression in every fold of the cross-validation.

B. Model evaluation

Aside from Pearson's correlation, the MLR model had the highest performance across metrics (RMSE = 3.35; R^2 = 0.46, SEE = 3.95). When assessing correlation, MLR had a performance equal to SVR, which was the second highest performing model across other metrics (Table 2). The XGBoost model was the worst performing across all metrics.

TABLE II
MODEL PERFORMANCE METRICS ACROSS THE FIVE FOLDS (MEAN $\pm$ STD). THE HIGHEST PERFORMANCE IN EACH METRIC IS EMBOLDENED.

Model	RMSE	R^2	Correlation	SEE
MLR	**3.35 $\pm$ 0.32**	**0.46 $\pm$ 0.13**	**0.68 $\pm$ 0.09**	**3.95 $\pm$ 0.39**
RF	3.82 $\pm$ 0.33	0.31 $\pm$ 0.13	0.61 $\pm$ 0.13	7.94 $\pm$ 0.61
XGBoost	3.86 $\pm$ 0.42	0.30 $\pm$ 0.14	0.59 $\pm$ 0.12	8.04 $\pm$ 0.82
SVR	3.40 $\pm$ 0.23	0.45 $\pm$ 0.11	**0.68 $\pm$ 0.08**	4.01 $\pm$ 0.27
MLP	3.69 $\pm$ 0.34	0.32 $\pm$ 0.05	0.63 $\pm$ 0.11	7.68 $\pm$ 0.76

IV. Discussion

This study compared the performance of several ML models in predicting VO_2max from wearable sensor data in preoperative participants. Results indicate that MLR performed best overall, achieving the highest R^2 (R^2 = 0.46 $\pm$ 0.13) and the lowest RMSE (3.35 $\pm$ 0.32) and SEE (3.95 $\pm$ 0.39). SVR had similar predictive performance. This study is the first to directly compare commonly used ML models in a clinical population.

The improved performance of MLR over ML models may be attributed to sample characteristics. Previous studies in healthy populations have reported that MLP models outperform linear models [10]. However, this may not generalize well to a heterogeneous clinical cohort. Age and gender are known to be among the strongest predictors of VO_2max, and the wide distribution of age and step count values in this

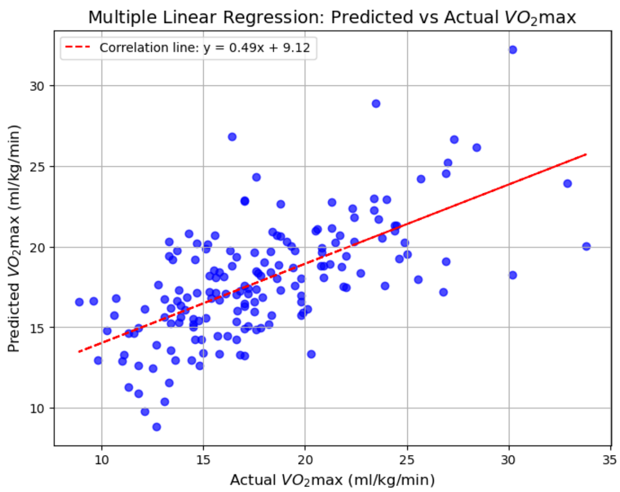

Fig. 2. Plot to show the correlation between the predicted and actual VO_2max values from the MLR model across participants.

cohort may have reinforced the linearity of these relationships. Additionally, in a clinical sample, a physical activity measure like step counts may not only quantify fitness but also indirectly reflect a participants overall health status. For example, individuals with fewer co-morbidities may have a higher functional status and therefore achieve higher daily step counts, rather than step count purely representing behavioural activity.

A predominance of male participants and the presence of varying co-morbidities may have influenced model performance, potentially favouring simple models that better capture dominant linear trends across the cohort. We considered developing separate models by sex; however, this approach has not been commonly implemented in previous research and would limit comparability. Additionally, clinical thresholds for VO_2max used in preoperative risk assessment do not differ by sex, and sex was included as an input feature across models to account for sex-specific differences. Maintaining a single model supports clinical application while ensuring sex-related variability is incorporated into model development.

The prediction performance in this study was lower than the highest-performing ML models reported in previous research (R>0.8) [10]. This discrepancy may be explained by differences in predictor variables; many high-performing VO_2max models include exercise test-derived variables. By including variables obtained during some form of exercise, even if sub-maximal, these are more likely to provide useful indication or capture aspects of maximal exercise capacity more directly than wearable-derived resting or free-living data. However, in clinical populations scheduled for major abdominal surgery, exercise test-derived variables are not always feasible and wearable sensors provide an accessible alternative despite this limitation. When comparing against other research using only free-living data in healthy populations, results are more comparable (r = 0.8) [6]. Additionally, this study's dataset represents a broader demographic with a wider range in fitness levels and health conditions, which may introduce further variation.

Despite lower average performance, this study achieved lower RMSE and comparable performance to some previous models using healthy participants in controlled or semi-controlled settings [21]. Notably, it is the first to incorporate HRV features from wearable sensors into VO_2max prediction. HRV parameters were consistently selected in LASSO regression for the highest-performing models, suggesting that they add meaningful information from free-living data. Since HRV reflects autonomic function, incorporating HRV features may further enhance predictive models.

This study has several limitations. While the sample includes diverse ages and health conditions, it is limited to a preoperative population, restricting generalisability. A gender imbalance may further impact this. Additionally, HRV features were collected in free-living conditions, where movement artifacts could affect measurements. Future research should further explore HRV's role in CRF prediction and its reliability in free-living settings.

This study also contributes to the development of robust ECG preprocessing pipelines for wearable data by outlining how an SQI can be implemented to support accurate HR estimation. While this provides a starting point, future work could benefit from a more formal framework for SQI development. In particular, signal quality assessment should be tailored to the specific goals of the processing pipeline—whether beat detection, HRV analysis, or signal morphology classification—rather than relying on general-purpose quality metrics. Embedding task-specific SQI frameworks into analysis pipelines could enhance both the interpretability and reliability of wearable ECG research, especially in clinical contexts.

V. Conclusion

Linear models outperformed ML models in predicting VO_2max in this preoperative population, contrasting with findings in healthy groups. This suggests that demographic and physiological predictors may exhibit stronger linear relationships in clinical settings. Future research should explore whether adding clinical variables improves predictions and further assess the role of HRV in VO_2max estimation.

References

[1] R. Ross, S. N. Blair, R. Arena, T. S. Church, J. P. Després, B. A. Franklin, W. L. Haskell, L. A. Kaminsky, B. D. Levine, C. J. Lavie, J. Myers, J. Niebauer, R. Sallis, S. S. Sawada, X. Sui, and U. Wisløff, "Importance of assessing cardiorespiratory fitness in clinical practice: A case for fitness as a clinical vital sign: A scientific statement from the american heart association," *Circulation*, vol. 134, pp. e653–e699, 12 2016. [Online]. Available: https://www.ahajournals.org/doi/10.1161/CIR.0000000000000461

[2] J. J. Lang, S. A. Prince, K. Merucci, C. Cadenas-Sanchez, J. P. Chaput, B. J. Fraser, T. Manyanga, R. McGrath, F. B. Ortega, B. Singh, and G. R. Tomkinson, "Cardiorespiratory fitness is a strong and consistent predictor of morbidity and mortality among adults: an overview of meta-analyses representing over 20.9 million observations from 199 unique cohort studies," *British Journal of Sports Medicine*, vol. 58, pp. 556–566, 5 2024. [Online]. Available: https://bjsm.bmj.com/content/58/10/556https://bjsm.bmj.com/content/58/10/556.abstract

[3] D. Z. Levett, S. Jack, M. Swart, J. Carlisle, J. Wilson, C. Snowden, M. Riley, G. Danjoux, S. A. Ward, P. Older, and M. P. Grocott, "Perioperative cardiopulmonary exercise testing (cpet): consensus clinical guidelines on indications, organization, conduct, and physiological interpretation," *British journal of anaesthesia*, vol. 120, pp. 484–500, 3 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29452805/

[4] A. Syversen, A. Dosis, D. Jayne, and Z. Zhang, "Wearable sensors as a preoperative assessment tool: A review," *Sensors*, vol. 24, p. 482, 1 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/2/482/htmhttps://www.mdpi.com/1424-8220/24/2/482

[5] M. Guazzi, K. Dickstein, M. Vicenzi, and R. Arena, "Six-minute walk test and cardiopulmonary exercise testing in patients with chronic heart failure: A comparative analysis on clinical and prognostic insights," *Circulation: Heart Failure*, vol. 2, pp. 549–555, 11 2009. [Online]. Available: https://www.ahajournals.org/doi/10.1161/CIRCHEARTFAILURE.109.881326

[6] M. C. M. Frade, T. Beltrame, M. de Oliveira Gois, A. Pinto, S. C. G. de Moura Tonello, R. da Silva Torres, and A. M. Catai, "Toward characterizing cardiovascular fitness using machine learning based on unobtrusive data," *PLOS ONE*, vol. 18, 8 2023. [Online]. Available: /pmc/articles/PMC9980797//pmc/articles/PMC9980797/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9980797/

[7] A. Neshitov, K. Tyapochkin, M. Kovaleva, A. Dreneva, E. Surkova, E. Smorodnikova, and P. Pravdin, "Estimation of cardiorespiratory fitness using heart rate and step count data," *Scientific Reports 2023 13:1*, vol. 13, pp. 1–13, 9 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-43024-x

[8] V. Vukomanovic, J. Suzic-Lazic, V. Celic, C. Cuspidi, T. Petrovic, S. Ilic, D. Skokic, D. A. Morris, and M. Tadic, "Association between functional capacity and heart rate variability in patients with uncomplicated type 2 diabetes," *Blood pressure*, vol. 28, pp. 184–190, 5 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30836775/

[9] D. Morelli, A. Rossi, L. Bartoloni, M. Cairo, and D. A. Clifton, "Sdnn24 estimation from semi-continuous hr measures," *Sensors 2021, Vol. 21, Page 1463*, vol. 21, p. 1463, 2 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/4/1463/htmhttps://www.mdpi.com/1424-8220/21/4/1463

[10] A. Ashfaq, N. Cronin, and P. Müller, "Recent advances in machine learning for maximal oxygen uptake (vo2 max) prediction: A review," *Informatics in Medicine Unlocked*, vol. 28, p. 100863, 1 2022.

[11] "Schedule of procedures," *Bupa*, 12 2022. [Online]. Available: https://codes.bupa.co.uk/home

[12] A. Syversen, Z. Zhang, J. Batty, M. Kaisti, D. Jayne, and D. C. Wong, "Assessment of ecg signal quality index algorithms using synthetic ecg data," 12 2024. [Online]. Available: https://www.cinc.org/archives/2024/pdf/CinC2024-270.pdf

[13] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring," *IEEE journal of biomedical and health informatics*, vol. 19, pp. 832–838, 5 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25069129/

[14] R. Walmsley, S. Chan, K. Smith-Byrne, R. Ramakrishnan, M. Woodward, K. Rahimi, T. Dwyer, D. Bennett, and A. Doherty, "Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease," *Br J Sports Med*, vol. 0, pp. 1–10, 2021. [Online]. Available: http://bjsm.bmj.com/

[15] M. Straczkiewicz, N. L. Keating, E. Thompson, U. A. Matulonis, S. M. Campos, A. A. Wright, and J. P. Onnela, "Open-source, step-counting algorithm for smartphone data collected in clinical and nonclinical settings: Algorithm development and validation study," *JMIR cancer*, vol. 9, 1 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37966891/

[16] C. Tudor-Locke, H. Han, E. J. Aguiar, T. V. Barreira, J. M. Schuna, M. Kang, and D. A. Rowe, "How fast is fast enough? walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review," *British Journal of Sports Medicine*, vol. 52, pp. 776–788, 6 2018. [Online]. Available: https://bjsm.bmj.com/content/52/12/776https://bjsm.bmj.com/content/52/12/776.abstract

[17] C. Speed, T. Arneil, R. Harle, A. Wilson, A. Karthikesalingam, M. McConnell, and J. Phillips, "Measure by measure: Resting heart rate across the 24-hour cycle," *PLOS Digital Health*, vol. 2, p. e0000236, 4 2023. [Online]. Available: https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000236

[18] M. Malik, T. J. Bigger, J. A. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Shwartz, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, pp. 354–381, 3 1996. [Online]. Available: https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

[19] B. Israel, D. J. Buysse, R. T. Krafty, A. Begley, J. Miewald, and M. Hall, "Short-term stability of sleep and heart rate variability in good sleepers and patients with insomnia: For some measures, one night is enough," *Sleep*, vol. 35, pp. 1285–1291, 9 2012. [Online]. Available: https://dx.doi.org/10.5665/sleep.2088

[20] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, 9 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29034226/

[21] M. M. H. Shandhi, W. H. Bartlett, J. A. Heller, M. Etemadi, A. Young, T. Plotz, and O. T. Inan, "Estimation of instantaneous oxygen uptake during exercise and daily activities using a wearable cardio-electromechanical and environmental sensor," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, pp. 634–646, 3 2021.