Sensor Applications

Predicting Health-Related Physical Fitness from Smartwatch Data

Maiko Lie[1], Thomas Beltrame[1], Vinicius T.D.M. da Costa[1], Gabriel T. Alves[1], Mariana de O. Gois[1], Eduardo K. Chikui[1], Juscelino L. de O. Junior[1], Paula R. Pinto[1], and Marcus de Assis Angeloni[1]

[1] *Samsung R&D Institute Brazil (SRBR), Campinas, SP, Brazil*

Manuscript received xxxx x, xxxx; revised xxxx xx, xxxx; accepted xxxx x, xxxx. Date of publication xxxx xx, xxxx; date of current version xxxx xx, xxxx.

Abstract—Measuring health-related physical fitness (HRPF) is important due to its association with overall health, enabling disease prevention, personalized exercise prescription, among other applications. HRPF measurement is often performed on a laboratory setting, using clinical equipment under professional supervision, but has become more accessible due to wearable technology. For instance, current generation smartwatches are capable of predicting two of the five HRPF components acknowledged by the American College of Sports Medicine (ACSM), namely cardiorespiratory fitness ($\dot{V}O_2$max) and body composition (body fat %). In this paper, we present approaches for prediction of the remaining three HRPF components: muscular endurance, muscular strength, and flexibility, allowing a complete characterization of health-related physical fitness solely from smartwatch data. We adopted a component-specific approach, as opposed to more general approaches used in previous work, and show that our approach is more effective on data from wearable devices. Furthermore, we show that free-living activity data, exclusively accessible through wearables, is highly discriminative for HRPF prediction. This demonstrates that current consumer-grade wearables enable more than merely approximating clinical data — they also complement it with real-world activity information.

Index Terms—Biomarkers, fitness, health, health-related physical fitness, smartwatch, wearables.

I. INTRODUCTION

The ability to perform activities of daily living with vigor characterizes health-related physical fitness (HRPF), which is important due to its association with overall health [1]. The American College of Sports Medicine (ACSM) acknowledges five HRPF components (cardiorespiratory fitness, body composition, muscular endurance, muscular strength, and flexibility), which are measured by standardized fitness tests that are performed using clinical equipment and under the professional supervision. Due to its relevance as health indicator, HRPF prediction has been extensively investigated [2]–[4], but restricted to anthropometric data or measurements from clinical equipment. While most work is limited to individual HRPF domains, recent studies have leveraged large-scale clinical data to learn predictors for all five HRPF components acknowledged by the ACSM [5], [6], using general approaches such as multiple linear regression (MLR) and multilayer perceptrons (MLP).

Wearable sensing devices are capable of non-invasive estimation of biomarkers under free-living conditions, which enable several mobile health applications, including HRPF tracking. For instance, current smartwatches are capable of predicting cardiorespiratory fitness (maximum oxygen uptake, $\dot{V}O_2$max) and measuring body composition (body fat %) [7], two of the five HRPF components acknowledged by the ACSM [1]. However, only all five components together can provide a complete characterization of health-related physical fitness. In this paper, we present prediction models for the remaining three components — muscular strength, muscular endurance, and flexibility — using only smartwatch data, enabling the characterization of all five HRPF components under free-living solely from the smartwatch. We present domain-specific prediction models (Fig. 1) and demonstrate that, in our setting, this outperforms more general approaches employed in previous work. We show that while for muscular strength prediction a simple linear regression on profile and bioelectrical impedance analysis (BIA) features is already effective, muscular endurance and flexibility prediction can be substantially improved by leveraging free-living activity data, as long as appropriate preprocessing is employed. This highlights the advantage of wearable devices to go beyond just approximating clinical data and actually complementing them with free-living information.

Our main contributions are: i) predicting, from consumer-grade smartwatch data, HRPF components that have been previously predicted only from clinical equipment; ii) demonstrating that free-living activity data from smartwatches can improve HRPF prediction; iii) introducing HRPF prediction models designed to leverage activity features, which outperform previous approaches on smartwatch data and are appropriate for wearables since they are just linear regressions, thus interpretable, computationally efficient, and compact.

Corresponding author: M. Lie (e-mail: maiko.l@samsung.com).
Associate Editor: .
Digital Object Identifier xx.xxxx/LSENS.xxxx.xxxxxxx

II. PROPOSED APPROACH

A. Background

Previous work used the same kind of model for all HRPF components, without considering component-specific design [5], [6]. However, they leveraged measurements from clinical equipment and a large-scale dataset (190k+ participants). In contrast, we use measurements from smartwatches, which are noisier and lead to a relatively small-scale dataset (less than 500 participants per HRPF component) due to the high cost of data collection with smartwatch data monitoring. This motivates the design of component-specific models, which introduce constraints explicitly since learning them from small-scale data is challenging and often unreliable.

TABLE 1. Feature groups adopted in the experiments. These features include variables available from the *Samsung Galaxy Watch 4* and some of their interactions. Unlike previous work [6] we did not have access to actual waist circumference measurements since we use only smartwatch data, so we used prediction equations based on user profile data [8]. BMI: body mass index.

Group	Features
Profile	age, sex, height, weight, BMI, (BMI · age)$^{-1}$, waist circumference
BIA	body fat %, basal metabolic rate/weight, skeletal muscle
Activity	exercise calorie (max) daily active calorie (min, max, mean, std) daily active time (min, max, mean, std) daily step count (min, max, mean, std) daily step speed (min, max, mean, std)

B. Problem statement

We are interested in learning prediction models for three HRPF components (muscular endurance, muscular strength, and flexibility) solely from features extracted from smartwatch data, which include user profile, bioelectrical impedance analysis (BIA), and activity data (Table 1). Each participant is represented by a single feature vector, where activity variables are aggregations (min, max, mean, std) over the free-living period. BIA variables are computed from the measurement made on the last day that the ground-truth HRPF tests were performed. Our models are linear regressions of the form:

$$\hat{y} = \boldsymbol{x}^\top \boldsymbol{\beta}, \tag{1}$$

where $\hat{y}$ is a real-valued prediction of an HRPF component, $\boldsymbol{x}$ is an N-dimensional vector representing an instance of features preprocessed from smartwatch inputs (and a constant bias term), and $\boldsymbol{\beta}$ is an N-dimensional vector of regression coefficients. Since each HRPF component has different data characteristics and assumptions, their preprocessing might also differ, thus the dimensionality of $\boldsymbol{x}$ and $\boldsymbol{\beta}$ might also differ between component predictors.

C. Model definition

Our models are linear regressions with component-specific preprocessing and optimization criteria. A diagram with an overview of the proposed HRPF prediction models are presented in Figure 1, and a detailed description of the preprocessing and optimization functions used to learn the regression coefficients of each HRPF component is presented next.

For each HRPF component, let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ be the matrix of preprocessed feature vectors of M instances with N features, including a bias constant, and $\boldsymbol{y} \in \mathbb{R}^{M \times 1}$ be the vector of ground-truth values. The coefficients $\boldsymbol{\beta}$ of the muscular strength regression are learned by solving:

$$\min_{\boldsymbol{\beta}} \frac{1}{2M} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_1, \tag{2}$$

where M is the number of instances, the second term is the sparsity-inducing L_1 (Lasso) penalty [9], and α is the regularization weight. We adopted this penalty because there was no improvement for a wide range of model design choices, suggesting that it is reasonable to simply adopt a more parsimonious model. For this predictor, preprocessing consisted of simply z-scoring the features.

Muscular endurance prediction improved with activity features but was more effective with specific preprocessing. Thus, besides z-scoring all features, we found that transforming activity features using cubic spline smoothing [10] enabled leveraging them more effectively.

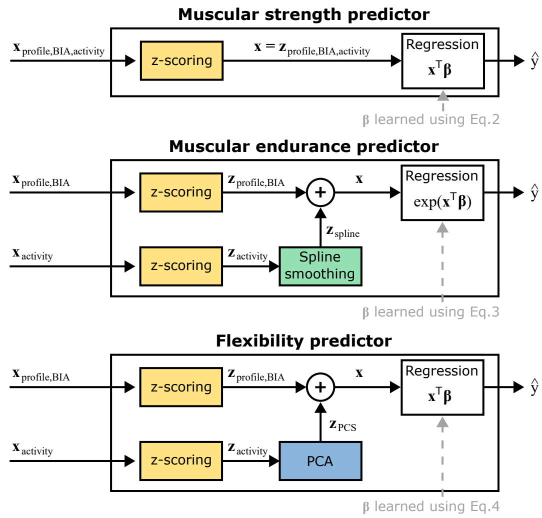

Fig. 1. Overview of our proposed Health-related Physical Fitness (HRPF) prediction models.

This transformation is often recommended to incorporate non-linear relationships between continuous features and the target variable [11]. Moreover, since the ground truth for this HRPF component is the number of push-ups, which is a count variable, we use Poisson regression since it assumes skewed, non-negative, distributions that arise from this type of data [11]. Thus, the coefficients $\boldsymbol{\beta}$ of the muscular endurance regression are learned by solving:

$$\min_{\boldsymbol{\beta}} \frac{1}{M} \sum_{i}^{M} \left(y_i \log \frac{y_i}{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})} - y_i + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \right) + \frac{\alpha}{2} \|\boldsymbol{\beta}\|_2^2, \tag{3}$$

where M is the number of instances, the second term is the L_2 (Ridge) penalty [12], and α is the regularization weight. We avoid L_1 penalty on this model since feature selection on spline basis functions can compromise interpretability and introduce discontinuities.

For flexibility prediction, we noticed that activity features improve results but, similarly to muscular endurance, need specific preprocessing to be leveraged effectively. Thus, besides z-scoring all features, we preprocess activity features by projecting them onto latent features via Principal Component Analysis (PCA) [13] before fitting the regression. We set the number of components as the smallest number that explains at least 80% of the variance in the data. The coefficients $\boldsymbol{\beta}$ of the flexibility regression are then learned by solving:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_2^2, \tag{4}$$

where the second term is the L_2 (Ridge) penalty [12], and α is the regularization weight. This choice of penalty was motivated by the fact that we project activity features via PCA, thus already performing dimensionality reduction. Reducing it further through sparseness via L_1 penalty would not only reduce interpretability, since higher order principal components could be eliminated without the same happening to lower order components, but could also degrade performance.

III. EXPERIMENTS AND RESULTS

A. Dataset

Our dataset is composed of data from three distinct data collections, in which the participants wore a *Samsung Galaxy Watch 4* [7]

TABLE 2. Characteristics of the dataset for each HRPF domain. Note that since the dataset was compiled from three data collections, not all physical tests were performed on all of them, leading to different sample sizes for each component.

HRPF component	Muscular strength	Muscular endurance	Flexibility
Test	Handgrip	Push-up	Sit-and-reach
Unit	kg	# of push-ups	cm
Mean (std)	37.23 ± 11.19	18.06 ± 14.16	22.07 ± 9.78
Age (yrs)	37.97 ± 10.31	37.92 ± 10.28	38.05 ± 10.96
Height (cm)	168.76 ± 9.55	168.79 ± 9.57	168.80 ± 9.71
Weight (Kg)	75.34 ± 15.99	75.37 ± 15.97	74.95 ± 15.68
Sex and sample size	M (234), F (247)	M (237), F (248)	M (183), F (207)

smartwatch during free-living for a period between 7 and 30 days and performed the guided physical tests following ACSM protocols [1] under controlled environment [14]. Free-living data was used as input to train our HRPF prediction models, while the physical tests were used as ground truth. The first and second data collections (CAEs: 53305121.5.0000.0068 and 61304822.7.0000.0068) were conducted at the *Instituto do Coração (InCor), Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo*, Brazil, between January 2022 to March 2022 and between September 2022 and March 2023, respectively. The third data collection (CAE: 58367822.5.0000.5599) was conducted at the *Samsung R&D Institute Brazil*, between October 2022 and May 2023. All experiments in these studies were performed in a clinical setting under the supervision of health professionals, followed the Declaration of Helsinki, Brazilian General Data Protection Law, and were approved by the pertinent institutional ethics committees. All participants provided informed consent. Note that not all data collections contained physical tests for all HRPF components that we predicted, so the sample size available for training each component was different. Moreover, participants were excluded if their data had values out of physiologically plausible range, missing input variables (Table 1), or did not use the smartwatch consistently (i.e., less than two days of activity, less than two exercise sessions, or less than 1,000 steps). A summary of the physical tests and their distributions, as well as the characteristics of the sample, is presented in Table 2.

B. Evaluation protocol

We evaluated all methods using the mean and standard deviation of 10 random train-test splits (80% for training and 20% for testing). The metrics of interest were the Pearson correlation (R) and mean absolute error (MAE) between the predictions and their ground truth. We tuned the hyperparameters of each approach using 5-fold cross-validation on each training set, and evaluated the compared methods under three different settings: i) using the feature sets from the original papers, ii) using profile and BIA features, and iii) using profile, BIA, and activity features. These feature groups are described in Table 1.

C. Comparative evaluation

We compare our approach to two recent works that proposed predictors for multiple HRPF components and to XGBoost [15], which is arguably the current state-of-the-art method for tabular data [16]. Kim et al. [5] used a multiple linear regression (MLR) with height, weight, body fat %, and body mass index (BMI) as input variables, which were selected using stepwise selection. Lee et al. [6] used a multilayer perceptron (MLP) with age, gender, height, weight, body fat %, waist circumference, and BMI as input variables, which were selected using a random forest-based approach. Since Lee et al. [6] evaluated several neural network configurations, for fairness, we considered all the configurations in their analysis (number of layers between 1–7, nodes per layer between 4–28) and report the best model selected via grid search. Moreover, since our dataset is different from theirs, we adapted the training hyperparameters to ensure model convergence. The networks were trained with the Adam optimizer ($LR = 10^{-4}, \beta_1 = 0.9, \beta_2 = 0.999$) and batch size = 8 for 1,000 epochs. For XGBoost, we report the best model using grid search (max_depth $\in [2, 4, 6]$, n_estimators $\in [50, 100, 200]$, min_child_weight $\in [1, 3, 5]$, subsample $\in [0.6, 0.8, 1.0]$, colsample_bytree $\in [0.6, 0.8, 1.0]$, and $LR = 5 \times 10^{-3}$). The results are presented in Table 3.

TABLE 3. Comparison of the proposed approach to state-of-the-art approaches used for HRPF prediction. The best results for each feature group are underlined, while the best results for each HRPF component, considering all feature groups, are indicated in gray.

		Original		Profile/BIA		Profile/BIA/Activity	
		R ↑	MAE ↓	R ↑	MAE ↓	R ↑	MAE ↓
Strength	MLR[5]	0.80 ± 0.02	5.15 ± 0.39	0.82 ± 0.02	4.94 ± 0.31	0.81 ± 0.02	4.97 ± 0.24
	MLP[6]	0.81 ± 0.02	5.13 ± 0.21	0.80 ± 0.03	5.15 ± 0.35	0.79 ± 0.03	5.29 ± 0.43
	XGBoost [15]	—	—	0.80 ± 0.02	5.86 ± 0.43	0.80 ± 0.02	5.89 ± 0.46
	Ours	—	—	0.82 ± 0.02	4.91 ± 0.32	0.81 ± 0.02	5.03 ± 0.29
Endurance	MLR[5]	0.54 ± 0.05	9.26 ± 0.47	0.60 ± 0.05	8.53 ± 0.56	0.64 ± 0.05	8.18 ± 0.81
	MLP[6]	0.60 ± 0.07	8.87 ± 0.63	0.62 ± 0.05	8.70 ± 0.53	0.64 ± 0.05	8.38 ± 0.75
	XGBoost [15]	—	—	0.57 ± 0.06	9.68 ± 0.44	0.63 ± 0.05	8.94 ± 0.42
	Ours	—	—	0.62 ± 0.06	8.41 ± 0.55	0.67 ± 0.05	7.97 ± 0.66
Flexibility	MLR [5]	0.46 ± 0.04	6.95 ± 0.41	0.51 ± 0.04	6.80 ± 0.37	0.50 ± 0.06	7.03 ± 0.35
	MLP [6]	0.46 ± 0.04	7.14 ± 0.43	0.46 ± 0.06	7.13 ± 0.30	0.44 ± 0.07	7.19 ± 0.42
	XGBoost [15]	—	—	0.43 ± 0.06	7.27 ± 0.53	0.49 ± 0.04	7.07 ± 0.48
	Ours	—	—	0.51 ± 0.04	6.77 ± 0.36	0.56 ± 0.05	6.59 ± 0.31

For muscular strength, all methods performed similarly. While our model achieved the best results, it was only marginally better than a multiple linear regression (MLR). MLP was more susceptible to overfitting, increasing its prediction error the more features were available, while XGBoost performed the worst among all methods. Including activity features did not improve, and actually degraded, the performance of all muscular strength models. For muscular endurance, on the other hand, all methods improved with the addition of activity features, with the proposed approach achieving the largest improvement. Similarly, flexibility prediction improved substantially with the addition of activity features, but only for XGBoost and our proposed model. For MLR and MLP, performance degraded, highlighting the need for adequate preprocessing to avoid overfitting when using activity features.

D. Ablation analysis

To understand the effect of each component in our models, we present an ablation analysis of our muscular endurance and flexibility prediction models on Table 4. We show the performance of the base regression model and its combination with preprocessing for two cases: 1) on all features, 2) just on activity features. For the two predictors, preprocessing improves results over their base regressors, but it is more effective when limited to activity data. Moreover, for the flexibility model, preprocessing on all features increases error, despite improving correlation. These results suggest that activity data enables substantial improvement over using just profile and BIA data, but that it also needs specific preprocessing to be leveraged effectively.

TABLE 4. Ablation analysis of muscular endurance and flexibility predictors. The best results are indicated in gray.

			R↑	MAE↓
Endurance	Poisson regression		0.64 ± 0.04	7.75 ± 0.57
	+ Spline	all features	0.65 ± 0.03	7.62 ± 0.54
		activity only	0.67 ± 0.03	7.53 ± 0.58
Flexibility			R↑	MAE↓
	Ridge regression		0.47 ± 0.06	6.97 ± 0.34
	+ PCA	all features	0.49 ± 0.06	7.11 ± 0.33
		activity only	0.56 ± 0.05	6.59 ± 0.31

IV. DISCUSSION

Predicting HRPF domains from wearable devices is challenging, especially those addressed in this work, since they are not directly measured by sensors available in consumer-grade smartwatches. Body composition can be measured directly by current smartwatches through BIA sensors [7], while cardiorespiratory fitness ($\dot{V}O_2$max) prediction is mature enough to be available and widely used in several wearable devices [17]–[19]. In contrast, predicting the remaining HRPF components acknowledged by the ACSM (muscular strength, muscular endurance, and flexibility) has not been tackled outside the clinical setting. In this work, for the first time, we have addressed the prediction of these components from wearable data obtained from consumer-grade smartwatches during free-living, thus providing a valuable tool for health promotion and engagement by enabling continuous and non-intrusive fitness tracking without the need to test on a clinical setting. Our results show that our proposed linear models with component-specific preprocessing provide predictions that achieve moderate to high correlation with the ground truth [20] and outperform general approaches previously used to predict these components from large-scale clinical data.

We have demonstrated that wearable devices are not limited to providing approximations to clinical data — they enable leveraging free-living activity data, which are unavailable on clinical settings and provide complementary information that is highly discriminative of HRPF, as demonstrated by the substantial increase in performance on muscular endurance and flexibility prediction. We have also showed how to preprocess these features to leverage them effectively with linear models, which is not only helpful but necessary to avoid degrading performance in the case of flexibility prediction.

It is worth noting that the relation between activity features and HRPF status is more meaningful when these features reflect habits instead of occasional short-term activity. Only chronic exposition to activity patterns affects HRPF status, such that on the very short term they might be misleading and require careful consideration. A limitation of our experiments was that the participants in our data wore the smartwatch from 7 to 30 days, which does not allow validating if the activity features actually reflect habits. However, their discriminative power for HRPF prediction suggests that it was predominantly the case.

V. CONCLUSION

In this paper, we proposed predictors for all health-related physical fitness components acknowledged by the ACSM that are not provided by current consumer-grade smartwatches. Our predictors enable a complete characterization of health-related fitness components from free-living data based on component-specific models designed for this setting. This enables leveraging smartwatch sensors to estimate higher-level fitness components that are used in well-established guidelines for exercise prescription [1], increasing the effectiveness and applicability of wearable devices for health improvement. Moreover, we show that activity data, which is available only through wearable devices, can improve the discriminability of HRPF prediction, which highlights the importance of such devices as more than just tools for approximating clinical data, but to actually complement them.

ACKNOWLEDGMENT

The results presented in this work were funded by Samsung Eletrônica da Amazônia Ltda., under the terms of Brazilian Informatics Law 8.248/91.

REFERENCES

[1] American College of Sports Medicine, *ACSM's guidelines for exercise testing and prescription*, 11th ed. Lippincott Williams & Wilkins, 2022.

[2] C. Akalan, R. Robergs, and L. Kravitz, "Prediction of vo2max from an individualized submaximal cycle ergometer protocol," *J. of Exercise Physiol. Online*, vol. 11, no. 2, pp. 1–17, 2008.

[3] A. T. Tveter, H. Dagfinrud, T. Moseng, and I. Holm, "Health-related physical fitness measures: reference values and reference equations for use in clinical practice," *Arch. of Phys. Med. and Rehabil.*, vol. 95, no. 7, pp. 1366–1373, 2014.

[4] J. Lopes, S. T. Grams, E. F. da Silva, L. A. de Medeiros, C. M. M. de Brito, and W. P. Yamaguti, "Reference equations for handgrip strength: Normative values in young adult and middle-aged subjects," *Clin. Nutrition*, vol. 37, no. 3, pp. 914–918, 2018.

[5] S.-W. Kim, H.-Y. Park, H. Jung, J. Lee, and K. Lim, "Estimation of health-related physical fitness using multiple linear regression in korean adults: national fitness award 2015–2019," *Frontiers in Physiol.*, vol. 12, 2021.

[6] S.-H. Lee, H.-S. Ju, S.-H. Lee, S.-W. Kim, H.-Y. Park, S.-W. Kang, Y.-E. Song, K. Lim, and H. Jung, "Estimation of health-related physical fitness (HRPF) levels of the general public using artificial neural network with the national fitness award (NFA) datasets," *Int. J. of Environ. Res. and Public Health*, vol. 18, no. 19, 2021.

[7] Samsung, "Galaxy Watch4 Smartwatch | Samsung US," 2024, [Access: May-15-2025]. [Online]. Available: "https://www.samsung.com/us/app/watches/galaxy-watch4/"

[8] S. R. Bozeman, D. C. Hoaglin, T. M. Burton, C. L. Pashos, R. H. Ben-Joseph, and C. S. Hollenbeak, "Predicting waist circumference from body mass index," *BMC Med. Res. Methodol.*, vol. 12, 2012.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. of the Royal Stat. Soc. Ser. B: Stat. Methodol.*, vol. 58, no. 1, 1996.

[10] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

[11] M. H. Katz, *Multivariable analysis: a practical guide for clinicians*. Cambridge University Press, 2006.

[12] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, 1970.

[13] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary rev.: comput. stat.*, vol. 2, no. 4, pp. 433–459, 2010.

[14] M. P. Bayles, *ACSM's exercise test. and prescription*. Lippincott williams & wilkins, 2023.

[15] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. of the ACM SIGKDD Int. Conf. on Knowl. Discovery and Data Mining*, 2016.

[16] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci, "Deep neural networks and tabular data: A survey," *IEEE Trans. on neural netw. and learn. sys.*, 2024.

[17] Garmin, "Fitness Tracking and VO2 Max," https://www.garmin.com/en-US/garmin-technology/health-science/fitness-tracking/, 2024, Access: Sep-13-2024.

[18] Apple, "Using Apple watch to estimate cardio fitness with VO_2max," Tech. Rep., 2021.

[19] Samsung, "Work out with your Wear OS Galaxy Watch," https://www.samsung.com/us/support/answer/ANS10003401/, 2024, Access: May-15-2025.

[20] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.